THE HONG KONG
POLYTECHNIC UNIVERSITY
香港理工大學
Pao Yue-kong Library
包玉剛圖書館

## Copyright Undertaking

This thesis is protected by copyright, with all rights reserved.

**By reading and using the thesis, the reader understands and agrees to the following terms:**

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

**IMPORTANT**

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact lbsys@polyu.edu.hk providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

# A MULTI-TIERED AI-ENHANCED APPROACH FOR OPERATIONAL LANDSLIDE PREDICTION AND STABILITY ASSESSMENT

**KYRILLOS MAGDY PHILIPE EBRAHIM**

**PhD**

**The Hong Kong Polytechnic University**

**2026**

**The Hong Kong Polytechnic University**

**Department of Building and Real Estate**

# A Multi-Tiered AI-Enhanced Approach for Operational Landslide Prediction and Stability Assessment

**Kyrillos Magdy Philipe EBRAHIM**

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

**December 2025**

# CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

(Signed)

Kyrillos Magdy Philipe EBRAHIM (Name of student)

# DEDICATION

With deepest humility and gratitude, I dedicate this thesis to my Creator – God – for the abundant grace that sustained me throughout this work. I also lovingly dedicate this to my wife and daughter, whose unwavering support, patience, and encouragement have been indispensable to the completion of this study.

# ABSTRACT

Hong Kong, with 60% of its terrain on natural slopes, faces elevated landslide risks from extreme precipitation and typhoons, making it a critical case study for this hazard. Within this context, contemporary landslide early warning systems are constrained by a critical efficacy–efficiency trade-off. Simplistic regional models (e.g., Intensity-Duration thresholds) generate high false-alarm rates by neglecting spatial variability and rainfall randomness, while sophisticated physical models are computationally challenging for real-time use. Persistent gaps include a focus on surface displacement over internal hydro-mechanical processes, vulnerability to data loss during extreme weather, and a lack of automation in stability-analysis workflows. To address these challenges, the primary objective of this research is to develop an operational, three-tiered framework that bridges the gap between data-driven and deterministic landslide models through: (1) a foundational critical review that synthesizes fragmented knowledge on monitoring technologies, prediction models and failure mechanisms; (2) a data-driven regional screening model to capture spatiotemporal trigger randomness; (3) a deep learning-enhanced subsystem for forecasting critical subsurface precursors; and (4) an automated stability tool that resolves the parameterization bottleneck and delivers end-to-end Factor of Safety analysis.

The pursuit of these objectives is operationalized through a tripartite framework founded on a comprehensive review of monitoring and prediction paradigms: **a)** Tier 1 establishes a multi-data-driven classifier that moves beyond binary decisions to probabilistically categorize slopes as low-risk, high-risk, or ambiguous “warning” cases requiring deeper analysis. The model was developed from a uniquely curated 1984–2022 regional dataset of Hong Kong natural hillside incidents, combining antecedent rainfall sequences and spatial features. Methodologically, it addresses real-world spatial–temporal imbalance with tailored preprocessing and applies a dual-threshold strategy for nuanced risk gradation; **b)** Ambiguous cases identified by the probability model activate the deep learning-

enhanced subsurface tier 2. Two methodological innovations are central: (1) multi-parallel LSTM networks with a novel independence-oriented time-series windowing for operationality; and (2) a multi-node prediction framework technique designed to handle data loss and forecast critical precursors (volumetric water content, matric suction, earth pressure). This tier is operationalized as OWHK (Operational VWC Forecasting Model for Shallow Layers in Hong Kong), which fuses LSTM-based forecasting; and **c)** Tier 3 introduces the Automated Parameterization Tool (APHs), which uses multi-LSTM networks to derive advanced soil parameters directly from standard lab data. Building on APHs, a hybrid multi-deep learning pathway predicts a regional Soil Saturation Index (SSI) from real antecedent rainfall sequences and terrain factors and computes the infinite-slope Factor of Safety, directly linking hydrologic response to geotechnical mechanics.

The efficacy of this framework is demonstrated through rigorous testing and validation. The Tier 1 classifier achieved an Area Under the ROC Curve of 0.92, successfully categorizing slopes while reducing false alarms. The Tier 2 OWHK model forecasted VWC with high accuracy (MAE < 0.6%, $R^2$ > 0.92) across 15 sensors at three instrumented sites. Complementary LSTMs predicted earth pressure and matric suction with RMSEs of 1.62 kPa and 3.18 kPa, respectively. An extended VWC framework provides both point and interval predictions (MAE 0.35–1.22%) and propagates VWC into FOS calculations ($R^2$ 0.89). The APHs workflow achieved a parameterization accuracy of 99.4% through seven novel multi-parallel deep LSTM networks. This accuracy was validated against a conventional reference case (99.6%) and an experimental laboratory study, both showing strong agreement between simulated and measured results. The tool significantly accelerates parameterization, delivering precise results in seconds. The Tier 3 hybrid SSI-FOS pathway was externally validated against the deterministic GeoStudio model for the 2005 Fei Ngo Shan landslide, showing excellent agreement (FOS MAE 0.06–0.07). It also correctly assessed the stability conditions of the independent 2023 Yiu Hing Road landslide.

Collectively, these results lead to several key conclusions: (a) A tiered structure integrating data-driven AI with physics-based modeling is essential for achieving both regional scalability and geotechnical accuracy in operational early warning systems; (b) The 1D-CNN architecture demonstrates superior performance for prediction, offering optimal balance between precision and computational efficiency compared to LSTM and GRU alternatives; (c) For subsurface forecasting, a multi-parallel LSTM structure with specialized networks targeting distinct hydrological regimes proves more effective than single-model approaches; (d) Volumetric Water Content emerges as the most critical and physically consistent predictive feature addressing rainfall spatiotemporal uncertainty; (e) The APHs tool represents a breakthrough in automation with accuracy surpassing traditional empirical methods; (f) A hybrid pathway fusing data-driven Soil Saturation Index with physics-based Factor of Safety calculations enables successful regional-scale, time-aware stability assessment; (g) Hillslope geometry constitutes a first-order control mechanism, with convergent slopes showing particular vulnerability to rapid moisture buildup and antecedent-condition-driven failure timing; (h) The "effective non-rainfall duration" provides a crucial metric for quantifying slope recovery time.

In summary, the significance of this work is threefold, directly addressing the initial research gaps: a) Theoretical: It establishes a direct, learned link between rainfall triggers, subsurface hydrology, and geotechnical stability mechanics; b) Methodological: It pioneers novel techniques including a dual-threshold risk grader, independence-oriented windowing for operational forecasting, multi-parallel LSTM architectures, and a hybrid SSI-to-FOS coupling; c) Practical: It delivers a suite of operational tools: a multi-model, multi-event landslide forecaster, OWHK, a subsurface response predictor, APHs, and an SSI–FOS automated stability tool.

# PUBLICATIONS ARISING FROM THE THESIS

**Journal Articles (Accepted/Published)**

1- Ebrahim, K., Taiwo, R. & Zayed, T. (2026). Data-driven interface for shallow rainfall-induced landslide forecasting considering trigger randomness and spatiotemporal dependencies. Landslides. https://doi.org/10.1007/s10346-025-02679-5

2- Ebrahim, K, Taiwo, R., & Zayed, T. (2025). OWHK: Operational volumetric water content forecasting model for shallow rainfall-induced landslides in Hong Kong. Engineering Geology, 108228, 355. https://doi.org/10.1016/j.enggeo.2025.108228

3- Ebrahim, K, Taiwo, R., Zayed, T., & El-Hamalawi, A. (2025). An automatic parameterization tool for hardening soil model. Journal of Rock Mechanics and Geotechnical Engineering. https://doi.org/10.1016/j.jrmge.2025.07.041

4- Ebrahim, K., Abdelkader, E., M., Zayed, T., & Meguid, M. A. (2025). Enhancing Landslide Prediction with Deep Learning: Insights into Soil Moisture Dynamics. Geoenvironmental Disasters, 12, 36 (2025). https://doi.org/10.1186/s40677-025-00333-9

5- Ebrahim, K., Gomaa, S. M. M. H., Zayed, T., & Alfalah, G. (2025). Rainfall-Induced Landslide Prediction Models, Part I: Empirical–Statistical and Physically Based Causative Thresholds. Water, 17(22), 3273. https://doi.org/10.3390/w17223273.

6- Ebrahim, K., Gomaa, S. M. M. H., Zayed, T., & Alfalah, G. (2024). Rainfall-induced landslide prediction models, part ii: deterministic physical and phenomenologically models. Bulletin of Engineering Geology and the Environment, 83(3), 85. https://doi.org/10.1007/s10064-024-03563-7

7- Ebrahim, K., Gomaa, S. M. M. H., Zayed, T., & Alfalah, G. (2024). Recent Phenomenal and Investigational Subsurface Landslide Monitoring Techniques: A Mixed Review. Remote Sensing, 16(2), 385. https://www.mdpi.com/2072-4292/16/2/385

8- Ebrahim, K., Fares, A., Faris, N., & Zayed, T. (2024). Exploring Time Series Models for Landslide Prediction: A Literature Review. Geoenvironmental Disasters, 11, 25. https;//doi.org/10.1186/s40677-024-00288-3

**Journal Articles (Under review)**

1- Ebrahim, K., Abdelkader, E., M., Zayed, T., & Rashid, A. S. A. (Under 1st round review). Predicting Landslide Subsurface Hydrological and Mechanical Responses Using LSTM Deep Learning. Disaster Prevention and Management.

2- Ebrahim, K., Ali, E., Shahzada, N., & Zayed, T. (Under 1st round review). A hybrid deep learning–physics framework for regional rainfall-induced landslide early warning: application to Hong Kong's natural slopes. Geomatics, Natural Hazards and Risk.

**Conference Papers**

1- Ebrahim, K., & Zayed, T. (2024). Factors controlling the triggering randomity for rainfall-induced landslides. In Proceedings of the 18th African Regional Conference on Soil Mechanics and Geotechnical Engineering (ISSMGE), Algiers, Algeria, 6-9 October 2024. https://www.issmge.org/publications/publication/factors-controlling-the-triggering-randomity-for-rainfall-induced-landslides

2- Ebrahim, K., & Zayed, T. (2025). Evaluating the impact of hillslope geometry on landslide vulnerability amid climate change.In Proceedings of the 48th Australasian Universities Building Education Association (AUBEA 2025) Conference, Canberra, Australian Capital Territory, Australia, November 26–28.

# ACKNOWLEDGMENTS

First and foremost, I thank God, my pillar of strength, for the countless blessings throughout this journey.

I wish to express my sincere gratitude to my research supervisor, Prof. Tarek Zayed, for his unwavering support, guidance, and encouragement during my studies. I am deeply grateful for the wisdom, humility, and kindness you have shown me, Prof. Zayed, and I will always cherish the knowledge and insights gained under your supervision. My heartfelt thanks also go to my colleagues at the SIMS Lab for their collaboration and support.

I am indebted to the Pao Yue-kong Library for providing invaluable access to reference books, journals, and e-resources that have been instrumental in my research. I am equally thankful to the Department of Building and Real Estate and the Department of Civil and Environmental Engineering for granting access to state-of-the-art facilities, equipment, and laboratories that are essential in completing this study. I gratefully acknowledge the financial support provided by The Hong Kong Polytechnic University.

I extend my deepest appreciation to the Head of the Geotechnical Engineering Office and the Director of Civil Engineering and Development, the Government of the Hong Kong Special Administrative Region, for permitting me to use the material produced by them. Furthermore, I gratefully acknowledge the Innovation and Technology Support Programme (ITSP) [Grant No. ITS/033/20FP], Hong Kong SAR

Last but certainly not least, I wish to thank my family for instilling in me a sense of pride and purpose. Your unconditional love and unwavering support have been the foundation of my success, and I am forever grateful for everything you have done for me.

# TABLE OF CONTENTS

**CERTIFICATE OF ORIGINALITY** ..... i

**DEDICATION** ..... ii

**ABSTRACT** ..... i

**PUBLICATIONS ARISING FROM THE THESIS** ..... iv

**ACKNOWLEDGMENTS** ..... vi

**TABLE OF CONTENTS** ..... vii

**LIST OF FIGURES** ..... xii

**LIST OF TABLES** ..... xx

**LIST OF NOTATIONS AND ABBREVIATIONS** ..... xxiii

**CHAPTER 1: INTRODUCTION** ..... 1

1.1. INTRODUCTION ..... 1

1.2. PROBLEM STATEMENT ..... 6

1.3. RESEARCH GAPS ..... 8

1.4. RESEARCH OBJECTIVES ..... 9

1.5. RESEARCH METHODOLOGY ..... 9

1.6. THESIS ORGANIZATION ..... 12

**CHAPTER 2: LITERATURE REVIEW** ..... 13

2.1. INTRODUCTION ..... 13

2.2 SEARCH STRATEGY AND FRAMEWORK ..... 13

2.3. LANDSLIDE MONITORING TECHNIQUES ..... 16

*2.3.1. Surface VS Subsurface Monitoring Techniques* ..... 16

*2.3.2. Subsurface Mechanical Response Monitoring.* ..... 18

*2.3.3. Subsurface Hydrological Monitoring.* ..... 22

*2.3.4. Subsurface Warning Monitoring.* ..... 24

*2.3.5. Data Collection Systems.* ....26

*2.3.6. Subsurface Deployment Techniques.* ....27

2.4. LANDSLIDE PREDICTION TECHNIQUES....31

*2.4.1. Local VS National Prediction Models.* ....31

*2.4.2. Empirical-Statistical I-D Thresholds.* ....32

*2.4.3. Physically Based Causative Thresholds.* ....34

*2.4.4. Physically Deterministic Models.* ....43

2.5. RESEARCH GAPS ....50

2.6. SUMMARY ....51

**CHAPTER 3: METHODOLOGY** ....54

3.1. INTRODUCTION ....54

3.2. LANDSLIDE PREDICTION MODEL (TIER 1) ....54

*3.2.1. Landslide Prediction Model Development* ....54

*3.2.2. Landslide Prediction Model Assumptions* ....67

3.3. SUBSURFACE SENSOR-BASED PREDICTION MODEL (TIER 2). ....69

*3.3.1. VWC Prediction Model Development* ....69

*3.3.2. VWC Prediction Model Assumptions* ....78

*3.3.3. Hypertuning Strategy for Enhanced Model Performance* ....80

*3.3.4. Distributed Multi-Sensor (Subsurface) Prediction Model* ....91

3.4. SLOPE STABILITY ANALYSIS AND AUTOMATION FRAMEWORK (TIER 3) ..95

*3.4.1. Automated Parametrization Model* ....95

*3.4.2. Numerical Slope Stability Modeling (Focused Site Specific Analysis)* ....107

*3.4.3. Automated Slope Stability Analysis Model (A Hybrid DL–Physics Model)* ....112

3.5. SUMMARY ....117

**CHAPTER 4: DATA COLLECTION AND CASE STUDY** ....118

4.1. INTRODUCTION ..... 118
4.2. HISTORICAL LANDSLIDE RECORDS DATASET ..... 118
4.3. GEOTECHNICAL INVESTIGATION AND LABORATORY DATASET ..... 124
4.4. AUGMENTED DATASET ..... 126
*4.4.1. Subsurface Landslide Response Augmentation.* ..... 126
*4.4.2. Laboratory Tests Augmentation.* ..... 127
4.5. SENSOR-BASED DATASET ..... 130
4.6. SUMMARY ..... 135
**CHAPTER 5: RESULTS AND DISCUSSION** ..... 136
5.1. INTRODUCTION ..... 136
5.2. LANDSLIDE PREDICTION MODEL (TIER 1) ..... 136
*5.2.1. Model Tuning and Training* ..... 136
*5.2.2. Fei Ngo Shan Landslide Validation* ..... 138
*5.2.3. Model Evaluation with Feature Analysis* ..... 140
*5.2.4. Model Warning Thresholds* ..... 146
*5.2.5. Model Validation and Testing* ..... 148
*5.2.6. Optimal Model Selection, Benchmarking, and Uncertainty* ..... 151
5.3. SUBSURFACE SENSOR-BASED PREDICTION MODEL (TIER 2). ..... 153
*5.3.1 VWC Prediction Operational Model (OWHK)* ..... 153
*5.3.2 Hypertuning Strategy for Enhanced Model Performance.* ..... 168
*5.3.3 Distributed Multi-Sensor (Subsurface) Prediction Model* ..... 187
5.4. SLOPE STABILITY ANALYSIS AND AUTOMATION FRAMEWORK (TIER 3) 206
*5.4.1. Automated Parametrization Model (APHs)* ..... 206
*5.4.2. Automated Slope Stability Analysis Model (A Hybrid DL–Physics Model)* ..... 217
*5.4.3. Factors Controlling Rainfall-Induced Landslides: A Parametric Study* ..... 222

*5.4.4. Geometry Effects on Rainfall-Induced Landslides: A Parametric Study* .....230

*5.4.5. Numerical Slope Stability Modeling (Focused Site Specific Analysis)* .....234

5.5. SUMMARY .....241

**CHAPTER 6: IMPLEMENTATION AND DEPLOYMENT OF WEB-BASED PREDICTIVE TOOLS** .....244

6.1. INTRODUCTION .....244

6.2. LANDSLIDE PREDICTING TOOL .....244

6.3. OPERATIONAL VWC FORECASTING TOOL (OWHK) .....250

6.4. AUTOMATED PARAMETRIZATION TOOL (APHs) .....259

6.5. AUTOMATED STABILITY ANALYSIS TOOL .....266

6.6. SUMMARY .....271

**CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS** .....272

7.1. INTRODUCTION .....272

7.2. SUMMARY OF THE KEY FINDINGS .....272

*7.2.1. Mechanisms, Predicting Approaches, and Monitoring Systems* .....272

*7.2.2. Landslide Prediction Model* .....275

*7.2.3. Subsurface Sensor-Based Predictions* .....276

*7.2.4. Slope Stability Analysis and Process Automation* .....279

*7.2.5. Synthesis of Key Findings and Conclusions* .....285

7.3. SIGNIFICANCE AND CONTRIBUTION OF THE RESEARCH .....287

*7.3.1. Theoretical Contributions* .....287

*7.3.2. Practical Contributions, Cost-Benefit Considerations, and Policy Integration* ..289

*7.3.3. Broader Contributions* .....291

7.4. RESEARCH LIMITATIONS .....292

7.5. RECOMMENDATIONS FOR FUTURE WORK .....293

*7.5.1. Enhancements for the Developed Research* .....294

*7.5.2. Extensions and Broader Applications*.....296

*7.5.3. Novel Research Directions and Spin-Off Technologies*.....297

7.3. SUMMARY .....298

**APPENDIX A**.....299

**REFERENCES**.....317

# LIST OF FIGURES

Figure 1.1. Landslide causative factors.....2

Figure 1.2. Po Shan Road's 1972 collapse (Cheng et al., 2009)(Photo by Geotechnical .....3

Figure 1.3. Fatal non-earthquake landslides from 2004 to 2016 (Froude & Petley, 2018) .....3

Figure 1.4. Schematic representation of shallow rainfall-induced landslides. .....4

Figure 1.5. Sample landslide subsurface responses: VWC (WTDR1A), suction (TC1C), earth pressure (EP1A), and groundwater (BH7) (Geotechnical Engineering Office 2010a-c). .....8

Figure 1.6. Research methodology workflow covering all study objectives. .....11

Figure 2.1. Methodological flow for systematic review. .....15

Figure 2.2. Methodological flow for scientometric analysis. .....15

Figure 2.3. PRISMA: Screening and selection flow diagram. .....16

Figure 2.4. Surface monitoring techniques. .....17

Figure 2.5. Subsurface monitoring systems. .....18

Figure 2.6. Schematic representation of several subsurface movement systems. .....21

Figure 2.7. Sliding force monitoring system (Adapted with permission from Tao et al., 2019). .....22

Figure 2.8. An optimized version of the optical soil moisture instrument (Leone et al., 2017) .....24

Figure 2.9. Comparison between elastic wave velocities and conventional methods (Adapted with permission from (Irfan et al., 2017), Elsevier) .....25

Figure 2.10. a) Shear wave signal versus time; b) Elastic wave velocity during rainfall events (Adapted with permission from (Irfan et al., 2017), Elsevier).....26

Figure 2.11. Wireless network structure (Modified from (Jeong et al., 2020; Yueshun & Wei, 2013)).....27

Figure 2.12. Subsurface deployment flowchart. .....29

Figure 2.13. Subsurface monitoring system deployment visualization (Sheikh et al., 2021)..29
Figure 2.14. Subsurface monitoring system deployment visualization (Sheikh et al., 2021)..31
Figure 2.15. Landslide prediction techniques classification. ..... 32
Figure 2.16. Rainfall pattern analysis in terms of IETD. ..... 33
Figure 2.17. Prediction accuracy (AUC) and false positive rate (FPR) for I-D and API thresholds ..... 33
Figure 2.18. Displacement response with rainfall and groundwater. ..... 35
Figure 2.19. Artificial models in landslide displacement prediction. ..... 36
Figure 2.20. Accuracies in predicting surface displacement. ..... 37
Figure 2.21. The coupled hydrological-slope instability model. ..... 41
Figure 2.22. Typical pore water pressure profile (Modified from Zhang et al., 2011) ..... 44
Figure 2.23. Green Ampt profile ..... 45
Figure 2.24. Stability analysis of a shallow infinite landslide ..... 45
Figure 2.25. Graphical summary for landslide prediction strategies ..... 53
Figure 3.1. Detailed methodology, deployments, and processing steps for all study objectives. ..... 56
*Figure 3.2. Spatiotemporal modeling workflow* ..... 59
Figure 3.3. a) LSTM cell structure ..... 61
*Figure 3.4. Schematic of the spatiotemporal DL model architecture (LSTM as an example)*. 65
*Figure 3.5.Subsurface VWC operational modeling workflow* ..... 71
*Figure 3.6. Simplified operational VWC LSTM architecture with the parameter search space.* ..... 72
*Figure 3.7. Data sampling approach using independence-oriented time series windowing.*..73
*Figure 3.8. The operational multi-LSTM framework* ..... 75

*Figure 3.9. a) Seepage under rainfall; b) Moisture content-based prediction; c) VWC versus displacement and velocity inverse (modified from Fukuhara, 2022; Tohari et al., 2007)*.....81
Figure 3.10. Time series modeling process. .....83
Figure 3.11. Time series model hypertuning flow chart .....84
Figure 3.12. Schematic representation of the shared network model structure .....88
Figure 3.13. Multi-sensor prediction modeling workflow. .....92
*Figure 3.14. Traditional parameterization workflow: (a–c) the experimental approach; (d–e) strength constants parameterization; and (f–j) stiffness constants parameterization.* .....98
Figure 3.15. The workflow of the proposed APHs tool. .....99
*Figure 3.16. Axisymmetric finite element model (SIGMA/W)*.....102
Figure 3.17. APHs model architecture for a single parameter. .....105
Figure 3.18. Schematic view of the parametric study highlighting the converging and diverging boundaries. .....109
Figure 3.19. 3D schematic view of curvature features (modified from Giuseppe et al., 2016). .....110
Figure 3.20. Schematic view of the rainfall controlling features. .....110
Figure 3.21. Schematic view of the hybrid model design and workflow. .....114
Figure 3.22. Schematic diagram of the DL model (CNN as an example) architecture. .....115
*Figure 4.1. Map of Hong Kong showing locations of the validation case study, raingauges, landslides, and non-landslide incidents. Basemap imagery provided by Esri and its partners.* .....121
*Figure 4.2. Frequency distribution of rainfall intensity per time step. Blue indicates non-landslide cases, and orange indicates landslide events.* .....121
*Figure 4.3. Suction dynamics at a Hong Kong landslide (Tung Chung as an example), illustrating the Initial Month concept (Geotechnical Engineering Office, 2010b)*.....122

*Figure 4.4. Cont. b) Key topographic attributes from the DEM characterizing the study area's geomorphology: (a) Elevation; (b) Slope; (c) Aspect; (d) Curvature; (e) TRI; (f) TWI*.....123

*Figure 4.5. Fei Ngo Shan Landslide soil profile cross-sectional elevation (modified from Maunsell Geotechnical Services Ltd., 2007)* .....125

*Figure 4.6. Sample subcases of the numerically simulated VWC responses.* .....128

*Figure 4.7. HS parameters frequency distribution (augmented dataset)*.....128

*Figure 4.8. HS parameters pair plot relationships (augmented dataset)* .....129

*Figure 4.9. The augmented stress-strain curves* .....130

*Figure 4.10. Sensor-based study area map and representative sensor record* .....132

*Figure 4.11. The sharing network with a monitoring array among several spatial locations* .....133

*Figure 5.1. Model coarse-to-fine tuning*.....137

*Figure 5.2. Fei Ngo Shan landslide boundary conditions* .....139

*Figure 5.3. Numerically simulated vs. monitored groundwater levels* .....140

*Figure 5.4. Numerically simulated vs. monitored suction at T2*.....140

*Figure 5.5. Multi-rainfall thresholds for several features: CR1W; CR2W; CR3W; CR4W; MRI; RD (definitions in Table 4.1). Figure A6 details the confusion matrix analysis.*.....141

*Figure 5.6. Comparison of a non-temporal vs. temporal models: a) Evaluation metrics*.....142

*Figure 5.7. Spatiotemporal LSTM model permutation-based feature importance analysis*..145

*Figure 5.8. Threshold curves and corresponding confusion matrices (LSTM as an example)* .....148

*Figure 5.9. Model Monte Carlo mean predictions vs. the numerical stability* .....150

*Figure 5.10. Monte Carlo predictions for the landslide event alongside the mean standard deviations across all events*.....151

*Figure 5.11. (a) Kernel density estimate of VWC measured at 0.50 and 1.50 m; (b) Frequency distribution of rainfall patterns for the corresponding subcases* ..... 153

*Figure 5.12. VWC operation model coarse-to-fine model tuning* ..... 155

*Figure 5.13. Comparison of predictions among different scenarios: a) the traditional method; b) the traditional method with improved features; c) the proposed improved method* ..... 157

*Figure 5.14. Scatter plot: predictions vs. actual data (a) single LSTM (15 sensors); (b) LSTM for 0.50 m depth; (c) LSTM for 1.50 m depth* ..... 160

*Figure 5.15. Predicted vs. actual VWC across the multi-operational models* ..... 161

*Figure 5.16. Operational VWC permutation-based evaluation of feature importance* ..... 164

*Figure 5.17. Site-specific vs. operational predictions for some sensors as an example* ..... 164

*Figure 5.18. Deterministic and stochastic evaluation of the operational model performance* ..... 166

*Figure 5.19. Statistical characteristics of the input Features: a) Pa Mei; b) Tung Chung* ... 169

*Figure 5.20. Data distributions for training (red) and validation (blue) sets from some sensors in the Pa Mei as an example. Y-axis: rainfall and evaporation (mm/day); VWC (%)* ..... 170

*Figure 5.21. Training and validation learning curves* ..... 171

*Figure 5.22. Model selection comparison: a) MSE; b) RMSE; c) MAE; d) R2* ..... 173

*Figure 5.23. Antecedent window size (W) selection: a) depth A; and b) depth B: (right panel) Taylor diagram; (left panel) $R^2$ vs. benchmarks* ..... 176

*Figure 5.24. Independent tuning search: a) dropout (1st run); b) dropout (2nd run); c) Leaky ReLU (VWC_2_C).* ..... 177

*Figure 5.25. Dependent tuning search heat map using random search (VWC_2_C): a) 1st trial random search* ..... 178

*Figure 5.26. Comparison between several training and validation sets (VWC_2_C).* ..... 180

*Figure 5.27. MAE comparison between several optimization techniques (right) and several loss functions (left) (VWC_2_C).* ..... 180

*Figure 5.28. Interval prediction and scatter plots for the VWC at three depths: (a) 0.50 m, (b) 1.50 m, (c) 2.50 m.* ..... 181

*Figure 5.29. FOS based on monitored and predicted VWC: a) sample of estimated suction values; b) FOS with 95% confidence interval and absolute error; c) FOS scatter plot; and d) residual distribution of FOS.* ..... 184

Figure 5.30. The corrected VWC (WTDR1A, B) - Tsing Shan. The Y-axis unit WTDR is %. ..... 189

Figure 5.31. a) Actual versus corrected JFT1B; b) SWCC fitted using Fredlund-Xing eq. .. 190

Figure 5.32. Data distribution in the training and validation sets with a ratio of 73%:27%, for WTDR sensors as an illustration – Tung Chung. ..... 190

Figure 5.33. The training learning curve with different batch sizes: a) 500 and b) 16. ..... 191

Figure 5.34. Pearson correlation coefficient among the input features and the output target. ..... 191

Figure 5.35. Prediction analysis of EP2A at Tung Chung. ..... 194

Figure 5.36. Prediction analysis of the EP2P at Tsing Shan. ..... 195

Figure 5.37. Comparative analysis between the predictions of the six locations at the Tung Chung and Tsing Shan case studies. ..... 196

Figure 5.38. Prediction analysis of the TC1B at Tung Chung. ..... 197

Figure 5.39. Prediction analysis of the TC1B at the Tung Chung extended case study. ..... 197

Figure 5.40. Prediction results of TC at Tung Chung with two different dataset lengths..... 199

Figure 5.41. TC1 comparative analysis at different depths and different triggering conditions. ..... 199

Figure 5.42. Prediction analysis of the JFT: a, b) at Tung Chung; c, d) at Tsing Shan..... 201

Figure 5.43. Performance comparison of various sensors at two different case studies.....202
Figure 5.44. JFT comparative analysis at different depths, triggering conditions, and two cases. .....202
Figure 5.45. Prediction analysis of the WTDR2C at Tung Chung. .....203
Figure 5.46. Density maps of hyperparameters vs. validation loss (APHs). .....208
*Figure 5.47. Model evaluation process (1st sampling dataset).* .....210
*Figure 5.48. APHs model training learning curves.* .....210
Figure 5.49. Deterministic (training and validation sets) and stochastic (test set) predicted strength parameters versus actual HS parameters with stochastic test MAE and residuals. .212
Figure 5.50. Deterministic (training and validation sets) and stochastic (test set) predicted stiffness parameters versus actual HS parameters with stochastic test MAE and residuals ..213
Figure 5.51. Comparison of simulated results: APHs vs. conventional parameterization (GEOSLOPE International Ltd., 2021). .....215
Figure 5.52. Comparison of simulated results: APHs vs. laboratory test results (Zhou, 2008). .....216
Figure 5.53. Comparative Performance Evaluation of LSTM and CNN Models. .....219
Figure 5.54. Permutation-based feature importance scores. .....220
Figure 5.55. Comparison of predicted vs the numerical FOS for the Fei Ngo Shan Landslide. .....221
Figure 5.56. Monte Carlo predictions and mean standard deviations for the August event. .222
Figure 5.57. Schematic illustration of different rainfall patterns investigated.....225
Figure 5.58. Rainfall parameters studied: intensity, duration, and recovery time. .....225
*Figure 5.59. Slope dimensions and boundary conditions*. .....226
Figure 5.60. FOS for different rainfall patterns. .....227
*Figure 5.61. FOS for different initial suctions versus non-rainfall duration.* .....228

Figure 5.62. FOS for different rainfall intensities versus non-rainfall duration. .....229
*Figure 5.63. FOS for different rainfall durations versus non-rainfall durations.* .....229
Figure 5.64. FOS for different hydraulic conductivities versus non-rainfall duration. .....230
Figure 5.65. Rainfall pattern with target event and antecedent moisture windows. .....232
Figure 5.66. Normalized slope-toe displacement across subsurface curvature and rainfall intensity–antecedent moisture scenarios. .....233
Figure 5.67. The perched GWL variation with time (Windows 1-5): a) for the diverging geometry (R/r = 1.4); b) for the converging geometry (R/r = 1.4). .....234
Figure 5.68. The model boundary condition and meshing for MPM. .....236
Figure 5.69. The temporal variation of the FOS: a) rainfall between 19 and 21 August 2005; b) LEA (Phi-B)-sensitivity analysis; c) LEA (Phi-B)-different rainfall.....237
Figure 5.70. The critical zone map: a) LEA-Phi-B; b) FEA. .....238
Figure 5.71. Post-failure simulation of the debris flows (t numerical = 2012 s) with the maximum displacement. .....242
Figure 5.72. Normalized average velocity versus travel horizontal distance. .....243
*Figure 6.1. Multi-event and multi-model operational forecasting tool* .....249
*Figure 6.2. Graphic user interface of the OWHK app*.....254
Figure 6.3. The APHs tool. .....259
*Figure 6.4. Permutation feature importance: Blue (Multi-model) vs. Grey (Single model).*.261
Figure 6.5. Comparison of simulated results: (APHs) vs. lab test results (Jin et al., 2016). .262
Figure 6.6. The automated slope stability tool.....267
Figure 6.7. The predicted FOS of the Yiu Hing landslide using the proposed models.....269
Figure 6.8. Single-event predictions for the Yiu Hing Road landslide between 6–8 September 2023. Illustrations adapted from PBS LearningMedia .....270

# LIST OF TABLES

Table 2.1. Summary of existing literature reviews. ..... 14

Table 2.2. A comparison of Wi-GIM, RTS, and GB-In-SAR (Mucchi et al., 2018)..... 18

Table 2.3. The pros and cons of each movement monitoring sensor. ..... 19

Table 2.4. Comparison among GPS, Extensometer, IN, TDR, and AE. ..... 20

Table 2.5. A comparison between IN, TDR, SOF, COFT, PSCFODS, and FBG-based inclinometer..... 20

Table 2.6. Literature-based studies regarding landslide monitoring data-related issues. ..... 28

Table 2.7. Literature multi-point monitoring field system..... 30

Table 2.8. Summary of the time series modeling main findings in the literature ..... 38

Table 2.9. Research gaps and suggested recommendations..... 51

Table 3.1. DL models design parameters and search space. ..... 65

Table 3.2. LR model design parameter and search space. ..... 66

Table 3.3. Comparison with available literature ..... 82

Table 3.4. LSTM hyperparameter search space ..... 88

Table 3.5. Multi-sensor model feature selection matrix ..... 93

Table 3.6. Multi-sensor model Hypertuning search space ..... 94

Table 3.7. APHs model design parameter search space ..... 106

Table 3.8. DL models design parameters and search space ..... 115

Table 4.1. Overview of key features and their justification ..... 120

Table 4.2. Overview of key features extracted from the DEM ..... 123

Table 4.3. Soil characteristics at Fei Ngo Shan (Maunsell Geotechnical Services Ltd., 2007). ..... 125

Table 4.4. Summary of direct shear test statistics. ..... 126

Table 5.1. The optimized model parameters. ..... 138

Table 5.2. The optimized spatiotemporal LR model parameters. .....138
Table 5.3. Alternative feature importance assessment: Feature's minimum vs maximum value. .....146
Table 5.4. Comparative assessment of the proposed data-driven models.....152
Table 5.5. The evaluation metrics for the VWC site-specific model: NMAE, R2, and MAE. .....159
Table 5.6. The evaluation metrics for the VWC operational model: NMAE, R2, and MAE.162
Table 5.7. The overall evaluation metrics for the operational vs. site-specific model.....165
Table 5.8. Statistical correlation among the input features and the output target at Pa Mei..170
Table 5.9. Pearson correlations for different rainfall antecedent periods at the Pa Mei. .....171
Table 5.10. Feature selection comparison. .....174
Table 5.11. Comparative evaluation metrics and benchmark against related literature.....182
Table 5.12. Main statistical characteristics of the dataset (Tung Chung). .....189
Table 5.13. Pearson coefficient among the input amended features and the output target. ...192
Table 5.14. The APHs final multi-LSTM model parameters. .....208
Table 5.15. The APHs model evaluation metrics: NMAE, MAPE, R2, and MAE. .....214
Table 5.16. APHs vs conventional parameterization (GEOSLOPE International Ltd., 2021). .....214
Table 5.17. HS model parameters with 99% CI lower and upper bounds using the APHs tool. .....217
Table 5.18. The optimized SSI model parameters. .....218
Table 5.19. Slope mechanical and hydrological characteristics. .....226
Table 5.20. The parametric study variables and results. .....232
Table 5.21. Material characteristics utilized in Anura3D. .....235
Table 6.1. Landslide prediction tool target improvement. .....250

Table 6.2. Summary of OWHK's current achievements and future direction. .....256

Table 6.3. Summary of the subsurface multi-sensor prediction model achievements. .....258

Table 6.4. The single model vs. multi-model evaluation metrics: MAE. .....262

Table 6.5. HS model parameters for Hostun sand with 99% CI using the APHs tool. .....263

Table 6.6. Summary of APHs current achievements and future direction. .....266

Table 6.7. Comparative assessment of data-driven models. .....268

Table 6.8. Current achievements, limitations, and future research directions. .....271

Table 7.1. Summary of cost-benefit and policy integration pathways. ..... 290

# LIST OF NOTATIONS AND ABBREVIATIONS

NOTATIONS

| Symbol | Description | SI Unit |
|---|---|---|
| $S_e$ | Effective degree of saturation | - |
| $K_x$ | Hydraulic conductivity in the x-directions | m/s |
| $K_y$ | Hydraulic conductivity in the y-directions | m/s |
| $d\sigma$ | Stress increase vector | kPa |
| $d\varepsilon$ | Strain increment vector | - |
| $K_i$ | Intrinsic permeability | m² |
| $V_L$ | Liquid phase volume | m³ |
| $V_S$ | Solid phase volume | m³ |
| $v_L$ | Liquid velocity vector | m/s |
| $v_S$ | Solid velocity vector | m/s |
| $K_e$ | Liquid elastic bulk modulus | kPa |
| $m_v$ | Water retention curve slope | $kPa^{-1}$ |
| $a_S$ | Solid acceleration | m/s² |
| $a_L$ | Liquid acceleration | m/s² |
| $\tau$ | Shear strength | kPa |
| $\gamma_w$ | The unit weight of water | kN/m³ |
| $\gamma_t$ | Wet unit weight | kN/m³ |
| $C$ | Cohesion | kPa |
| $C'$ | Effective cohesion | kPa |
| $\phi$ | Internal friction angle | (°) |
| $n$ | Porosity | - |
| $\phi'$ | Effective internal friction angle | (°) |
| $\phi^b$ | Internal friction angle corresponding to ψ | (°) |
| $\psi$ | Matric suction | kPa |
| $v$ | infiltration velocity | m/s |
| $u_a$ | Atmospheric pressure | kPa |
| $u_w$ | Pour water pressure | kPa |
| $h_w$ | Water head | m |
| $A_{in}$ | Inflow radial area | m² |
| $A_{out}$ | Outflow radial area | m² |
| $t'$ | Normalized time | - |
| $T$ | Field observed time | s |
| $t_{numerical}$ | Numerically simulated time | s |
| $V'$ | Normalized velocity | - |
| $r$ | Local radius | m |
| $R_f$ | Failure ratio | - |

| Symbol | Description | SI Unit |
|---|---|---|
| $k$ | Unsaturated hydraulic conductivity | m/s |
| $k_{sat}$ | Saturated hydraulic conductivity | m/s |
| $E$ | Modulus of elasticity | kPa |
| $\upsilon$ | Poisson's ratio | - |
| $I$ | Identity matrix | - |
| $b$ | Body force vector | kN/m³ |
| $\rho_m$ | Mixture density | kg/m³ |
| $\mu_L$ | Liquid dynamic viscosity | Pa·s |
| $F$ | Nodal load | kN |
| $d$ | Nodal deformation | m |
| $Q$ | Nodal flow | m³/s |
| $u$ | Nodal pore water pressure | kPa |
| $D$ | Rainfall duration | d or h |
| $I$ | Rainfall intensity | mm/d |
| $E_N$ | Rainfall event | - |
| $API_N$ | Antecedent precipitation index | mm |
| $S$ | Displacement | m |
| $q$ | Surface flux | m/s |
| $Q$ | Discharge | m³/s |
| $\sigma$ | Normal stress | kPa |
| $\sigma_s$ | Suction stress | kPa |
| $\sigma_{net}$ | Net normal strength | kPa |
| $\sigma_n$ | Normal strength | kPa |
| $g$ | Gravitational acceleration | m/s² |
| $z_\psi$ | Matrix suction head | m |
| $W_t$ | Weight of soil | kN |
| $H$ | Thickness above the bedrock layer | m |
| $z$ | Unsaturated thickness | m |
| $h$ | Saturated thickness | m |
| $\beta$ | Slope angle | (°) |
| $\theta_w$ | Volumetric water content (VWC) | m³/m³ |
| $\theta_r$ | Residual VWC | m³/m³ |
| $\theta_s$ | Saturated VWC | m³/m³ |
| $\beta$ | The slope of the critical state line | (°) |
| $P_o$ | Reference pressure | kPa |
| $R$ | Total radius | m |

| | | | | | |
|---|---|---|---|---|---|
| $E_{50}^{ref}$ | Secant stiffness in drained triaxial tests | kPa | $\lambda$ | Fitting parameter | - |
| $E_{ur}^{ref}$ | Unloading/reloading stiffness | kPa | *a, n, m* | Fredlund – Xing fitting parameters | a: kPa; n,m: - |
| $m$ | Power for the stress level dependency of stiffness | - | $\varepsilon_1$ | Axial strain | - |
| $q$ | Deviatoric stress | kPa | $\varepsilon_v$ | Volumetric strain | - |
| $Z'$ | Vertical sliding depth | m | $\sigma_3$ | Lateral stress | kPa |
| $H'$ | Vertical height from slope toe to crest | m | $R^2$ | Coefficient of determination | - |

ABBREVIATIONS

| | | | |
|---|---|---|---|
| *AE* | Acoustic Emission | *LEA* | Limit Equilibrium Analysis |
| *AI* | Artificial Intelligence | *LHS* | Latin Hypercube Sampling |
| *ANN* | Artificial neural networks | *LiDAR* | Light detection and ranging |
| *APHs* | Automatic Parametrization of Hardening Soil (HS) | *LR* | Logistic regression |
| *AUC* | The area under the ROC curve | *LSTM* | Long short-term memory |
| *BA* | Bat algorithm | *LURR* | Loading-Unloading-Response Ratio |
| *BGRU* | Bayesian gated recurrent unit network | *M/HDT* | Moderate to Highly Decomposed Tuff |
| *BLSTM* | Bayesian long short-term memory network | *MAE* | Mean Absolute Error |
| *BOTDR* | Brillouin optical time domain reflectometry | *MAPE* | Mean Absolute Percentage Error |
| *BPNN* | Backpropagation neural networks | *MC* | Monte Carlo |
| *CD* | Consolidated Drained | *MKGM* | Multi-kernel grey model |
| *CDG* | Completely Decomposed Granite | *MLR* | Multilinear regression |
| *CDT* | Completely Decomposed Tuff | *MPM* | Material Point Method |
| *CDV* | Completely Decomposed Volcanic | *MSE* | Mean Squared Error |
| *CEED* | Civil Engineering and Development Department | *MSI* | Multiple Swarm intelligence |
| *CEEMD* | Complete ensemble empirical mode decomposition | *NMAE* | Normalized Mean Absolute Error |
| *CI* | Confidence Interval | *PSCFODS* | Parallel-series connected fiber-optic displacement sensor. |
| *CNN* | Convolutional neural network | *PS-InSAR* | Persistent scatterer interferometry |
| *COFT* | Combined optical fiber transducer | *PSL* | Production storage level |
| *CREST* | Coupled Routing and Excess STorage | *PWP* | Pore Water Pressure |
| *CSL* | Critical State Line | *RF* | Random forest model |
| *DA* | Dragonfly optimization algorithm | *RNN* | Recurrent neural networks |
| *DES* | Double exponential smoothing | *RS* | Random Search |
| *DInSAR* | Differential (SAR) interferometry | *RTS* | Robotized total station |
| *DL* | Deep Learning | *SAR* | Synthetic aperture radar |
| *DMA* | Double moving average | *SAs* | Smart aggregates |
| *EEMD* | Ensemble empirical mode decomposition | *SDSs* | Soil deformation sensors |
| *ELM* | Extreme learning machine | *SDT* | Slightly Decomposed Tuff |
| *EMD* | Empirical mode decomposition | *SHETRAN* | Système Hydrologique Européen TRANsport |
| *ENN* | Evolutionary neural network | *SLIDE* | SLope-Infiltration-Distributed Equilibrium |
| *ERT* | Electrical resistivity tomography | *SLIP* | Shallow Landslides Instability Prediction |
| *ESP* | Effective Stress Path | *SOF* | Sensing optical fiber |
| *EWS* | Early warning systems | *SSA* | Sparrow search algorithm |
| *FBG* | Fiber Bragg Grating | *SSI* | Soil saturation index |

| | |
|---|---|
| *FEA* | Finite Element Analysis |
| *FOS* | Factor of Safety |
| *FPR* | False positive rate |
| *GA* | Green-Ampt model |
| *GB-InSAR* | Ground-based SAR |
| *GCO* | Geotechnical Control Office |
| *GEO* | Geotechnical Engineering Office |
| *GI* | Ground Investigation |
| *GInfo* | Geotechnical Information Infrastructure |
| *GIS* | Global information system |
| *GNSS* | Global navigation satellite system |
| *GOA* | Grasshopper optimization algorithm |
| *GPS* | Global positioning system |
| *GRU* | Gated recurrent unit |
| *GWL* | Groundwater level |
| *GWO* | Grey wolf optimization |
| *HCF* | Hydraulic Conductivity Function |
| *HIRESSS* | High-Resolution Slope Stability Simulator |
| *HPC* | High-Performance Computing platform |
| *IETDs* | Inter-event time definitions |
| *IN* | Inclinometers |
| *InSAR* | Interferometric synthetic aperture radar |
| *KGM* | Kernel-based grey model |
| *Std. Dev* | Standard Deviation |
| *SVC* | Support vector classifier |
| *SVM* | Support vector mechanism |
| *SVR* | Support vector regression |
| *SWCC* | Soil Water Characteristic Curve |
| *SWI* | Soil wetness index |
| *TDR* | Time Domain Reflectometry |
| *TENG* | Triboelectric nanogenerators |
| *TOPMODEL* | TOPography-based hydrological MODEL |
| *TP* | Trial Pit |
| *TRIGRS* | Transient Rainfall Infiltration and Grid-based Slope Stability |
| *TTEFBS* | Timbo-like triboelectric force and bend sensor |
| *TWI* | Topographic Witness Index |
| *UHF RFID* | Ultrahigh-frequency radio-frequency identification |
| *USDA* | United States Department of Agriculture |
| *VMD* | Variational mode decomposition |
| *WHO* | World Health Organization |
| *WI* | Wetness increase |
| *Wi-GIM* | Wireless Sensor Network for Ground Instability Monitoring |
| *WMKGM* | Weighted multi-kernel grey model |
| *WOA* | Whale optimization algorithm |
| *WSN* | Wireless Sensor Network |

# CHAPTER 1: INTRODUCTION

This chapter serves as the foundational component of the thesis, establishing the research landscape, defining the scope of the investigation, and outlining the strategic approach for the study. It begins by presenting the broad context and specific problem of rainfall-induced landslide forecasting, using Hong Kong as a critical case study. The chapter then systematically articulates the core Problem Statement, identifying the critical efficacy-efficiency trade-off in existing early warning systems. Following this, specific Research Gaps are detailed to pinpoint the precise limitations in current literature and practice. The chapter subsequently presents the study's Objectives, which are designed to directly address these gaps, and provides a brief overview of the overarching Methodology and the novel three-tiered framework developed to achieve these aims. Finally, it concludes with a summary of the Thesis Organization, providing a roadmap for the reader by outlining the content and sequence of the remaining chapters.

## 1.1. INTRODUCTION

Landslides represent a pervasive and destructive global geological hazard, defined as the downward movement of slope-forming materials when shear stresses exceed the available shear strength of the ground. These events can be triggered by various mechanisms, including earthquakes, volcanic activity, and rainfall, as illustrated in Figure 1.1. Among these, rainfall-induced landslides are particularly consequential in regions experiencing intense precipitation events, which are becoming more frequent and severe due to climate change (Wu et al., 2020). Hong Kong stands as a paramount case study for this specific hazard. Its unique topographic and climatic profile, a mountainous city with over 60% of its land area comprising natural slopes (Geotechnical Engineering Office, 2016), subjected to an average of 2398 millimeters of annual rainfall from typhoons and rainstorms (HKO, 2019), creates a high-risk environment that manifests in over 300 reported landslides annually. The catastrophic potential is grimly exemplified by historical events such as the 1972 Po Shan Road landslide (Figure 1.2), which

obliterated apartment blocks and claimed 156 lives (HKO, 2015). Over the past half-century, landslides have been responsible for approximately 500 fatalities in Hong Kong, mirroring a devastating global trend that saw nearly 60,000 deaths from non-seismic landslides between 2004 and 2016 (Froude & Petley, 2018), as visualized in Figure 1.3. This enduring threat, despite extensive mitigation efforts, underscores the critical and continuous need for advanced research into rainfall-induced landslide mechanisms (Song et al., 2017).

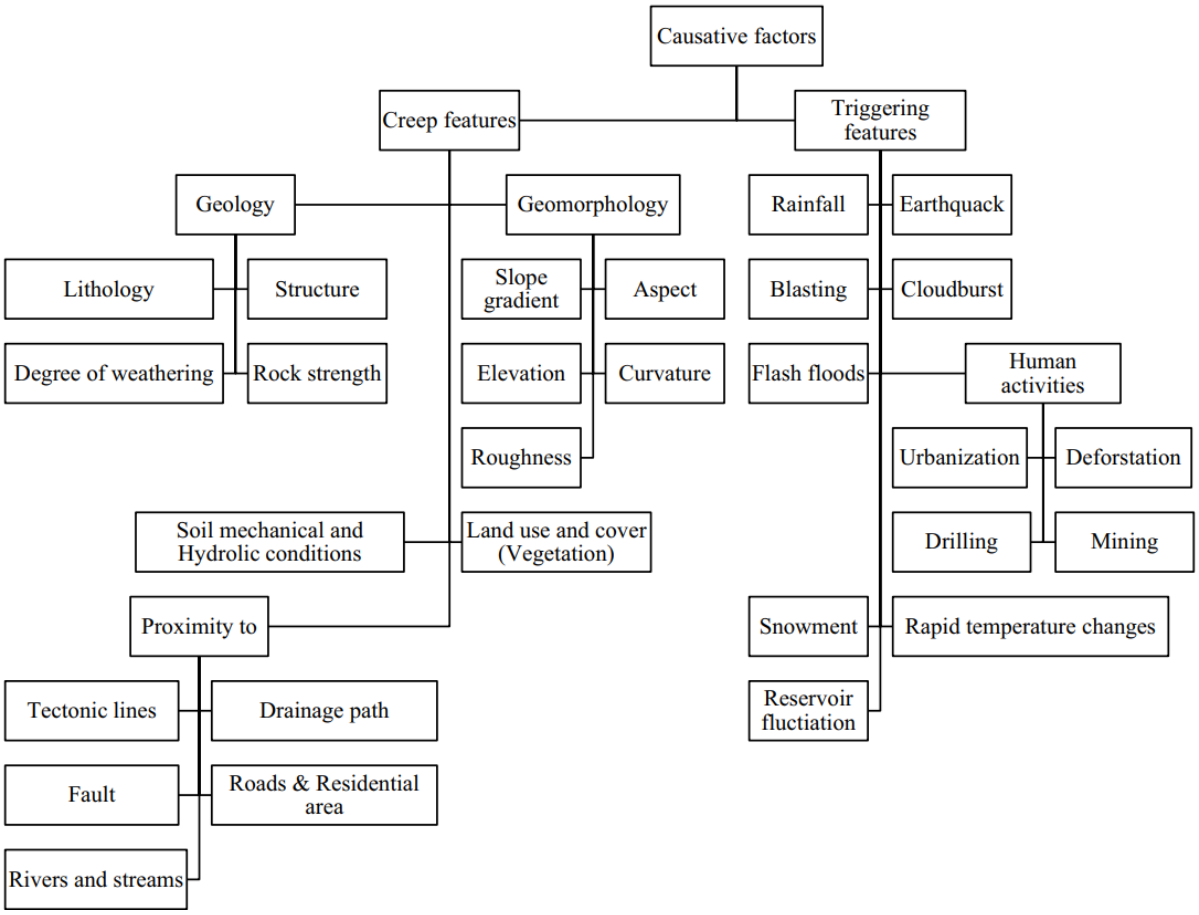



*Figure 1.1. Landslide causative factors*



*Figure 1.2. Po Shan Road's 1972 collapse (Cheng et al., 2009)(Photo by Geotechnical Control Office, Hong Kong Government)*

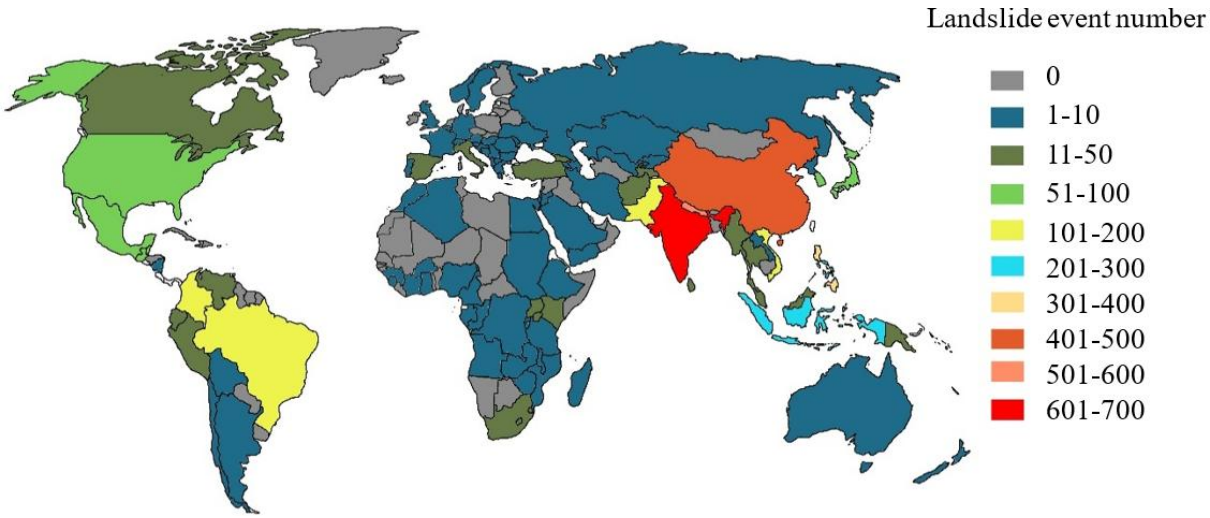



*Figure 1.3. Fatal non-earthquake landslides from 2004 to 2016 (Froude & Petley, 2018)*

The diversity of landslide phenomena is captured by detailed classification systems, such as the one by Hungr et al. (2014), which categorizes 32 distinct types based on material composition and movement mechanisms. While this provides a valuable framework, this dissertation narrows its focus to shallow, rainfall-induced landslides (Figure 1.4), which are

among the most frequent and destructive types in subtropical regions like Hong Kong (Formetta & Capparelli, 2019). These failures typically involve a relatively thin mantle of soil and colluvium, often 3 to 5 meters thick (Zhang et al., 2011), and their initiation is governed by complex hydro-mechanical processes within the vadose zone. A naive approach to risk mitigation involves the widespread reinforcement of all slopes using engineering structures like soil nails and retaining walls (Huang & He, 2023). However, this is economically and logistically impractical given the vast number of slopes. A more sophisticated and efficient paradigm advocates for a proactive cycle of monitoring, data analysis, and prediction to enable targeted and timely intervention.

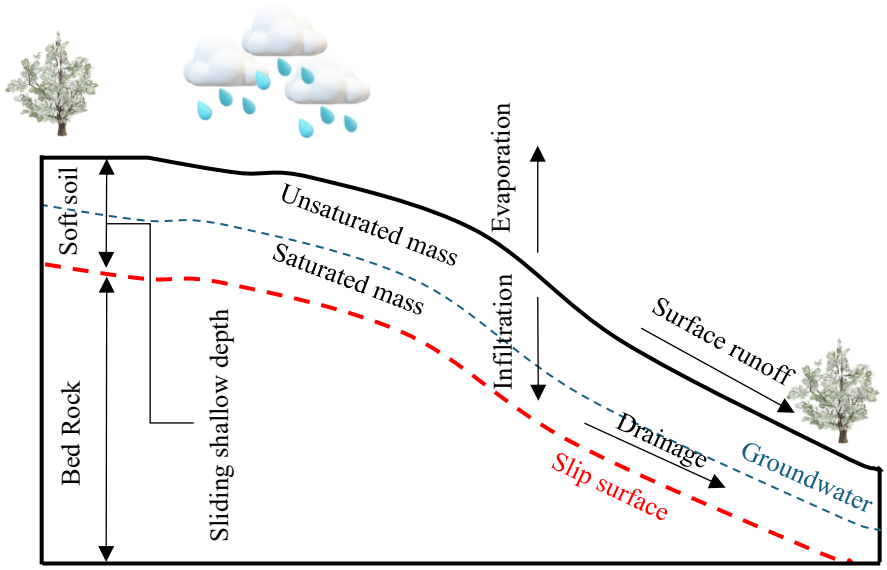



*Figure 1.4. Schematic representation of shallow rainfall-induced landslides.*

The successful implementation of this paradigm is hampered by the inherent complexity of slope systems. Slope engineering is an unpredictable, non-linear, and dynamic domain where stability is influenced by a multitude of geological and environmental factors (Dai et al., 2021), many of which exhibit significant randomness and uncertainty (Wu et al., 2021). In this context, Artificial Intelligence (AI) and, more specifically, Deep Learning (DL) have emerged as transformative tools. DL algorithms are capable of constructing complex, non-linear mappings between input data and outputs, making them exceptionally suited to model

the intricate behaviors of geotechnical systems (Ma et al., 2020; Phoon & Zhang, 2023). The strengths of DL, including its versatility, its ability to autonomously extract salient features from raw data, and its independence from rigid prior assumptions (Zhang et al., 2022), have led to its successful application across geotechnical engineering, from landslide susceptibility mapping and displacement prediction to slope stability analysis (Huang et al., 2023; Azmoon et al., 2021; Song et al., 2019; Kumar et al., 2020a; Chiang et al., 2022; Huang et al., 2023; Dai et al., 2022; Filipović et al., 2022).

Physically, rainfall-induced landslides constitute a time-dependent problem where slope stability deteriorates during rainfall infiltration and recovers during subsequent drainage periods. The hydrological process is governed fundamentally by the soil's hydraulic conductivity, which controls the rate at which infiltrating rainwater moves through the unsaturated zone. Soils with higher hydraulic conductivity permit rapid infiltration and deeper wetting front propagation, while lower conductivity soils promote surface runoff and localized pore-pressure buildup near the surface. This process is further modulated by surface flux conditions and storage capacity, where infiltrating rainwater progressively fills soil pores and displaces air (Zhang et al., 2011). The loss of matric suction upon wetting is a key destabilizing mechanism in unsaturated soils (Valentino et al., 2014), and failure can occur due to localized water accumulation even before full saturation is reached (Gerscovich et al., 2006). The slope's recovery, or "healing," is a function of the non-rainfall duration (the antecedent period), with longer dry periods allowing for greater drainage and strength recovery (Bontemps et al., 2020). While advanced numerical methods, including coupled hydro-mechanical analyses and the Material Point Method (MPM), can simulate these physical processes, they are often hamstrung by extensive parameter requirements and computational demands, rendering them unsuitable for real-time early warning. Consequently, there is a pressing need for a novel framework that

leverages the predictive power of data-driven models, capable of providing accurate, timely, and operational warnings to safeguard communities and infrastructure.

## 1.2. PROBLEM STATEMENT

The fundamental challenge in landslide risk reduction lies in accurately and reliably answering two critical questions: where will a landslide initiate, and when will it occur? While substantial research has been dedicated to this field, a definitive and operationally viable solution remains elusive. Contemporary approaches are bifurcated into two problematic camps. On one end of the spectrum are simplistic empirical models, such as Intensity-Duration (I-D) thresholds. While computationally efficient and easy to implement regionally, these models are notoriously prone to high false-alarm rates. This is primarily because they grossly oversimplify reality by neglecting spatial variability (e.g., differences in soil properties, slope geometry, and land cover) and the intrinsic randomness of real rainfall patterns, including the crucial role of antecedent moisture conditions (Hong et al., 2018). On the opposite end are sophisticated physically-based deterministic models. These models, which can simulate internal processes like pore-water pressure generation and suction loss with high fidelity, are computationally intensive and require extensive, often unavailable, geotechnical parameters. This makes them prohibitively slow and costly for real-time, regional-scale early warning applications.

This fundamental trade-off is exacerbated by several persistent and interconnected operational gaps, which this thesis directly confronts: **a)** Many existing systems and studies prioritize surface manifestations, such as displacement, while overlooking the critical internal hydro-mechanical precursors that are the true harbingers of failure. The monitoring and prediction of key subsurface variables, such as Volumetric Water Content (VWC), which is a direct indicator of infiltration and a precursor to failure (Tohari et al., 2007; Suk et al., 2022); matric suction, a key component of shear strength in unsaturated soils (Davar et al., 2022); and earth pressure, which reveals the kinematic forces within the slope (Chuan et al., 2013), are not

adequately integrated into operational forecasting frameworks; **b)** Rainfall alone is an insufficient predictor. Studies have shown a low direct correlation between rainfall and landslide movement (Hemalatha et al., 2019) because the infiltration process is modulated by a host of unmonitored factors like evapotranspiration, surface runoff, and soil characteristics (Granata et al., 2023). Furthermore, the subsurface response to rainfall is not instantaneous; it exhibits complex temporal lags and spatial variability due to the antecedent moisture state, soil hydraulic properties, and slope geometry (Uwihirwe et al., 2020; Sasahara, 2017). This is vividly demonstrated in field data from Tung Chung (Figure 1.5), where sensors for VWC, suction, and earth pressure at different locations and depths show distinctly different and lagged responses to the same rainfall event; **c)** The reliance on wireless sensor networks for monitoring introduces critical points of failure. These networks are susceptible to power depletion and communication breakdowns precisely during the extreme weather events they are meant to monitor, leading to data loss (Kumar et al., 2020b; Mucchi et al., 2018). Most proposed solutions to this problem come from a computer science perspective, neglecting the geotechnical implications of losing critical physical data; and **d)** The workflow from data collection to an actionable stability assessment is fragmented and manual. Key steps, particularly geotechnical parameterization and Factor of Safety (FOS) computation, lack automation, creating a delay that undermines the "early" aspect of early warning.

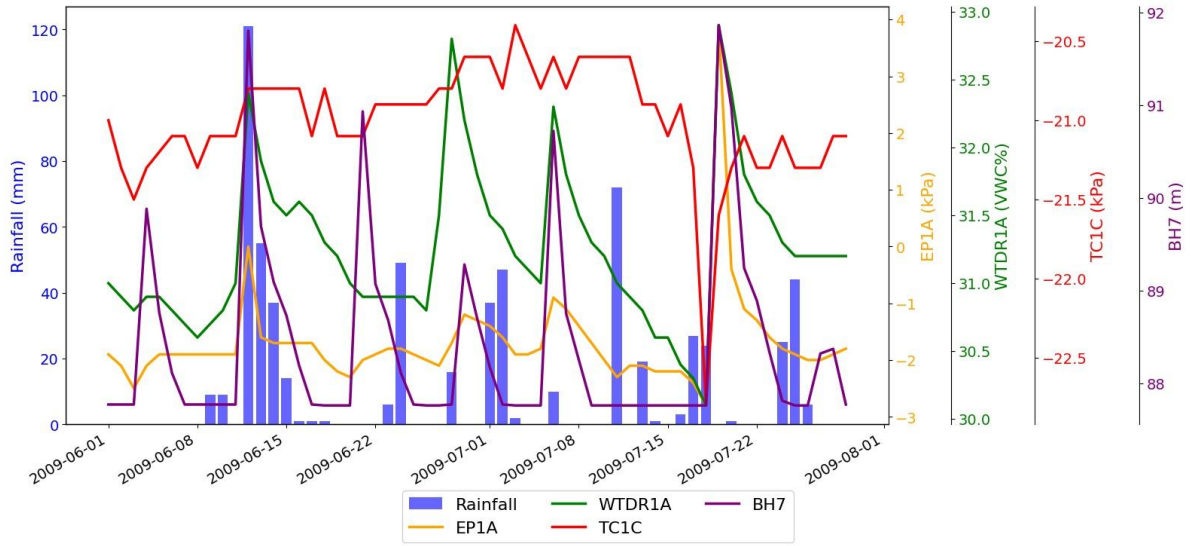



*Figure 1.5. Sample landslide subsurface responses: VWC (WTDR1A), suction (TC1C), earth pressure (EP1A), and groundwater (BH7) (Geotechnical Engineering Office 2010a-c).*

In summary, the central problem is the absence of an integrated framework that seamlessly bridges data-driven forecasting with a foundational understanding of landslide physics. Therefore, a critical need exists for an intelligent system that not only handles spatiotemporal complexities and operational vulnerabilities like data loss but also automates the workflow from monitoring to stability assessment. Such a system should be grounded in a comprehensive review of mechanisms and methods to achieve both high accuracy and operational practicality for landslide early warning.

## 1.3. RESEARCH GAPS

Despite advances in the available literature, this thesis directly addresses the following research gaps, which are derived from the outlined problem statement: **a)** Lack of a comprehensive review of subsurface monitoring techniques and prediction models for rainfall-induced landslides; **b)** Limited examination of rainfall's random effects, as models often reduce its complexity to cumulative totals, ignoring spatiotemporal patterns and antecedent conditions; **c)** Over-reliance on surface data (e.g., displacement) in many models, neglecting critical internal hydro-mechanical processes; **d)** Neglect of geotechnical perspectives in data-loss solutions, with most studies focusing solely on computer science aspects; **e)** Use of

oversimplified analytical methods like regression, which fail to capture the dynamic, non-linear behavior of slope systems; **f)** Rare implementation of automated, end-to-end workflows for geotechnical parameterization and stability analysis.

## 1.4. RESEARCH OBJECTIVES

To bridge the identified research gaps and resolve the critical efficacy-efficiency trade-off, this thesis is guided by a primary aim: the development of a multi-tiered AI-enhanced approach for operational landslide prediction and stability assessment. This overarching goal is operationalized through the following four specific objectives:

- Understand and study factors and mechanisms governing landslide initiation, and critically review and evaluate existing monitoring systems and predictive methodologies.
- Design a spatiotemporal model for shallow, rainfall-induced landslides using historical records and data-driven techniques to capture triggering randomness.
- Establish an integrated prediction framework that fuses landslide subsurface hydrological and mechanical responses to improve early-warning accuracy.
- Streamline and automate the stability-analysis workflow, including parametrization and FOS computation.

## 1.5. RESEARCH METHODOLOGY

To overcome the critical limitations of existing landslide prediction models, this study pioneers a three-tiered framework that strategically integrates data-driven and deterministic approaches. This hierarchical methodology is designed to intelligently allocate computational resources, ensuring both regional scalability and site-specific geotechnical accuracy. The framework, illustrated in Figure 1.6, operates as follows: The first tier leverages incident records and diverse geospatial data to formulate a preliminary decision model using multi-data-driven models (Long short-term memory (LSTM), gated recurrent unit (GRU), and 1D convolutional neural network (CNN), were employed alongside a logistic regression (LR) benchmark, with

forecasting probabilities compared to a deterministic seepage stability model). This model classifies forecasting into three categories: Failure not expected: low-risk zones; Warning signs: forecasting of potential instability is observable; and Failure expected: high-confidence predictions of landslide occurrence. This preliminary decision process ensures that cases with clear high or low risk are identified swiftly. However, due to uncertainties inherent in heterogeneous datasets, some cases will fall into the “warning” category, necessitating a more refined analysis.

For warning signs (ambiguous cases), a secondary analytical layer (Tier 2) that integrates subsurface monitoring techniques is introduced. Although monitoring is critical for tracking landslide processes and predicting disasters, it is often hampered by challenges such as data collection limitations, inherent randomness, and triggering uncertainties. This study leverages DL (novel multi-parallel LSTM networks with independence-oriented time-series windowing) to enhance the characterization of subsurface responses, effectively capturing spatiotemporal dependencies while addressing both triggering uncertainties and monitoring inaccuracies. The proposed model enhances early warning by providing real-time, dynamic updates for subsurface forecasts. These forecasts subsequently feed into a deterministic analysis during the third stage of focused evaluation, further refining predictive reliability and establishing a robust framework for hazard assessment. A key innovation is its multi-node prediction framework, which is explicitly designed to mitigate data loss from field sensors, ensuring robust operational forecasting.

The final tier directly addresses the primary bottleneck of deterministic models, their extensive input data requirements, by introducing a multi-faceted approach to automated stability evaluation. The cornerstone is the Automated Parameterization Tool (APHs), which employs LSTM networks to derive advanced soil parameters from standard laboratory tests with high accuracy, effectively overcoming the critical parameter acquisition hurdle. This

enables two distinct analytical pathways: First, a rapid hybrid approach that couples data-driven saturation forecasting (Soil Saturation Index, SSI) with infinite-slope stability analysis (FOS) for efficient, regional-scale early-warning. Second, for scenarios requiring high-fidelity assessment, a comprehensive deformable analysis that utilizes the automatically derived parameters for coupled hydro-mechanical modeling to simulate slope deformation and estimate debris volume and travel distance. By seamlessly embedding the subsurface forecasts from Tier 2 into these calibrated physical models, this tier creates a robust, closed-loop workflow that integrates the swift responsiveness of data-driven methods with the precision of geotechnical mechanics, thereby significantly mitigating uncertainties and enhancing predictive reliability for actionable risk assessment.

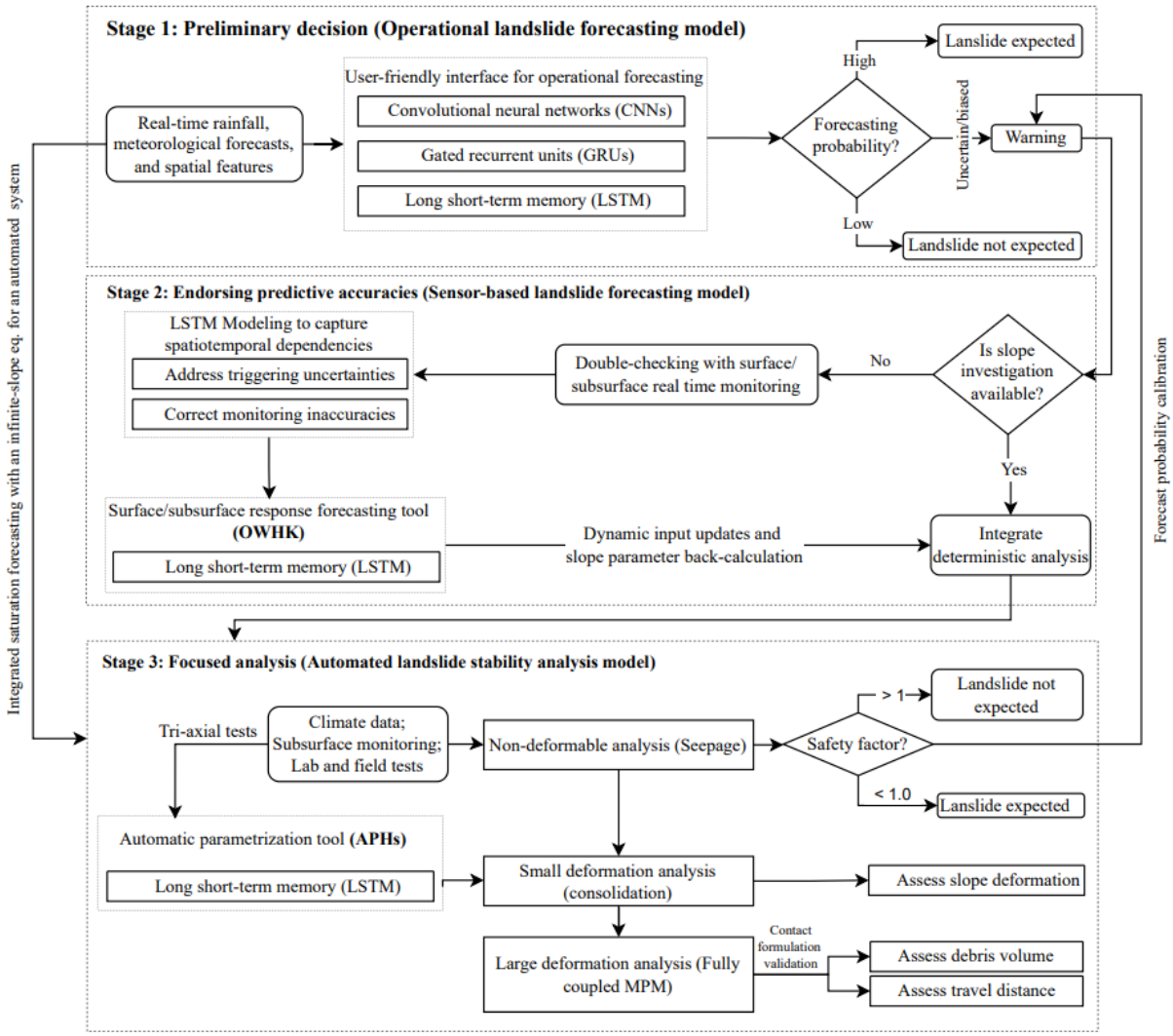



*Figure 1.6. Research methodology workflow covering all study objectives.*

## 1.6. THESIS ORGANIZATION

This thesis is structured to systematically address the research objectives. **Chapter 2** provides a comprehensive literature review, critically evaluating existing monitoring and prediction methodologies to formally identify the research gaps that this study aims to fill. **Chapter 3** delineates the overall research methodology, detailing the novel three-tiered framework and the specific analytical techniques developed to bridge these gaps. **Chapter 4** describes the data collection framework, including the instrumentation, curated historical datasets, and the specific case study sites in Hong Kong that form the basis of the analysis. **Chapter 5** presents the core analysis and results, validating the performance of each tier of the proposed framework against geotechnical benchmarks and historical events. **Chapter 6** translates these research findings into practical output, presenting the operational web-based applications and automated tools developed for landslide forecasting. Finally, **Chapter 7** concludes the thesis by synthesizing the key findings, underscoring the theoretical, methodological, and practical significance of the work, and providing recommendations for future research.

# CHAPTER 2: LITERATURE REVIEW

## 2.1. INTRODUCTION

This chapter presents a comprehensive systematic review of previous research related to landslides. This literature review proceeds as follows: First, a comprehensive analysis of subsurface monitoring systems is presented, covering displacement sensors, force and stress gauges, groundwater instrumentation, temperature probes, and early warning indicators. Next, the geotechnical and hydrological theory behind rainfall-induced landslides is outlined. Current prediction methodologies are then surveyed, including deterministic physical models, physically causative frameworks, and empirical-statistical thresholds, emphasizing their practical applications. Two stages of examination are undertaken: a macroscopic overview tracing each method's evolution in relation to technology and boundary conditions, followed by a microscopic deep dive into individual system features. The review concludes by identifying key research gaps, offering targeted recommendations, and presenting a graphical chart that synthesizes the reviewed literature.

## 2.2 SEARCH STRATEGY AND FRAMEWORK

The existing literature offers extensive reviews, as summarized in Table 2.1. However, it is revealed that many reviews focus on single methodologies; scientometric analysis is infrequently applied; and the knowledge remains fragmented across different approaches. Consequently, a comprehensive review synthesizing this fragmented knowledge into a unified framework is needed. With that in mind, this study adopts quantitative (scientometric) and qualitative (systematic) methods to investigate the available literature. The techniques above assist researchers in improving systematic review reporting (Wuni & Shen, 2020). Figure 2.1 illustrates the methodical flow for a systematic review, which consists primarily of three processes: identification, screening, and eligibility. Regarding scientometric analysis, Figure 2.2 demonstrates the steps of the scientometric analysis, which mainly consist of gathering

bibliometric data, exporting bibliometric data to an appropriate program (VOS software (van Eck & Waltman, 2010)), evaluating data, and discussing results. In the identification process, keywords, search databases, and inclusion and exclusion criteria were utilized to filter the papers acquired. This study focuses on Scopus and Web of Science as preliminary search database sources for landslide prediction. Keywords are identified after selecting a search database, such as (“landslide prediction” ), (“landslide monitoring”), and ("landslide or landslides" and "machine learning or artificial intelligence or deep learning" ) to consider several possible prediction and monitoring techniques.

*Table 2.1. Summary of existing literature reviews.*

| Study | Approach | Content |
|---|---|---|
| (Angeli et al., 2000) | Systematic | Address the management, challenges, and remedies of diverse systems. |
| (Shamshi, 2004) | Systematic | A brief overview of landslide monitoring equipment was provided. |
| (Zhang et al., 2011) | Systematic | An examination of the geotechnical and hydrological factors influencing rainfall-triggered landslides. |
| (De Graff, 2011) | Systematic | Explaining the process of acquiring and constructing an improved monitoring system with an emphasis on rating its effectiveness |
| (Eyo et al., 2014) | Systematic | Utilizing inexpensive GPS technology in various applications. |
| (Chae et al., 2017) | Systematic | Assessing landslide susceptibility, modeling runout scenarios, monitoring landslide activity, and implementing early warning systems. |
| (Segoni et al., 2018) | Systematic | Establishing thresholds for rainfall-induced events. |
| (Merghadi et al., 2020) | Systematic | Utilizing machine learning algorithms to assess landslide susceptibility. |
| (Shano et al., 2020) | Systematic | Various forecasting methods center around statistical models. |
| (Yanbin et al., 2022) | Systematic | Utilization of machine learning models for landslide susceptibility |
| (Zou & Zheng, 2022) | Scientometric | Conducting scientometric analysis, exploring constrained physical prediction methods, and examining case studies. |
| (Huang et al., 2022a) | Bibliometric | Assessing landslide susceptibility using Geographic Information Systems. |
| (Petrucci, 2022) | Systematic | Identifying the main factors contributing to landslide-related fatalities. |
| (Lapenna & Perrone, 2022) | Systematic | Explore the applications of time-lapse electrical resistivity tomography. |
| (Breglio et al., 2023) | Systematic | Utilizing photonic technology to monitor deformation in slopes and tunnels and temperature and soil humidity in agricultural soil. |
| (Huang et al., 2023) | Systematic | Continuous monitoring through Global Navigation Satellite Systems. |
| (Auflič et al., 2023) | Bibliometric | Evaluation of landslide monitoring methods through questionnaire analysis. |
| This study | Systematic & scientometric | Deterministic and susceptibility models for landslide prediction. |
| | | Thresholds are derived from empirical data and physical principles. |
| | | An overview of methods for monitoring subsurface conditions. |
| | | Exploring time series models for landslide prediction. |

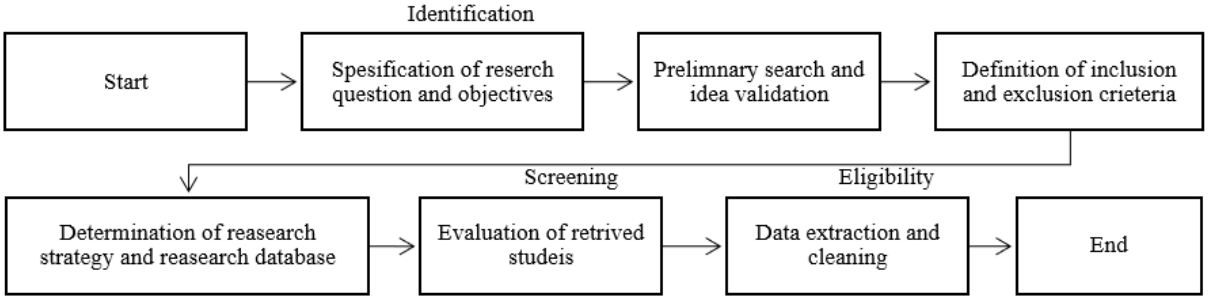



*Figure 2.1. Methodological flow for systematic review.*

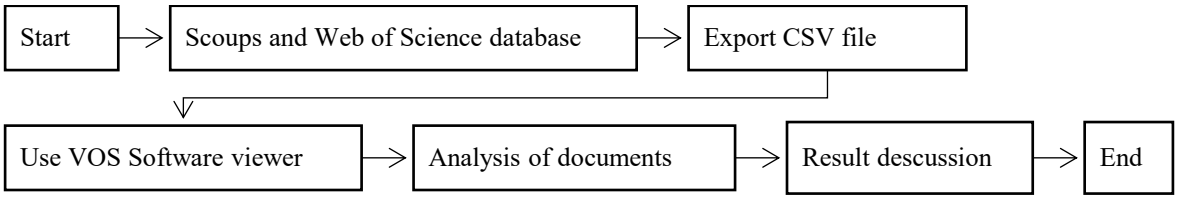



*Figure 2.2. Methodological flow for scientometric analysis.*

Inclusion and exclusion criteria were applied to filter the results and prioritize the most relevant studies to the research question. This research used the following criteria: (1) research focusing on the aforementioned keywords, (2) studies published after 2000, (3) articles published in peer-reviewed journals, and (4) studies published as articles and review submissions. The exclusion criteria were as follows: (1) papers published in a language other than English, (2) studies with no full text accessible, (3) manuscripts published in a subject area other than engineering, and (4) publications published in a source type other than a journal. The (PRISMA) procedure (Moher et al., 2009) was then used to evaluate and assess the collected articles (Figure 2.3). The backward and forward snowballing approach was then used to locate additional studies that could not be found using Scopus, Web of Science, and Google Scholar searches (Wohlin, 2014). The scientometric analysis began after performing a screening process to analyze the connections among authors, keywords, publications, and countries within certain research areas. The scientometric analysis was conducted with the open-source VOSviewer software application version 1.6.20 (van Eck & Waltman, 2010). This scientometric review was used to provide citation and co-authorship analyses of nations, organizations, authors, and keywords involved in the study topic. The primary goal of

scientometric analysis is to guarantee that the findings are relevant enough to be included in a systematic review (see Figure A1).

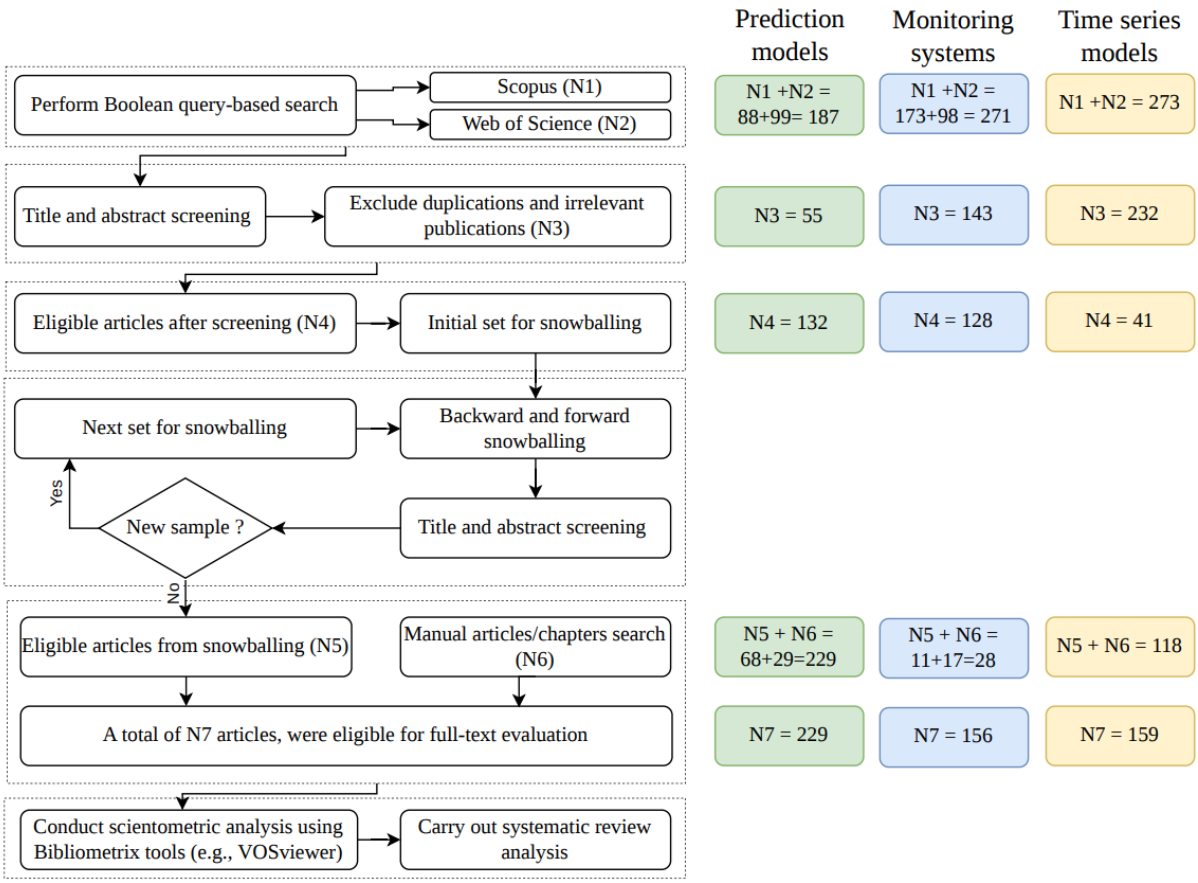



*Figure 2.3. PRISMA: Screening and selection flow diagram.*

## 2.3. LANDSLIDE MONITORING TECHNIQUES

### *2.3.1. Surface VS Subsurface Monitoring Techniques*

Surface and subsurface techniques have advantages and disadvantages, highlighting the need to utilize these systems best. For instance, surface monitoring (Figure 2.4) systems can provide information about ground movement (in millimeter-scale) and near-surface soil moisture with good spatial resolution and are helpful for wide-area surveillance. However, these systems have some limitations: a) it is challenging to collect data in real-time (Rosi et al., 2011), b) they are affected by dense vegetation and shadows, and c) they are highly affected by weather conditions such as fog or snow (Mucchi et al., 2018). Additionally, these systems cannot detect the subsurface mechanism of the slope (Auflic et al., 2023). To illustrate, slope instability arose

from the inner subsurface area, which surface systems could not detect. The slope surface can deform macroscopically only when the slope mass changes sufficiently (Deng et al., 2021). As a result, surface techniques are suitable for national-scale studies (i.e., hundreds to thousands of $km^2$ (Oguz et al., 2022). These methodologies are appropriate for creating landslide susceptibility maps. However, such maps cannot provide early warning indications or predict disasters. These objectives can only be met by knowing the inner mechanism and monitoring subsurface conditions.

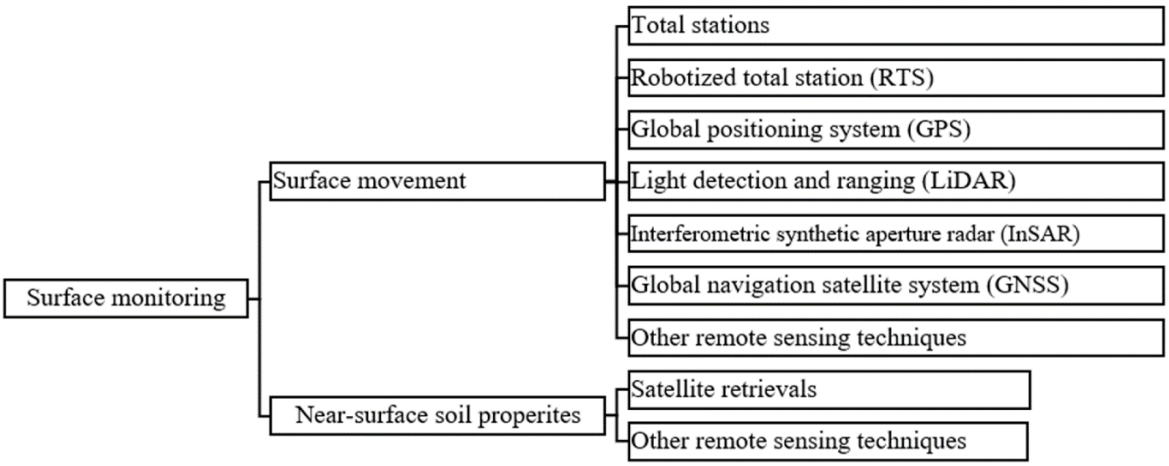



*Figure 2.4. Surface monitoring techniques.*

The subsurface monitoring systems (Figure 2.5) have the advantage of collecting the actual inner mechanism of the slope in real-time. Additionally, they have the following capabilities: a) monitoring deep displacement to locate where the landslide will initiate (Chuan et al., 2013) through several techniques, as illustrated in Figure 2.5. In addition, they can track the actual subsurface stresses and hydrological response and provide early warning signs for emergencies. However, they are incredibly costly when applied to national scales. As a result, they are suitable for local scales (i.e., single slope to 10 $km^2$ (Oguz et al., 2022). Additionally, they need additional research to improve their durability, precision, environmental impact, and maximum measuring range, as Mucchi et al. (2018) recommended when comparing other

surface methods (Table 2.2). As a result, this study will focus on local scales, the subsurface mechanism, and inner monitoring of the slope.

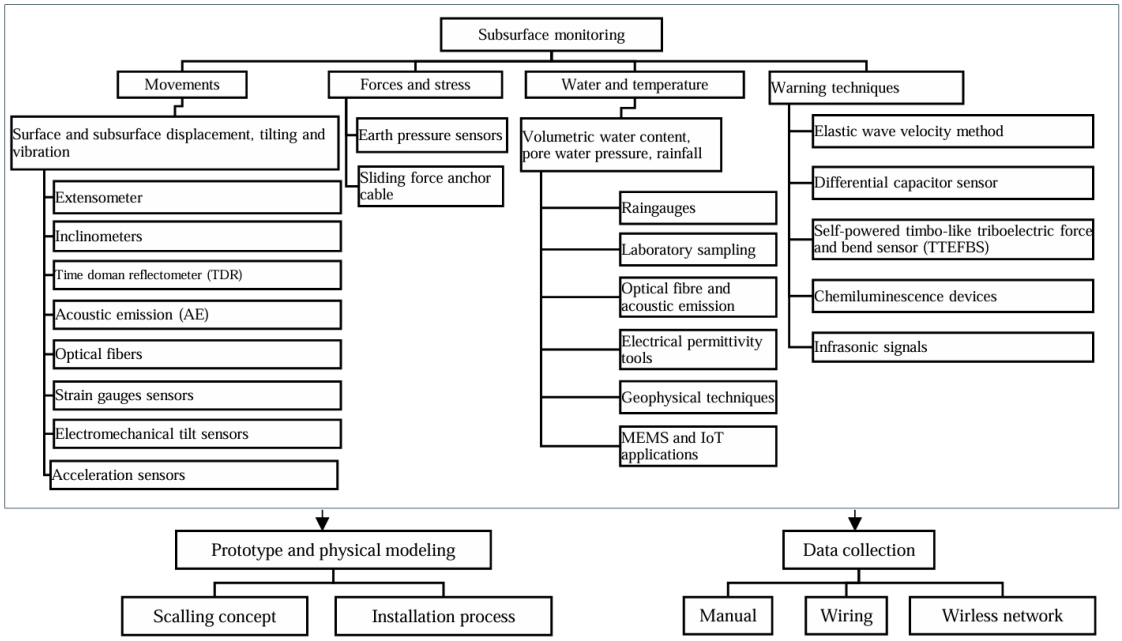



*Figure 2.5. Subsurface monitoring systems.*

*Table 2.2. A comparison of Wi-GIM, RTS, and GB-In-SAR (Mucchi et al., 2018)*

| Case | Wi-GIM | RTS | GB-In-SAR |
|---|---|---|---|
| Cost (Area = 100000 $m^2$) | 5220 euro | 18150 euro | 58100 euro |
| Environmental impact | Good | Very good | Very good |
| Installation effort | Excellent | Good | Very good |
| Influence of rain/snow | Very good/Very good | Good/Good | Poor/Very good |
| Completeness of measurement | Very good | Excellent | Excellent |
| Durability | Fair | Good | Excellent |
| Precision | Fair (2-3 cm) | Very Good | Excellent |
| Maximum range | Fair | Excellent | Excellent |

### *2.3.2. Subsurface Mechanical Response Monitoring.*

Movement monitoring is one of the most extensively used ways of determining a landslide's mechanical state. Table 2.3 highlights the advantages, disadvantages, and applications of each movement method. Tables 2.4 and 2.5 give more insights into movement approaches, taking into account other parameters such as cost, precision, range, initial accuracy, spatial resolution, maximum displacement, dynamic range, loading orientations, and sliding position. Extensometers, time domain reflectometry, acoustic emission (AE), fiber Bragg grating, and tilt systems are all illustrated in Figure 2.5.

*Table 2.3. The pros and cons of each movement monitoring sensor.*

| Sensor | Characteristics | Cons | Usage |
|---|---|---|---|
| Extensometers | • Have high temporal resolution (36 mm/s).<br>• Have high precision (0.011 ± 0.0083 mm).<br>• Measure surface movement at a single point. | • Cannot detect the sliding initiation depth or zone. | • Appropriate for translational landslides |
| Conventional inclinometers | • Outperforms extensometers, accounting for subsurface movement. | • Limited spatial resolution along its length (0.5 – 1 meter).<br>• Not suitable for sensitive deformation. | • Appropriate for narrow, shallow bandwidth landslides |
| Time domain reflectometry (TDR) and | • Able to extract information about the exact location of the sliding surface<br>• Has an excellent spatial resolution of 0.05 m<br>• The coaxial cable costs less than the inclinometer by 55% | • Quantifying the displacement is challenging | • Suitable for slow-moving landslides |
| AE | • More flexible compared to inclinometers<br>• Suitable for minor and sensitive movement | • Not suitable for fast-moving and large deformation monitoring | • Suitable for slow-moving landslides |
| Optical fiber-based inclinometers | • High initial accuracy of 0.98 mm<br>• Suitable measuring range of 36 mm<br>• Low cost of 0.45USD/m<br>• High spatial resolution of 10 mm<br>• Can be combined with FBG and BOTDA to monitor large areas | • Their application is limited as it is necessary to have too many wires. | • Best suited for rock landslides |
| Tilt sensors | • Able to detect the direction of a landslide with two-directional deformations<br>• Has a good accuracy of 0.0025 °<br>• Suitable measurement range of (-30 ° to +30 °)<br>• Can be combined with inclinometers to create multimodal systems. | • Title sensors are single points unsuitable for monitoring large areas. | • Sensors with short rods are suitable for circular slip surfaces.<br>• Sensors with long bottom fixed rods are suitable for shallow landslides. |
| Inclinometers based on strain gauges | • Can detect micro-movements.<br>• The Strain Gauge Deep Earth Probe (SG-DEP sensor) can give 360-degree directional measurements. | • A proper understanding of the initial conditions is required, as if the casing is stiffer than the surrounding soil, it will be difficult to track the minor movement or the movement in the shallow strata above the stiff layer. | • Suitable for both shallow and deep-seated landslides |
| Soil deformation sensors (SDS). | • They have low stiffness compared to conventional methods.<br>• Able to track minor movement of 1 mm with a suitable range of 25 mm. | | • Preferred for fast landslides |
| Acceleration sensors | • It can detect any movements regardless of the external triggering, where linear acceleration is the most influential characteristic. | • Necessitate advanced data processing to better correlate the accelerations (linear and gravitational) with motion. | • Suitable for fast transitional landslides without tilting |

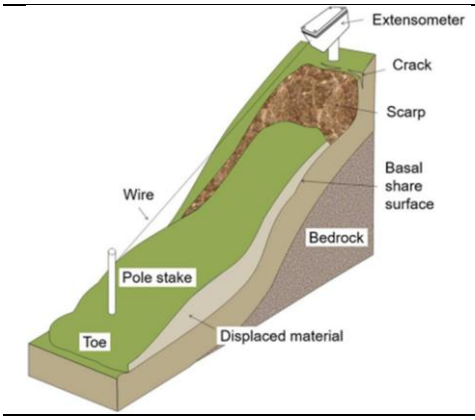
*Table 2.4. Comparison among GPS, Extensometer, IN, TDR, and AE.*

| System | Range | Precision | Displacement | Cost (USD) |
|---|---|---|---|---|
| GPS | - | 3 mm | Surface | High (6000-10000/station) |
| Extensometer (Setiono et al., 2023) | Up to 1000 mm | 0.011± 0.0083 mm | Surface | High (600-1500/station) |
| Inclinometer (IN) | < 50 mm (Dixon et al., 2015) | ±0.01 mm per 500 mm | Deep | High (600/sensor) |
| TDR | 60 mm (210-mρ) | 0.5 mm (Chung et al., 2022) | Deep | Low (6-10/m) |
| AE | > 500 mm | 0.0001 mm/h to 400 mm/h | Deep | Low |

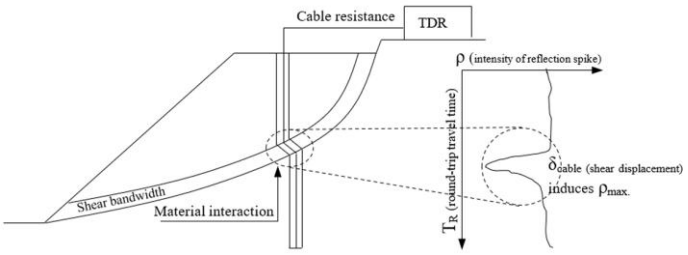
*Table 2.5. A comparison between IN, TDR, SOF, COFT, PSCFODS, and FBG-based inclinometer.*

| Method | Initial accuracy (mm) | Max. disp. (mm) | Spatial resolution | Dynamic range | Loading direction | Sliding location | Price (USD/m) |
|---|---|---|---|---|---|---|---|
| IN | 0.01 | < 50 | 500 mm | - | Yes | Yes | 30 |
| TDR | 0.5 mm | 60 (210-mρ) | 50 mm | 0–20.4 | Yes | Yes | 13.5 |
| SOF | 0.3 | 3.6 | - | 0-3.3 | No | No | 0.03 |
| COFT | 0.98 | 36 | - | 0-34 | Yes | Yes | 0.45 |
| PSCFODS | 0.98 | 36 | 250 mm | - | Yes | Yes | 0.2 |
| FBG | 0.02 | 50 | 100 mm | - | Yes | Yes | - |

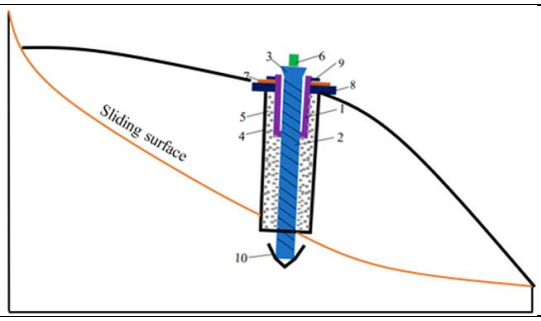
However, movement monitoring cannot thoroughly analyze the failure mechanism. Conversely, landslides may be predicted by tracking the ground pressure and sliding force in real-time, as sliding force is the most helpful measure for determining the kinematic features of landslides (Chuan et al., 2013). Instruments such as seismic vibrations and earth pressure cells can be buried in the soil (Askarinejad et al., 2018; Ma et al., 2017). Moreover, a cable anchor system can be installed, as shown in Figure 2.6, where the status of slope safety is determined by the actual force in the anchor (Tao et al., 2019). Figure 2.7 illustrates the monitoring system and monitoring curve phases, which are categorized into three groups: (1) the steady stage, (2) the steadily increasing stage, and (3) the stable stage (Tao et al., 2019).



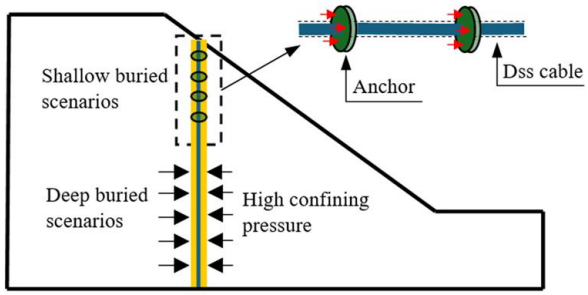

a) Extensometer Device (Setiono et al., 2023)



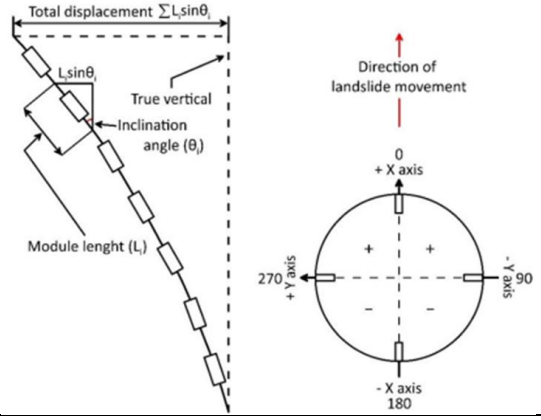

b) Time Domain Reflectometry (TDR) (Reprinted/adapted with permission from (Chung &Lin 2019), Elsevier)



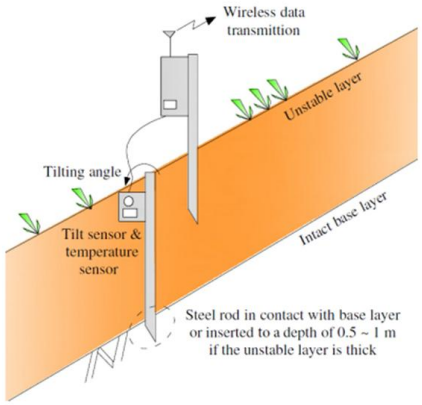

c) AE : (1) sleeve with inner wall threaded; (2) anchor cable; (3) conical metal head; (4) rubber tube; (5) backfill material; (6) AE transducer; (7) ring dynamometer; (8) pedestal; (9) nut; (10) anchoring end (Deng et al., 2021).




d) Fiber Bragg grating (FBG) (Reprinted/adapted with permission from Zhang et al., 2023), SpringerNature)




e) Multimodal biaxial tilt inclinometers (Ruzza et al. (2020))




f) Electromechanical (MEMS) tilt Sensors (Qiao et al., 2020)

*Figure 2.6. Schematic representation of several subsurface movement systems.*

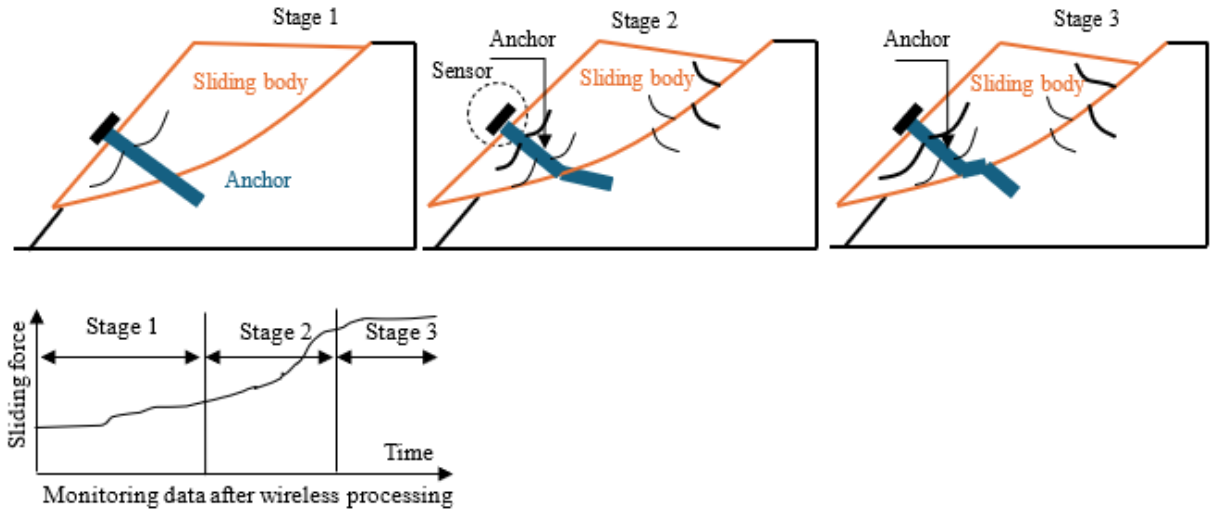



*Figure 2.7. Sliding force monitoring system (Adapted with permission from Tao et al., 2019).*

### *2.3.3. Subsurface Hydrological Monitoring.*

As for rain monitoring, which is considered to be the main triggering feature in Hong Kong, Rain gauges are widely adopted and can be classified into mechanical, optical, electrical, visual, and radar types, with mechanical rain gauges (tipping-bucket rain gauges - TBRs) being the most widely used and accurate. The mechanical type can monitor rainfall directly, whereas the other approaches use indirect measurement (Hu et al., 2022). Traditional TBRs suffer from limited measurement accuracy and significant abrasion under heavy rainfall conditions, making real-time rainfall monitoring and rain energy harvesting impossible. Hu et al. (2022) developed a novel TBR based on multiple triboelectric nanogenerators (TENG) units, a self-powered wireless sensor with a high measurement range (0 to 288 mm/d) and resolution (5.5 mm). However, a single issue with using rainfall data is that the rainfall criteria do not account for the underlying landslide process. Thus, better landslide prediction requires knowledge of subsurface variations in matric suction, moisture content, groundwater variability, etc.

Subsurface water monitoring includes site investigation and laboratory sampling, optical fiber and AE methods, electrical permittivity tools, geophysical techniques, and IoT applications. Tensiometers, for example, are commonly used to measure suction (Minardo et

al., 2021), whereas piezometers are utilized to monitor groundwater levels (Prabha et al., 2017; Ramesh & Rangan, 2014; Ramesh & Vasudevan, 2012). Recent research has focused on moisture dynamics as an appropriate technique for representing the antecedent effect and infiltration-surface run-off dynamics. The VWC measurement shows the antecedent influence of rainfall. Several strategies can be utilized for these purposes: The drying approach for evaluating soil moisture in a laboratory has remarkable accuracy, but it is a labor-intensive operation (Xiaochun et al., 2022); AE approaches have difficulties in correlating soil moisture to acoustic waves since it is influenced by soil density, void ratio, effective stress, etc (Kong et al., 2017); Fiber Bragg grating (FBG) may detect up to 37% VWC (Leone et al., 2017) as visualized in Figure 2.8; Ultrahigh-frequency radio-frequency identification (UHF RFID) sensors with self-charging capabilities may detect soil moisture levels as high as 16% (Pichorim et al., 2018); The smart aggregate (SAs) technique can monitor soil moisture up to 30 % (Kong et al., 2017); Geophysical approaches, such as electrical resistivity tomography (ERT), can provide information about large areas rather than isolated points with plot-scale variations. This method can detect soil moisture as deep as 2 meters (Wicki & Hauck, 2022; Xiaochun et al., 2022). The disadvantages of electrical resistivity are decreased resolution with depth and the lack of direct information (Chen et al., 2018a). IoT technology is becoming increasingly prevalent (Marino et al., 2023). Similarly, MEMS sensors offer a potential option for providing real-time data at a reasonable cost (Abraham et al., 2020a; Chu et al., 2021). However, these approaches use point sensors with low spatial resolution and considerable power consumption. Soil temperature monitoring, especially in the shear band zone, is found to be an essential subsurface feature to monitor as the thermal sensitivity plays a crucial role in the slope stability analysis (Segui & Veveakis, 2021; Seguí & Veveakis, 2022). However, this method necessitates a preliminary analysis to determine the position of the shear band before installing the temperature monitoring system (thermometers). Another approach that may be

utilized for early warning indications is surface temperature monitoring, since it has been observed that moving mass temperatures are significantly greater than nonmoving mass temperatures (Ma et al., 2017). However, the published literature on this subject is sparse and requires further study.

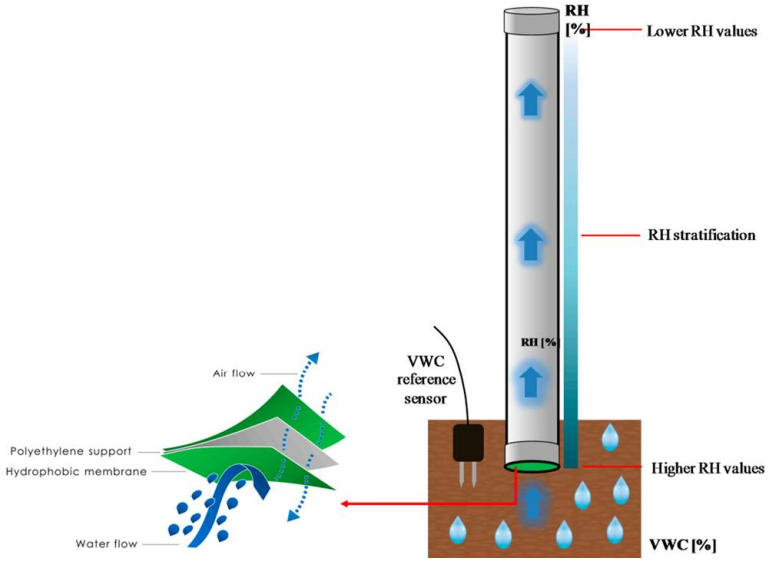



*Figure 2.8. An optimized version of the optical soil moisture instrument (Leone et al., 2017)*

### *2.3.4. Subsurface Warning Monitoring.*

The monitoring systems illustrated above focus on accurately quantifying the inner landslide response. However, they may fail to deal with emergency cases (Wang et al., 2017). Warning signs monitoring techniques aim to initiate signs without accurately quantifying the landslide response. The elastic wave techniques track subsurface moisture variations in the study area. It is preferred to single-point monitoring systems since it is damage-free and may be less expensive when covering a larger area (Chen et al., 2018a). Figure 2.9 compares traditional sensors (point sensors) with the elastic wave approach. Figure 2.10. a depicts how monitoring wave velocity can provide a failure alert. The difference in wave velocities shows a highly saturated zone that may be utilized to design a successful landslide prediction system since the internal structure of soil particles, shear modulus, void ratio, soil moisture content, and soil deformation impact it. Figure 2.10. b shows the elastic wave response throughout rainy seasons

(i.e., yielding). It was found that wave velocity decreased consistently with rainfall as moisture and movement increased, and a noticeable drop in wave velocity was observed after failure. However, this technique lacked field deployment due to the following challenges: a) this system requires a constant high-power source; b) the wave frequency must be investigated, as powerful waves may affect slope stability and weak waves may be affected by noise; and c) available studies investigate a single soil layer, necessitating investigation of layered soils. Low-frequency AE (infrasonic) signals can monitor multiple landslides in the same way as elastic waves can. Zhang et al. (2020) demonstrated that the infrasonic approach may detect subsurface microscopic variations produced by sliding forces (i.e., changes in void ratio or porosity). At failure, they can record a high-energy infrasonic signal that indicates macroscopic soil change and can be associated with sliding force. However, external noise, such as wind, thunder, and motor traffic, impacts this approach, which must be filtered out.

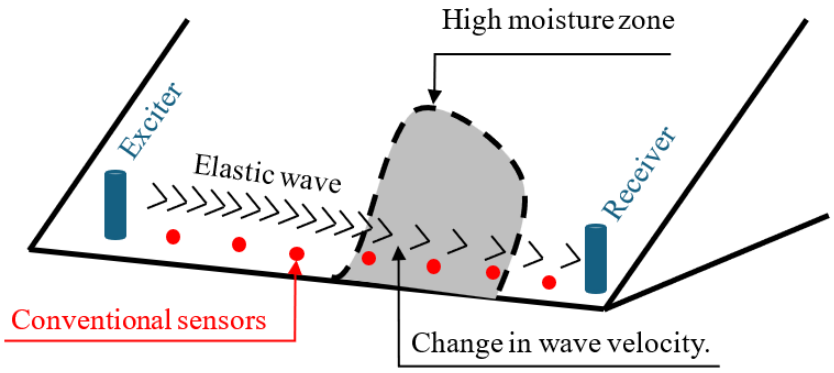



*Figure 2.9. Comparison between elastic wave velocities and conventional methods (Adapted with permission from (Irfan et al., 2017), Elsevier)*

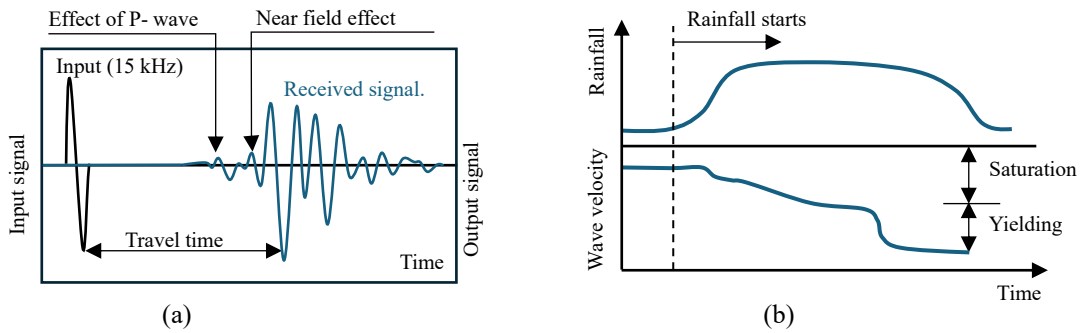



*Figure 2.10. a) Shear wave signal versus time; b) Elastic wave velocity during rainfall events (Adapted with permission from (Irfan et al., 2017), Elsevier)*

Motakabber and Ibrahimy (2015) designed a differential capacitor-type sensor warning system with an underground penetration wire and a capacitor gauge. When the soil begins to move, the development of the force-on-force plate and the penetration wire causes a change in the differential capacitor, providing warning signals. (Lin et al., 2018) used a self-powered timbo-like triboelectric force and bend sensor (TTEFBS) to detect subsurface movement. The sensor has a rapid reaction time (<6 ms), long-term durability (>40000 cycles), and high compression and bending sensitivity (5.20 V/N and 1.61 V/rad). Wang et al. (2017) used an accelerometer to send warning signals for any rapid movement. Kuang (2018) examined the chemiluminescence-based method, in which the reactants are stored in separate compartments and, when deformed, quickly produce light, making them simple to detect using low-cost optoelectronics. Since chemiluminescence is entirely passive, no electricity is required to function. On the other hand, this technology is sensitive to small deformations.

### *2.3.5. Data Collection Systems.*

If the monitoring systems fail to deliver data appropriately, it indicates that the system failed to fulfill its goal. Data can be collected manually or via wiring systems or wireless networks. The manual and wire techniques confront several obstacles, such as the difficulty of constructing in dangerous areas, human-caused destruction, natural disaster damage, etc (Yueshun & Wei, 2013). Conversely, wireless sensor networks (WSN) offer several benefits over conventional methods, including a) covering a greater area with less wiring, b) collecting

and analyzing data from multiple points, c) installing WSN without the need for pre-existing infrastructure, d) being less susceptible to environmental impacts; e) being more energy-efficient than conventional systems (Giri et al., 2019; Khoa & Takayama, 2018; Mucchi et al., 2018; Ramesh & Vasudevan, 2012; Rosi et al., 2011; Yueshun & Wei, 2013). Figure 2.11 depicts the WSN's components: sensor nodes, sensor topology structure, gateway, and monitoring center. The nodes deliver monitored data, the gateway connects the nodes to the internet, and the monitoring center stores, processes, and analyzes the data. Mesh, star, and tree architectures are commonly utilized (Jeong et al., 2020). LoRa modulation technology is a viable solution for node communication (Ragnoli et al., 2022). However, WSN has energy consumption, memory size, and communication difficulties (Kumar et al., 2020b; Mucchi et al., 2018), as numerous researchers have attempted to overcome these challenges, as illustrated in Table 2.6. However, among the methods illustrated, employing advanced methodologies such as DL algorithms and accounting for the geotechnical characteristics and the prediction accuracy to overcome such challenges is lacking.

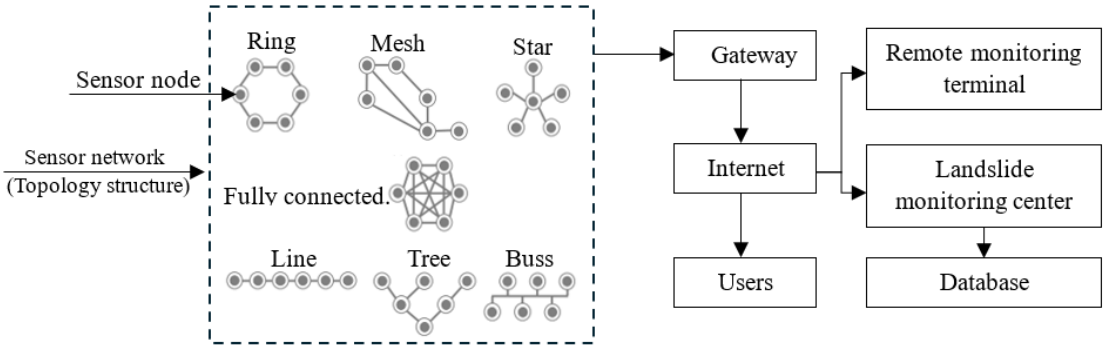



*Figure 2.11. Wireless network structure (Modified from (Jeong et al., 2020; Yueshun & Wei, 2013))*

### *2.3.6. Subsurface Deployment Techniques.*

One of the most important components of landslide monitoring is the placement of such instruments in the field. Dust, vibrations, accidents caused by falling boulders, and other hazards endanger the health and life of laborers who are installing monitoring systems. These

systems are sometimes costly to build and operate, limiting their use to well-funded initiatives. As a result, a MEMS system was recently developed that allows sensor modules to be lodged in the ground using a tiny hammer, making it appropriate for shallow landslide monitoring. Chuan et al. (2013) illustrated the deployment of any subsurface system where the process is outlined in Figure 2.12. The procedure is visualized and employed by Sheikh et al. (2021), in which the system consists of small-sized sensors (Figure 2.13). This process is as follows: (1) soil removal; (2) installing a borehole; (3) removal of the borehole casing; (4) inserting a steel rod; (5) mounting the tilt sensor;(6) powering and cabling the system.

*Table 2.6. Literature-based studies regarding landslide monitoring data-related issues.*

| Study | Issue | Solution | Notes |
|---|---|---|---|
| (Giri et al., 2019; Marino et al., 2023; Prabha et al., 2017; Wang et al., 2017; Yang et al., 2017) | Energy consumption (Pre-issue techniques) | Lowering the frequency of data. | • WSNs can reduce data transmission during dry seasons by establishing thresholds for sensor collection and shutdown, or by increasing frequency during wet seasons and decreasing during dry seasons. |
| (Jeong et al., 2020; Ramesh & Rangan, 2014) | | Limiting the number of active sensors. | • If a modest accuracy is accepted, there is no need to have many monitoring sensors.<br>• Geotechnical investigation is beneficial in creating a susceptibility map to install sensors only in vulnerable places. |
| (Lin et al., 2018; Qiao et al., 2020; Sheikh et al., 2021; Xu et al., 2020; Yang et al., 2017) | | Improving the self-rechargeability of the power system | • These systems include Solar batteries, Zigzag-structured triboelectric, Nanogenerators, Wireless power transfer, and Magnetic resonance wireless power transfer |
| (De Souza & Ebecken, 2012; Li et al., 2021; Shentu et al., 2023; Wang & Zhao, 2023) | Missing data (Post-issue techniques) | Superficial and shallow statistical correlations. | These models, such as:<br>• Second-order polynomial equations for trend time series.<br>• Paired adaptive regressors cumulative sum for periodic data.<br>• Principal component analysis with artificial neural networks.<br>• Small Feedback Optimizing Background Gray Model<br>• Mean-based low-rank autoregressive tensor completion<br>• Cubic spline interpolation |
| (Hemalatha et al., 2019; Long et al., 2022) | | Machine learning and multi-feature fusion transfer | • Assume similar geographical and geological characteristics are comparable but different in magnitude. |
| (De Graff, 2011) | | Parallel monitoring | • Install a spare system that can be used as a second alternative for missing data. |

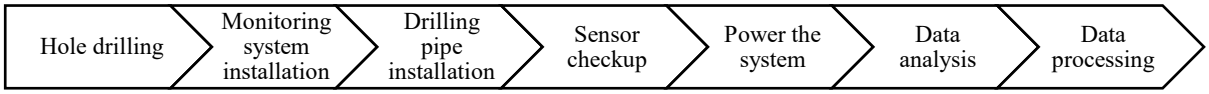



*Figure 2.12. Subsurface deployment flowchart.*

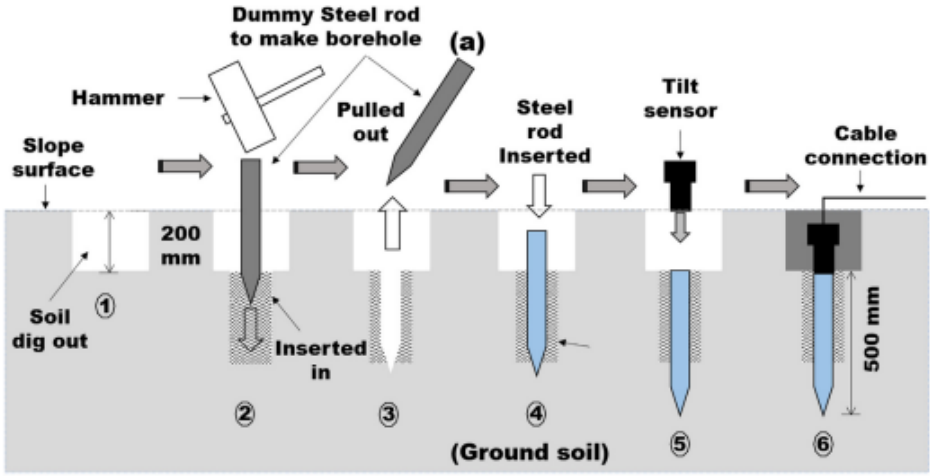


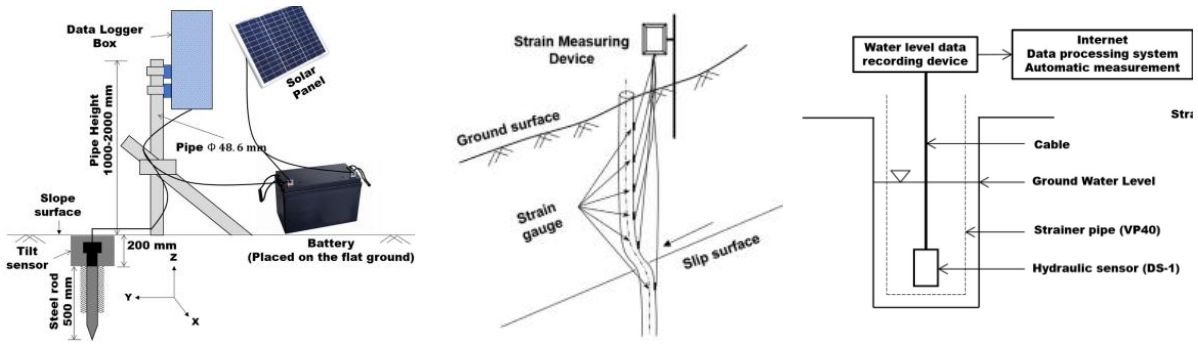

*Figure 2.13. Subsurface monitoring system deployment visualization (Sheikh et al., 2021).*

It should be mentioned that field studies and laboratory testing are necessary before placing any monitoring system in operation (Jeong et al., 2020). Basic details on the classification of landslides, the characteristics of the soil profile, the location of the sliding surface, etc., may be found in a site study. Four techniques of geographical distribution can be used to choose the monitoring locations: random, matrix, vulnerable, and hybrid. On a slope that is prone to landslides, the monitoring points in a random technique are placed at generic random positions. With the matrix approach, the whole deployment area is divided into a matrix of NxN cells, with one monitoring probe positioned in each matrix cell. To use the vulnerability approach, monitoring stations are positioned in vulnerable (also known as crucial zones) that are found by site inquiry, topographical mapping, and soil testing. The hybrid strategy uses both the vulnerable and matrix techniques, i.e., it starts with the matrix and modifies the monitoring device placements according to the most susceptible (i.e., crucial) areas. However, the literature lacks proper optimization of the sensor distribution, where advanced numerical modeling of the study area can predict the vulnerable location considering both the initial conditions and the external triggering effect.

Most of the preceding methods can offer a single measurement (e.g., displacement, soil moisture, etc.); however, the possibility of high false alarms limits their usage. Thus, such data should be correlated with other monitoring data, such as rainfall or soil moisture, to reduce such effects. As a result, multi-monitoring systems are strongly advised. Multimonitoring systems can be produced by combining the individual systems shown above or by designing a single sensor node with many functionalities. Multifunctional sensor devices that make use of MEMS sensors and WSNs are now commercially available. Table 2.7 reviews the literature on multi-point monitoring field system characteristics and their limitation. For instance, Sheikh et al. (2021) built prototype field experiments to investigate the relationship between tilt angle, displacement, strain, ground level, and rainfall using wireless sensors (tilt sensor, pipe strain gauge, water level gauge, and rain gauge) (refer to Figure 2.14).

*Table 2.7. Literature multi-point monitoring field system*

| Study | System | Notes/Gaps |
|---|---|---|
| (Chuan et al., 2013) | Pore water pressure sensors, a stress sensor, and a displacement sensor | Focuses on designing the mechanical components of the sensor. |
| (Ramesh & Vasudevan, 2012) | Piezometers, dielectric moisture sensors, strain gauges, tilt meters, and weather stations | The deep displacement system has considerable stiffness that may affect results. |
| (Yunus et al., 2015) | Soil moisture, vibration on land, slope angle, soil temperature, air temperature, humidity, and atmospheric pressure | It focuses on programming the wireless system itself. |
| (Quoc-Anh et al., 2018) | Soil moisture, temperature, tilt meter, geophone sensors, and weather station | |
| (Jeong et al., 2020) | Rain gauge, tensiometer, soil moisture sensor, and inclinometer | Lacks displacement monitoring. |
| (Yang et al., 2017) | Soil moisture, soil matric suction, ground vibration, tilt, and rainfall sensors | Simply correlate the data.<br>No displacement monitoring. |
| (Abraham et al., 2020b) | MEMS tilt sensor and VWC | Tilt and VMC only |
| (Sheikh et al., 2021) | Tilt sensor, pipe strain gauge, water level gauge, and rain gauge | Mainly focus on the tilting measurement. |




(a) (b) (c)

*Figure 2.14. Subsurface monitoring system deployment visualization (Sheikh et al., 2021).*

## 2.4. LANDSLIDE PREDICTION TECHNIQUES

### *2.4.1. Local VS National Prediction Models.*

Landslide prediction is the process of accurately anticipating the spatial (location) and temporal (timing) components of such occurrences (Valentino et al., 2014). Prediction is a crucial step in reducing the landslide risk, as decision-making can take the necessary action earlier to save civilians and infrastructure. Figure 2.15 depicts landslide prediction approaches, which are classified into two groups based on the study region: local (i.e., single slope to 10 $km^2$) and national (i.e., hundreds to thousands of $km^2$) scales (Oguz et al., 2022). The prediction approach differs within the same category depending on the available data inventory. Empirical techniques can be used locally with rainfall and historical landslide databases, physically based causative models can consider surface or subsurface responses, and deterministic models can simulate the physical process using field and laboratory investigations. At the national level, landslide susceptibility, risk, and vulnerability maps can aid in identifying susceptible locations within broad areas for development and planning reasons.

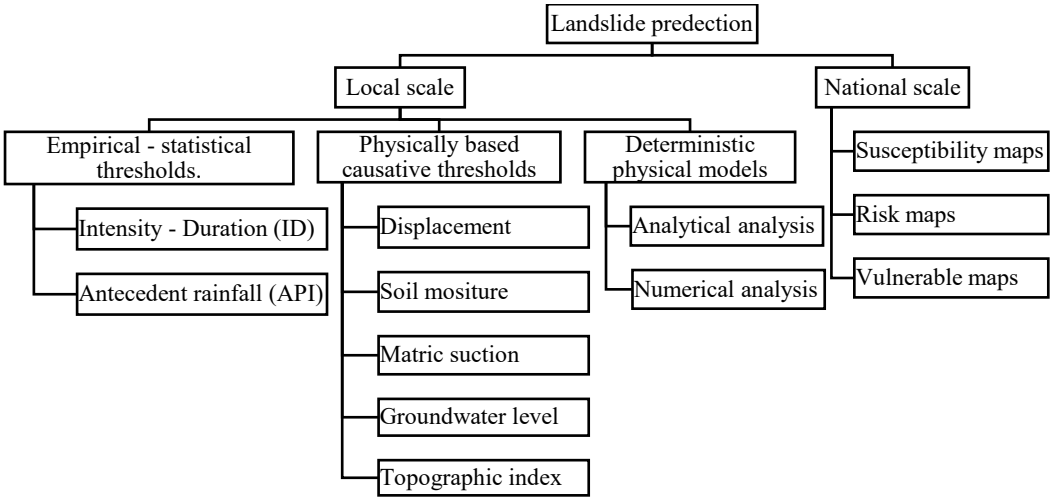



*Figure 2.15. Landslide prediction techniques classification.*

### *2.4.2. Empirical-Statistical I-D Thresholds.*

Rainfall thresholds are the lowest rainfall levels that could trigger landslides. This approach has been used in several investigations since it is simple and relies on rainfall and previous event statistics. For a specific location, the historical rainfall intensity (I) and duration (D) that causes the slope to fail are studied using statistical regression models to develop a warning threshold (Caine, 1980; Gariano et al., 2019; Hong et al., 2006; Zhao et al., 2019a). For instance, (Peruccacci et al., 2017) present a generalized form of rainfall thresholds in Equation 2.1 using a recent CTRL-T model to extract rainfall features automatically (Gariano et al., 2019; Melillo et al., 2018). While these thresholds may give acceptable accuracies, they may also produce high false alarms since they are fixed and do not account for antecedent effects.

Recently, the literature has investigated additional features to better comprehend the dynamic behavior of landslides. This is accomplished using the antecedent precipitation index (API). To illustrate, when evaluating rainfall-induced landslides, it is essential to account for the antecedent rainfalls, which can explain the temporal infiltration, storage, and evaporation processes (Hong et al., 2018; Uwihirwe et al., 2020). The antecedent duration varies with the slope characteristics and boundary conditions: this period might range from 1 hour to 30 days or even more (Hong et al., 2018; Lee et al., 2014; Zhao et al.,2019b). Hong et al. (2018)

efficiently demonstrated the antecedent effect using inter-event time definitions (IETDs), the time intervals between two separate rainfall events, as represented in Figure 2.16. According to the IETD, the effective antecedent effect is the rainfall quantity with a non-rainfall time of no more than the IETD. However, as shown in Figure 2.17, these systems generate false positive alarms (Uwihirwe et al., 2020; Zhao et al., 2019b).

$$E = (\varepsilon \pm \Delta\varepsilon) D^{(\xi \pm \Delta\xi)} \qquad \textit{(Equation 2.1)}$$

Where E is the rainfall event, ε is a scaling parameter, and ζ is the shape parameter, defining the intercept and slope of the power law curve, respectively. Moreover, Δε and Δζ represent the uncertainties associated with the two parameters.

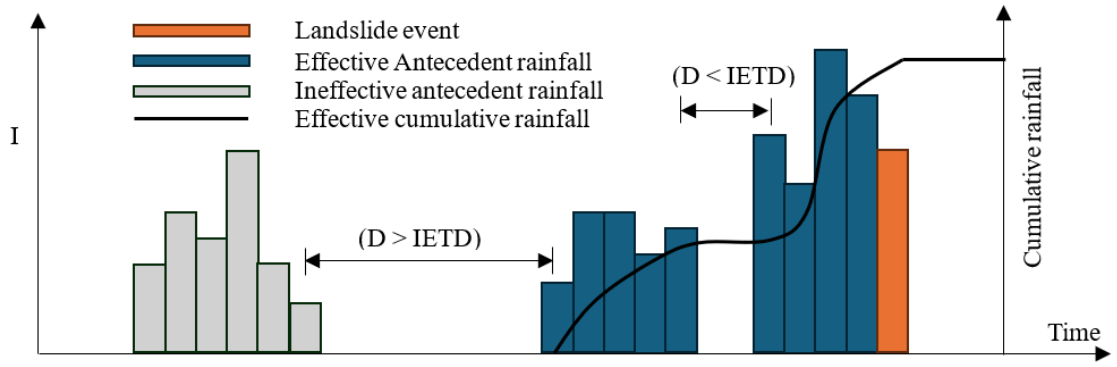



*Figure 2.16. Rainfall pattern analysis in terms of IETD.*

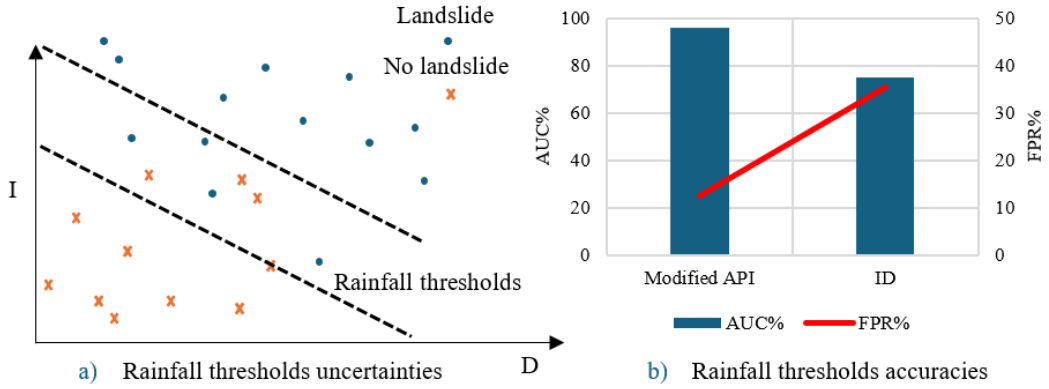



*Figure 2.17. Prediction accuracy (AUC) and false positive rate (FPR) for I-D and API thresholds*

### *2.4.3. Physically Based Causative Thresholds.*

Section 2.4.2 shows that empirical thresholds possess a high false positive rate. This could be illustrated as follows: Landslides are caused by severe rainfall or long-term antecedent effects. Empirical thresholds can consider this mechanism in the case of heavy rainfalls, in which the slope fails with a temporal reliance on the rainfall event. However, in the case of a long-term antecedent influence, when the landslide event might occur with minimal rainfall or even without rainfall, empirical thresholds are ineffective, emphasizing the need to consider hydrological and mechanical responses into account such as moisture content (Bezak et al., 2019), displacement (Bednarczyk, 2018), suction (Davar et al., 2022), groundwater level (Cao et al., 2020), and topographic features (i.e., spatial soil thickness variation) (Ho et al., 2012).

Figure 2.18.a depicts the response of surface displacement, rainfall, and reservoir level variation, whereas the landslide response exhibits temporal delay and a complicated process. As a result, statistical techniques like linear regression analysis require previous knowledge and skill to understand such complex behavior (Bednarczyk, 2018; Chen et al., 2018b; Sasahara, 2017; Zhang et al., 2016). For example, Li et al. (2021) used the short-term forecasting of landslides (STFL) formulation to predict deformation in the accelerated deformation stage based on data from the primary and secondary phases (Equation 2.2 and Figure 2.18.b). However, natural slopes are composed of more complex, non-uniform soil layers, necessitating more complex algorithms to comprehend these kinds of responses better.

$$S = P_1(P_2^{P_3 t} - 1) / (P_0 - t) \qquad \textit{(Equation 2.2)}$$

Where $P_0$, $P_1$, $P_2$, and $P_3$ are the simulated parameters, t denotes time, and S denotes displacement.

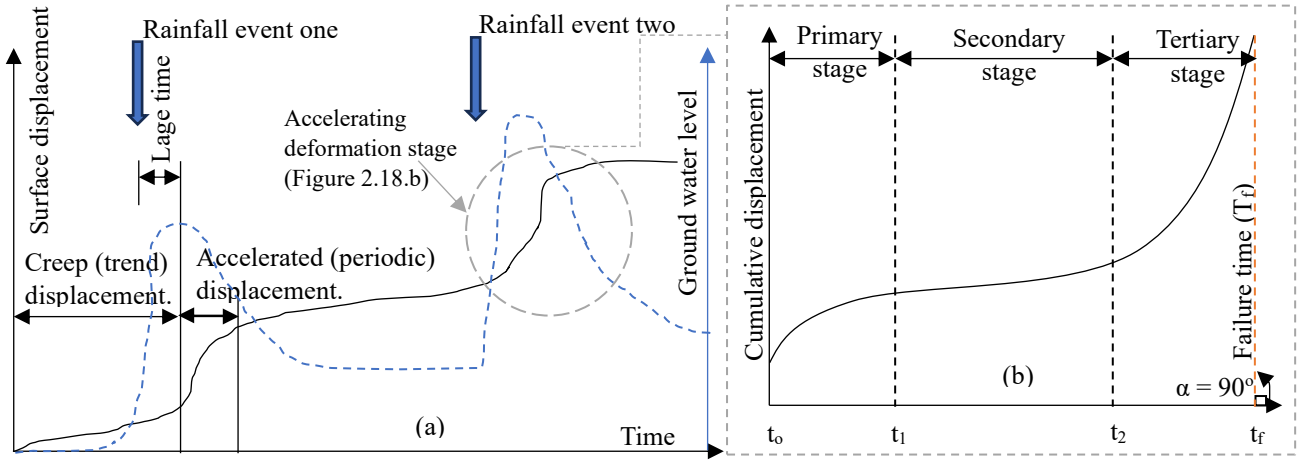



*Figure 2.18. Displacement response with rainfall and groundwater.*

AI models have demonstrated exceptional accuracy in forecasting non-linear responses (Tengtrairat et al., 2021). Figure 2.19 indicates the models for predicting surface displacement. These models can be single models, although they can be enhanced utilizing hybrid systems due to the dynamic nature of landslides and their triggers (Gao et al., 2020). To illustrate, surface displacement has been found to have trend and periodic behaviors: the trend components are simple to predict, whereas the periodic terms are more complex. The trend behavior of landslides is primarily determined by the creep effect, which includes lithology, structure, and stress state. In contrast, the periodic response is influenced by random events such as rainfall (Lian et al., 2016). The prediction model's performance depends not only on the intelligence model but also on the controlling factors. Discussing intelligence models is governed by their controlling features, such as the high dependency between the model and the features, dataset quality, sampling ratio, accuracy of inventory data, etc. As a result, there is no superior model but a superior prediction (Han et al., 2021). Outdated or non-related features can negatively affect the model, and it is better to remove these factors from the model (Cao et al., 2016). It has been found that BPNN (Liu et al., 2016) and MLR (Krkač et al., 2020) provide reasonable accuracy; however, these models are simple and cannot consider dynamic and non-linear relationships. This is because the well-established dataset closely represents the physical

mechanism. Marrapu et al. (2021) concluded that ANNs with an extensive dataset can achieve higher accuracy than ANNs with a limited dataset. However, if the dataset lacks valuable information, the model must be improved to fit these complex relations.

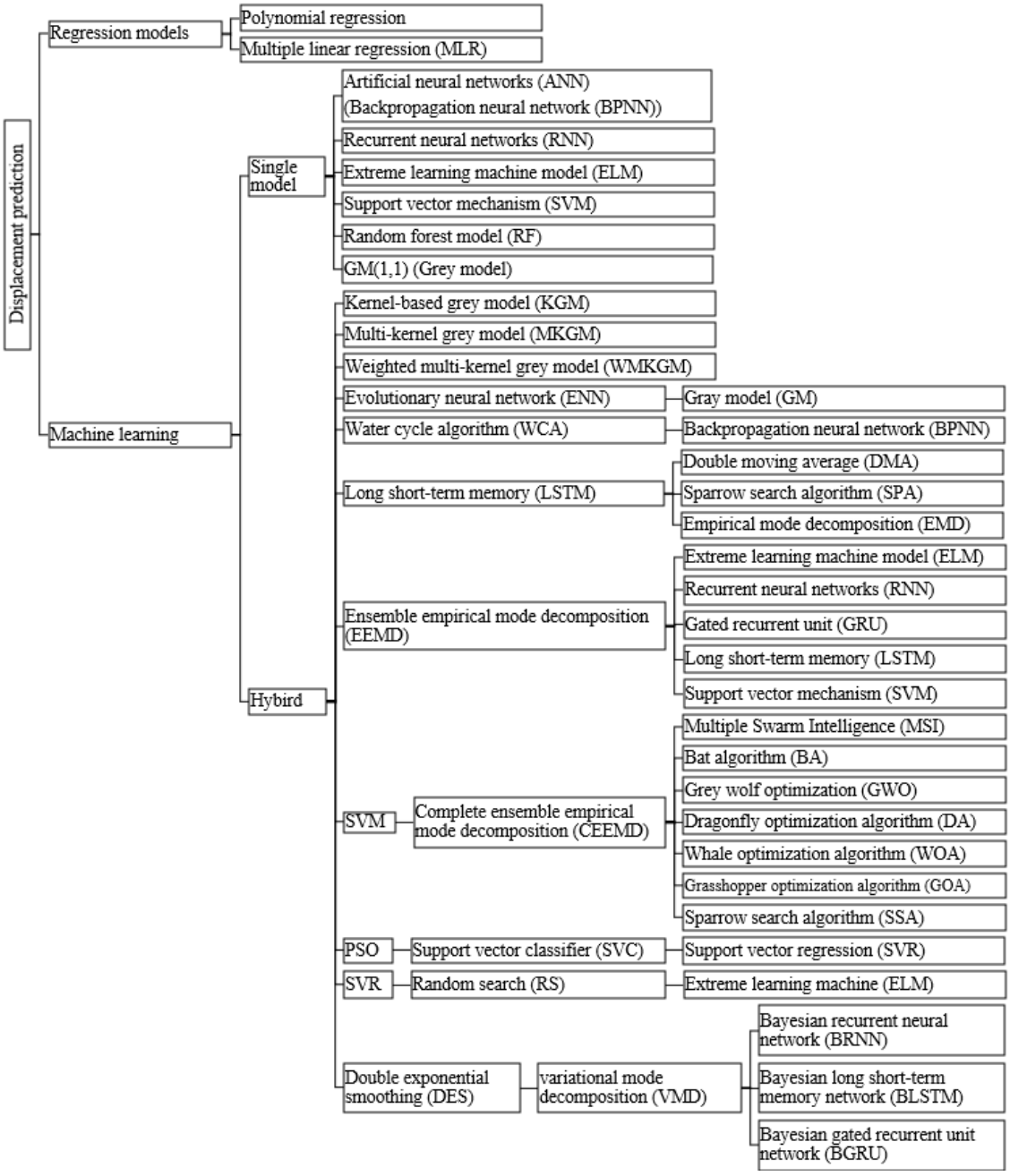



*Figure 2.19. Artificial models in landslide displacement prediction.*

A comparison between all models, considering only the final accuracy and performance of the models, is provided in Figure 2.20. The GM (1,1) model achieved the minimum RMSE among all single models for total displacement, while the PSO-SVR-SVC hit the minimum

RMSE among all hybrid models. The maximum $R^2$ can be achieved for creep displacement by adopting simple polynomial regression, whereas regarding periodic displacement, RS-SVR records the maximum $R^2$. Generally, trend displacement can be predicted using simple polynomial regression. However, the problem is the periodic displacement that exhibits complex and non-linear behavior, necessitating hybrid models such as RS-SVR.

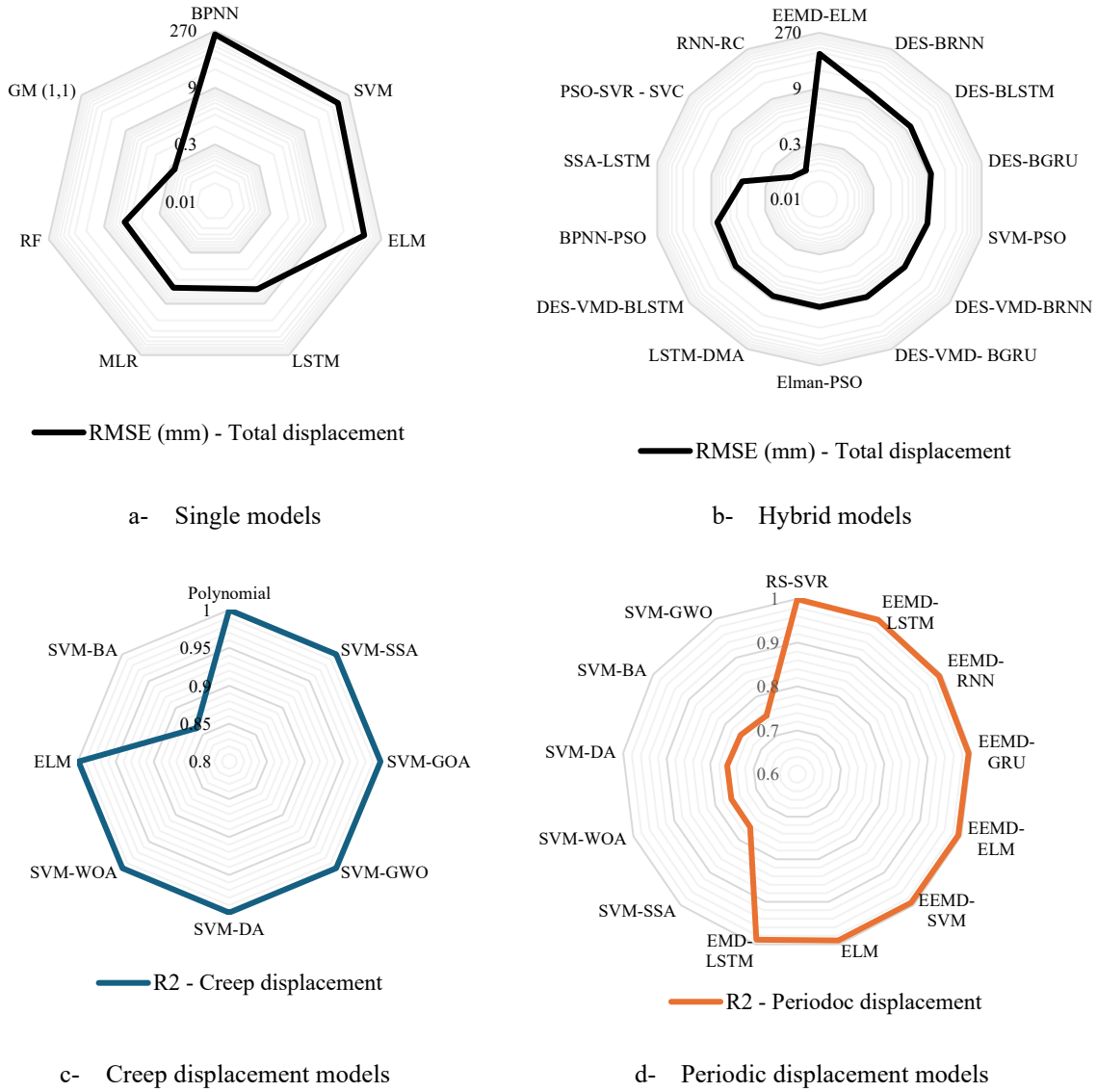



*Figure 2.20. Accuracies in predicting surface displacement.*

An exhaustive assessment of time series models for landslide prediction is conducted as illustrated in Table 2.8, revealing that DL networks perform very well in capturing dynamic subsurface reactions. DL's advantages include adaptability across domains, simplicity of

feature extraction, parameter optimization, and scalability. However, these skills are contingent on a detailed knowledge of the underlying physical processes of landslides, the quality of the monitoring data, and the empirical Hypertuning of the model parameters. This emphasizes the importance of an effective monitoring system capable of accurately capturing the physical response of the landslide, which is frequently delayed (i.e., the infiltration process depicted in the following lines), as well as overcoming harsh environmental conditions and data collection issues (refer to Table 2.6). The models above tried to predict the mechanical landslide response regarding surface displacement. It should be noted that several responses can be integrated to represent the slope response, such as the hydrological response (i.e., soil moisture, groundwater, suction, etc.). However, such hydrological responses are lacking due to the shortage of geotechnical data, while displacement prediction is widely adopted due to the availability of monitoring data.

*Table 2.8. Summary of the time series modeling main findings in the literature*

| | Criteria | Main finding |
|---|---|---|
| **DATA PREPARATION** | Data frequency | • Studies use two frequencies: monthly and daily time steps.<br>• Frequency selection is based on monitoring data availability, disregarding physical and computational backgrounds. |
| | Data splitting | • Two factors should be considered: the temporal sequence of splitting the dataset and the ratio of the training to validation sets:<br>• The temporal ordering should be maintained through a standard holdout strategy<br>• The testing set should represent at least the minimum expected temporal response. |
| | Decompositions | • Landslide response to landslides is complex, requiring various attempts to simplify the process.<br>• Decomposition can simplify the complex response into residual, trend, and seasonal behaviors.<br>• Differentiating methods outperform other decompositions, minimizing uncertainties and offering high prediction accuracy. |
| | Window size | • Considering an antecedent period (i.e., lagged sequence) improves model performance.<br>• The lagged period varies based on the slope's mechanical and hydrological characteristics.<br>• Research has not given enough consideration to the window size (sequence length) and how the antecedent value influences the prediction model's performance. |

*Table 2.8-Cont.*

| | Criteria | Main finding |
|---|---|---|
| **DATA PREPARATION-Cont.** | Feature selection | • Outdated or non-related features can negatively affect the model.<br>• Factors affecting landslides include creep and triggering features.<br>• Statistical methods for feature selection include gray relation analysis (GRA), partial autocorrelation function (PACF) algorithms, maximal information coefficient (MIC), kernel sHAP, Pearson correlation, R2-adj, Akaike information criterion (AIC), and least absolute shrinkage and selection operator (LASSO).<br>• Models often neglect temporal dependencies in landslide responses, leading to unrelated features.<br>• Knowledge-based methods and sensitivity analysis are recommended to consider temporal dependencies and select the best-related features. |
| | Statistical correlations | • Pearson's correlation coefficient is a common model used in feature selection algorithms due to its practical nature.<br>• The maximal information coefficient (MIC) outperforms Pearson's correlation coefficient as it extracts both linear and non-linear correlations, as well as complex correlations.<br>• The effective antecedent period can be sensitively investigated using statistical correlations. |
| **MODEL SELECTION** | Model selection | • Static models struggle to account for the temporal correlation between input features and external triggering.<br>• Dynamic and DL models can extract non-linear correlations between triggering and landslide response, where LSTM offers outstanding performance. |
| **OPTIMIZATIONS** | Hypertuning | • Techniques include grid search, random search, Bayesian optimization, particle swarm optimization (PSO), genetic algorithms (GA), successive halving (SH), sparrow search algorithm (SSA), etc.<br>• Grid search is a heuristic search method within a learning algorithm's hyperparameter space, but it is time-consuming.<br>• Random search uses a randomized sampling approach, but may not always yield the optimal hyperparameter combination.<br>• PSO refines solutions iteratively, evaluating their quality based on fitness.<br>• Genetic algorithms enhance learning efficiency and computational accuracy when tuning model hyperparameters.<br>• The SH model optimizes hyperparameters by efficiently allocating computational resources.<br>• The sparrow search algorithm (SSA) is an innovative method inspired by the foraging and anti-predatory behaviors observed in sparrows.<br>• The main concept can be simplified as the model searches from coarse to fine scales to accurately capture the best model structure while also converging faster. |
| | Training optimization | • Selecting the right number of iterations, learning rate, and monitoring metrics is crucial.<br>• Small iterations can cause bias issues, while long iterations can cause variance issues.<br>• A small learning rate requires large computational time, while a large learning rate may be misleading.<br>• Adam optimizers offer the best prediction accuracy, outperforming traditional gradient descent methods.<br>• The Adam algorithm dynamically adapts learning rates for each parameter, improving convergence speed and efficient exploration of the solution space. |
| | Loss functions | • Huber loss function integrates and refines both MSE and MAE, offering a balanced optimization approach.<br>• Monitoring and quantifying training and validation loss can help overcome underfitting and overfitting challenges. |

*Table 2.8-Cont.*

| | Criteria | Main finding |
|---|---|---|
| **OPTIMIZATIONS-Cont.** | Normalization | • Quantitative variation in triggering and landslide responses makes convergence challenging.<br>• Normalizing and scaling data is crucial and helps the model to converge more easily.<br>• Techniques include minimum and maximum normalization, mean normalization, and Z-score normalization. |
| | Regularizations | • Four techniques to solve this issue:<br>- Increasing data.<br>- Avoiding unrelated features.<br>- Monitoring the training process.<br>- Reducing the size of the training weights.<br>• L2 regularization or dropout layers can be added to reduce training weights while maintaining model output.<br>• The generalization term can be assigned a value greater than zero, where selecting and Hypertuning an appropriate value of regularization is critical |
| | Activation Functions | • Activation functions are higher-level features that convert input features to new ones that fit the labeled output.<br>• Several activations are used, including linear, ReLU, Leaky ReLU, tanh, and sigmoid.<br>• ReLU is preferred for hidden layers due to faster convergence, but is susceptible to the "dying ReLU" problem.<br>• Leaky ReLU addresses the dying ReLU by allowing a slight, non-zero gradient for negative inputs.<br>• Tanh & sigmoid offer bounded output ranges suitable for specific tasks but suffer from vanishing gradient problems and saturation issues.<br>• Softmax is used for multiclass classification. |
| **MODEL EVALUATION** | Model Evaluations | • Two techniques used for model evaluation: the unweighted method and the weighted method.<br>• The unweighted method assigns the same error weight to all datasets.<br>• The weighted method assigns different error weights for creep and mutual points.<br>• RMSE, MAE, and MAPE are the most commonly used metrics.<br>• R2 and absolute error are rarely used in literature.<br>• Other metrics like R2 adj, R, EF, PICP, RI, MASE, and SMAPE are rarely used. |
| **PREDICTIONS** | Predicted Target | • Surface displacement predictions are the most commonly predicted targets in studies of landslides.<br>• Hydrological response predictions, such as VWC, matric suctions, and groundwater level variation, are rarely considered.<br>• Reservoir landslides have received significant attention, but other types, like deep-seated landslides and rock slope types, have received less attention.<br>• Rainfall-induced shallow landslides are missing from the literature.<br>• More research is needed to account for the inner response of various types of landslides. |
| | Single and Interval Predictions | • Uncertainties in predictions arise from model assumptions, necessitating quantification.<br>• Studies often use single predictions, predicting a single time step ahead without quantifying corresponding uncertainties.<br>• Xing et al. (2020) provide a detailed derivation of calculating interval predictions, assuming a random variable with a zero mean Gaussian distribution and independent of the input variable $x$. |

To illustrate the physical process, Figure 2.21 illustrates the soil's moisture dynamic and saturation storage as provided by Lee & Ho (2009). The soil thickness above the bedrock

(*H*) is divided into the unsaturated zone (*z*) and saturated zone (*h*). When the cumulative rainfall exceeds the water storage in the unsaturated zone, the depth of the saturated zone increases and may reach ground level at full saturation based on the rainfall dynamic history. The saturated thickness can be calculated using Equation 2.3 (Ho & Lee, 2017). Bezak et al. (2019) developed a threshold (*R*) through a hydrological model called the production storage level (PSL). PSL considers the slope's wetness increases (WI) by accounting for the evaporation and discharge data (Equation 2.4). The R-PSL threshold produces meaningful results compared to I-D thresholds for long-duration rainfall. Kirkby (1975) proposed a topographic threshold that considers the spatial variation in soil thickness and accounts for the saturation thickness (Equation 2.5). Therefore, thresholds can be concluded based on these assumptions illustrated in Equations 2.6 and 2.7 (Ho et al., 2012). Zhao et al. (2019b) adopted the effect of antecedent soil moisture with recent rainfall using the Bayesian probabilistic approach. The soil moisture was calculated using the SHETRAN (Système Hydrologique Européen TRANsport) model for better accuracy than I-D thresholds. De Luca and Versace (2017) utilized the new GFM (Generalized FLaIR Model). The GFM considers the initial soil moisture that depends on antecedent rainfall. Additionally, several configurations can be defined to choose the most suitable threshold (Equations 2.8 and 2.9).

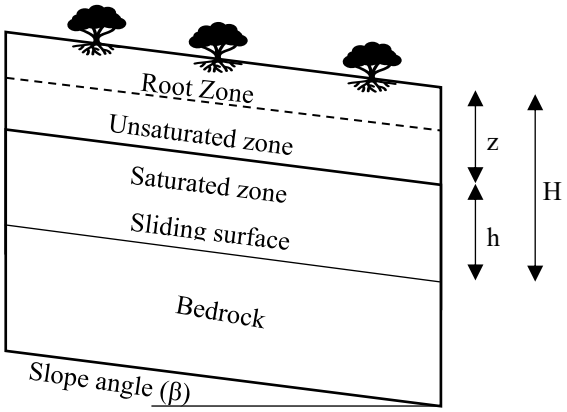



*Figure 2.21. The coupled hydrological-slope instability model.*

These physically based causative thresholds are built based on some assumptions: The relationship between the landslide and the features controlling this phenomenon will not change significantly in the future (Krkač et al., 2017); No sudden failure will occur (i.e., strain softening along the sliding surface) in the future (Xing et al., 2020); Causative thresholds consider only one limited feature, such as displacement or groundwater level, ignoring all other characteristics, such as spatial variation in land cover, soil type, topography, and geotechnical and hydrological parameters. Nonetheless, these models are favored where AI models can predict any causative features using available monitoring data; therefore, these models can provide a warning sign more accurately than empirical-statistical thresholds. These models provide a cost-effective method for landslide prediction because they do not require much geotechnical investigation.

$$h(t) = H - Z(t) - FRC\left[ ATI - \ln\left( \frac{W}{\tan\beta} \right) \right] \quad \textit{(Equation 2.3)}$$

$$R = a * PSL + b \quad \textit{(Equation 2.4)}$$

$$TWI = \ln(W / \tan\beta) \quad \textit{(Equation 2.5)}$$

$$\tan\beta \geq \left[ \frac{C}{\gamma_t g H \cos^2\beta} + \tan\phi \right] \text{ (lower thresholds)} \quad \textit{(Equation 2.6)}$$

$$\tan\beta < \left[ \frac{C}{\gamma_t g H \cos^2\beta} + \left( 1 - \frac{\gamma_w}{\gamma_t} \right) \tan\phi \right] \text{ (upper thresholds)} \quad \textit{(Equation 2.7)}$$

$$Y(t) = \int_{t-M}^{t} I(\mathrm{T}) \Xi(t - \mathrm{T}) d\mathrm{T} \quad \textit{(Equation 2.8)}$$

$$Y_{cr}(t) = \left( R^*_D(t-d)/D \right) + \left( f\left[ R^*_D(t-d) \right]/d \right) \quad \textit{(Equation 2.9)}$$

Where *h(t)* denotes the saturated water height at time *t*, *H* denotes the soil thickness, *W* represents the unit width of the collecting area, *tanβ* is the surface slope, and *ATI* means the average topographic index in the catchment. *FRC* is a model coefficient obtained through the

flow recession records. $C$ donates the cohesion, $g$ presents the gravitational acceleration, $\gamma_t$ is the bulk density of the soil, $\gamma_w$ is the density of water, $H$ is the soil thickness measured vertically, $\beta$ is the gradient of the hill slope, and $\phi$ is the angle of the soil's effective friction. $Y$ is a mobility function, $I$ is rainfall intensity, $\Xi$ is a filter function that can take different mathematical expressions, $t$ is time, $T$ is defined on the interval $[t-M; t]$, $M$ is the process's temporal memory, and $R^*_D$ is cumulative rainfall filtered on $D$ and $d$ durations and evaluated at the instants $(t-d)$ and $t$, respectively.

### *2.4.4. Physically Deterministic Models.*

Empirical thresholds and physically based causative models are widely adopted with early warning systems because of their simplicity compared to deterministic models. On the other hand, deterministic physical models show superior accuracy compared to these modes (Ho & Lee, 2017; Wang et al., 2020a) as they combine both hydrological processes (infiltration) and slope stability analysis (failure stress criteria). Hydrologically, Equation 2.10 shows the partial differential equation utilized to solve the infiltration moisture profiles using Darcy's law (Richards, 1931). Concerning slope stability analysis, Equation 2.11 shows the Mohr-Coulomb failure criteria for the unsaturated soil (Fredlund & Rahardjo, 1993). It should be emphasized that rainfall-induced fluctuations in pore water pressure affect soil stresses and deformation, causing a hydro-coupled effect. This complicates the infiltration and seepage investigation, necessitating sophisticated methods such as numerical modeling (Zhang et al., 2011).

Zhang et al. (2011) demonstrated the infiltration process as illustrated in the typical pore water pressure profiles in Figure 2.22. During the dry seasons, unsaturated soil exhibits larger matric suction, resulting in a greater slope safety factor. During the wet season, the suction decreases or diminishes due to infiltration, affecting the slope's stability. There are two categories for the pore water pressure distribution: transient states and steady-state conditions. The transient state is a transitional state between the initial state and the final state. The ground

surface flux, soil water storage, and hydraulic conductivity affect how long it takes to reach the final state. The matric suction disappears when the surface flux (q) exceeds or equals the saturated hydraulic conductivity (ksat). The matric suction decreases otherwise. Numerous attempts have been made through literature to simplify the process by understanding the initial boundary conditions. For example, in prolonged rainfall periods, some researchers assume the worst situation where the soil is represented as saturated soil, neglecting the effect of the infiltration (Equation 2.12) (Collins & Znidarcic, 2004; Ho et al., 2012). In some scenarios, it is reasonable to assume infinite slope stability because the depth of the failure surface to its length is shallow. Additionally, if the slope angle does not vary significantly along the slope height, the failure surface can be assumed to be parallel to the slope surface (Cho, 2016; Ho & Lee, 2017; Ho et al., 2012; Lee & Ho, 2009; Wang et al., 2020a).

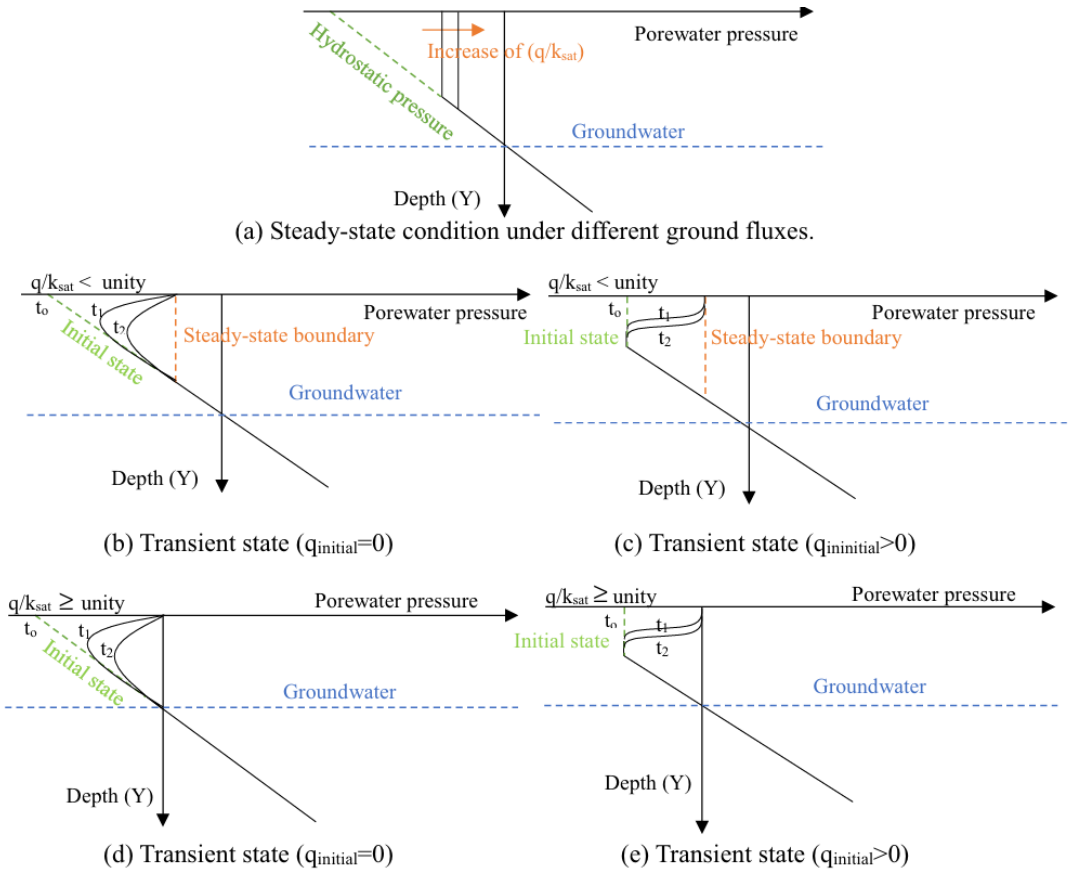



*Figure 2.22. Typical pore water pressure profile (Modified from Zhang et al., 2011)*

Another example, for thick or thin layers with free drainage from the bottom, the 1D infiltration model is suitable, whereas the Green Ampt (GA) model can provide a simple solution for the subsurface flow (Green & Ampt, 1911; Pagano et al., 2010). GA assumes the presence of a thin sheet of water at the ground surface. GA considers a consistent approach of the wetting front that divides the saturated wetting zone from the unsaturated soil zone with a constant value of soil matric suction (Figure 2.23). In contrast, Formetta and Capparelli (2019) showed that the infiltration process is a leading triggering for initially unsaturated soil, and Thang et al. (2022) and Valentino et al. (2014) demonstrated that the failure surface can be above or at the soil-bedrock interface. Lee and Ho (2009) account for the groundwater's time-dependent fluctuation, assuming a failure surface parallel to the slope surface (Figure 2.24 and Equation 2.13). Logically, the transient state provides reliable and accurate results compared with steady-state models. This finding is based on the comparison between dynamic and steady-state models developed by Liang and Uchida (2022) and Valentino et al. (2014).

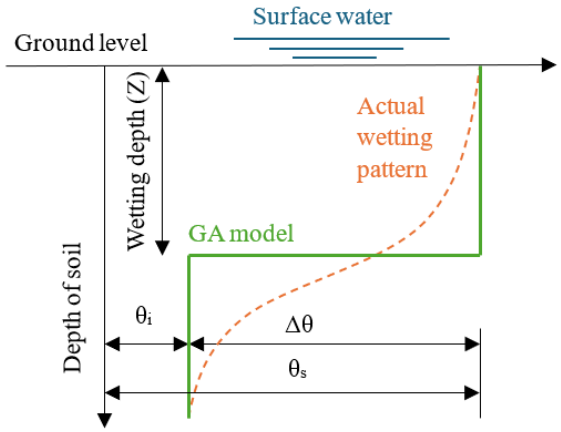



*Figure 2.23. Green Ampt profile*

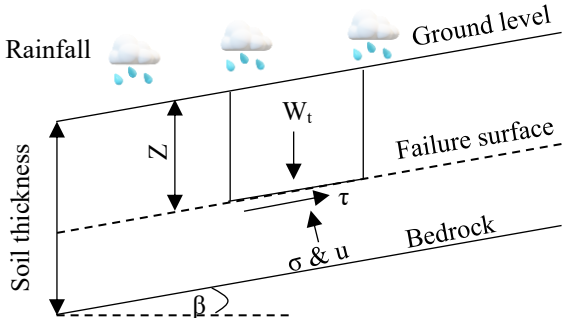



*Figure 2.24. Stability analysis of a shallow infinite landslide*

$$\frac{\partial}{\partial x}\left(k\frac{\partial h_w}{\partial x}\right)+\frac{\partial}{\partial y}\left(k\frac{\partial h_w}{\partial y}\right)+\frac{\partial}{\partial z}\left(k\frac{\partial h_w}{\partial z}\right)=-\frac{\partial \theta_w}{\partial t} \quad \textit{(Equation 2.10)}$$

$$\tau = c' + (\sigma - u_a)\tan\phi' + \psi\tan\phi^b \quad \textit{(Equation 2.11)}$$

$$FOS = \frac{C' + (H\gamma_t - h_w\gamma_w)\cos^2\beta\tan\phi}{\gamma_t H\sin\beta\cos\beta} \quad \textit{(Equation 2.12)}$$

$$FOS = \frac{C' + (W_t\cos^2\beta - \sigma_s)\tan\phi'}{W_t\sin\beta\cos\beta} \quad \textit{(Equation 2.13)}$$

$\theta_w$ is the VWC, $k$ is the unsaturated hydraulic conductivity, and $h_w$ is the total head. $c'$ is the effective cohesion; $\phi'$ is the effective friction angle; $\psi$=$u_a$-$u_w$ is the matric suction; $u_a$ is atmospheric pressure; $u_w$ denotes the pore water pressure; $\phi^b$ is the internal friction angle corresponding to the matrix suction; and $\sigma$ is the total stress. $C'$ is the effective cohesion; $\gamma_t$ is the soil bulk unit weight; $\gamma_w$ is the water unit weight; $h_w$ is the saturated soil thickness above the slip surface; $H$ is the soil thickness measured vertically; $\beta$ is the gradient of the hill slope; $\phi$ is the soil effective friction angle. $W_t$ is the weight of the soil slice, and $\sigma_s$ is the suction stress expressed in terms of the normalized VWC or effective degree of saturation $S_r$.

Numerous analytical models are available, each with its own assumption based on the initial boundary conditions of the investigated case study. Some of these analytical models include the TOPMODEL (Beven & Kirkby, 1979), Green & Ampt model (Green & Ampt, 1911), SHETRAN (Birkinshaw & Ewen, 2000), GR4J model (Perrin et al., 2003), TRIGRS (Baum et al., 2008), SIMTOP (Lee & Ho, 2009), Modified TOPMODEL (Lee & Ho, 2009), H SLIDER (Uchida et al., 2009), CREST (Wang et al., 2011), HIRESSS (Rossi et al., 2013), SLIP (Valentino et al., 2014), TiVaSS (An et al., 2016), Modified Green-Ampt (Cho, 2016), and TAG-FLOW (Thang et al., 2022). While these models provide notable computational efficiency and cost savings, their effective application depends on a thorough understanding of

the initial conditions. Furthermore, their underlying assumptions restrict generalizability across diverse slopes, necessitating case-specific solutions to complex governing equations.

Numerical analysis is time-consuming and costly, but it provides a more precise picture of the issue and can solve soil nonlinearity equations and complicated functions iteratively. Numerical analysis is facilitated through commercially available software such as GeoStudio (Calvello et al., 2009; Das et al., 2022; Huang & He, 2023; Pagano et al., 2010; Xu et al., 2022); MIDAS/GTS (Qiu et al., 2019; Yang et al., 2023); ABAQUS (He et al., 2021); GEOtop 2.0 (Formetta & Capparelli, 2019); Rocscience software (RS2) (Purnama et al., 2022); BGSlope (Abolmasov et al., 2015); SLIDE (He et al., 2016; Khan & Wang, 2021; Liao et al., 2010); and PLAXIS (Emmanuel et al., 2021; Oguz et al., 2022). Numerical models have several advantages compared with analytical models, which can be mentionedas follows: 1) These models can consider various triggering factors such as rainfall (Calvello et al., 2009; Das et al., 2022; Huang & He, 2023), reservoir level fluctuation (Xu et al., 2022), earthquakes (Yang et al., 2023), and energy provided by high waves (Hobbs et al., 2019) where these factors can be modeled together; 2) They include complex geometry such as deep soil depth, steep gradients, and material anisotropy (Liang & Uchida, 2022); 3) They include virtual monitoring points that replicate the real monitoring locations in the slope considering the three-dimensional impacts (Qiu et al., 2019); 4) They consider the delayed rainfall peak pattern effect (Liang & Uchida, 2022); 5) They can overcome the problem of unmonitored physical characteristics using back analysis of the available data (Abolmasov et al., 2015); 6) They can consider some asymmetric factors such as a surcharge (Sun et al., 2022); 7) They consider the nature of the bedrock layer, which can delay or decrease the failure time (Pagano et al., 2010).

Three categories of numerical analysis models are available: large deformation analysis, small deformation analysis, and non-deformable analysis. In other words, numerical modeling can be categorized as pre-failure analysis (limit equilibrium analysis and finite

element analysis) and post-failure analysis (large deformation analysis). Limit equilibrium analysis (LEA) can be used to study the non-deformable (uncoupled) analysis. The LEA methods are single free body methods (i.e., infinite slope, log spiral, and Swedish circle) or multiple free bodies (Fredlund & Krahn, 1977) (i.e., simplified Bishop's method (Bishop, 1955), Spencer's method (Spencer, 1967), Morgenstern and Price's method (Morgenstern & Price, 1965), Duncan's method (Duncan, 1996), Chen and Morgenstern's method (Chen & Morgenstern, 1983), and Sarma's method (Sarma, 1987)). LEA ignores soil deformation by assuming only seepage analysis and does not account for variations in vertical and shear stresses caused by pore water pressure variations (i.e., it considers only the weight of the slice) (Nguyen et al., 2019).

Coupled analysis (small deformation finite element analysis) considers both seepage and soil deformation over time (Matyas & Radhakrishna, 1968; Postill et al., 2021; Sitarenios et al., 2021; Song et al., 2020; Tang et al., 2019). However, this type of analysis is more complex and requires computational power compared with LEA. In large deformation, finite element analysis suffers from mesh distortion. The hydro-mechanical coupled process is seldom considered when forecasting landslides (Zhang et al., 2011). Some researchers have recently used large deformation (fully coupled) analysis (Cuomo et al., 2021; Lee et al., 2021; Liu et al., 2020). Large deformation analysis can be handled by several methods, including the MPM (Cuomo et al., 2021), Particle Finite Element Method, Finite Element Method with Lagrangian Integration Points, Smooth Particle Hydrodynamics, and Coupled Eulerian-Lagrangian methods (Chen et al., 2019). However, such models are hugely time-consuming and sometimes unfeasible as they are implemented in explicit schemes that necessitate small time steps (Yerro et al., 2022).

Some limitations of the current research can be illustrated as follows: It is seen that coupled analysis is not widely used, as the complexity of the landslide mechanism is simplified to be a

seepage problem, neglecting soil deformation, which may be misleading in terms of temporal and spatial prediction. The randomness of the rainfall pattern is not widely considered. Higher accuracy is needed in evaluating input geotechnical parameters, particularly regarding the influence of shear strength parameters (Bicocchi et al., 2019; Valentino et al., 2014). Furthermore, these models demand extensive geotechnical work. To illustrate, the prediction of landslides can be simplified to answer two questions: Where (local and regional)? And when? Several factors control the landslide mechanism, such as creep and periodic features. Among all the static and dynamic features that trigger landslides, the regional geometry that controls the radial subsurface flow is missing among these features. Some studies highlighted the topographic effect from a statistical and historical perspective (Gao et al., 2021). Recently, Song et al. (2020) adopted a vectorized finite element solver to account for the hydromechanical coupling effect on a regional scale. However, this study solves the water flow in one dimension, requiring more research regarding regional predictions. A series of parametric studies have been conducted through the literature to investigate the factors affecting rainfall-induced landslides, such as rainfall intensity, pattern, duration, and hydraulic, mechanical, and topographic characteristics (Oliveira et al., 2022; Ng & Shi, 1998; Ng et al., 2001; Rahimi et al., 2011; Sitarenios et al., 2021; Cheng et al., 2007; Feng et al., 2023). However, focusing more on the features that affect temporal prediction, such as the data frequency, may help researchers improve monitoring and data collection systems.

Widely, uncoupled analysis has been conducted (Ng & Shi, 1998; Ng et al., 2001; Feng et al., 2023; Oliveira et al., 2022). However, for landslides, the uncoupled analysis does not account for variations in stresses caused by pore water pressure variations (Cheng et al., 2007). Limited studies are available that consider the coupled hydro-mechanical analysis through seepage and slope deformation over time, since it complicates infiltration and seepage analysis (Postill et al., 2021; Sitarenios et al., 2021; Song et al., 2020; Tang et al., 2019). Nevertheless,

these models suffer from mesh distortion in large deformation problems such as landslides (Soga et al., 2016). Discrete Element Method, Smooth Particle Hydrodynamics, Particle Finite Element Method, Finite Element Method with Lagrangian Integration Points, and MPM are some methods that can handle large deformation analysis (Zhao et al., 2020; Chen et al., 2019). Several researchers have recently used large-deformation MPM analysis (Cuomo et al., 2021; Lee et al., 2021; Liu et al., 2020; Zhu et al., 2022). Although MPM can simulate the fully coupled hydro-mechanical behavior of the landslide process, it is time-consuming.

## 2.5. RESEARCH GAPS

Despite significant advancements in the field, several gaps remain (Table 2.9): there is limited comprehensive review of subsurface monitoring techniques and prediction models; rainfall-trigger randomness is reduced to cumulative totals or idealized patterns; physically based models focus on surface displacement and overlook internal processes; data-transfer during extreme weather leads to loss and ignores slope physics; intensity–duration (I–D) thresholds suffer high false-alarm rates by neglecting spatial variability; and stability-analysis workflows lack automation. The three-tiered framework is designed to directly address each of these shortcomings. The first tier employs a preliminary decision model, leveraging incident records and geospatial data, to rapidly classify zones as “failure not expected,” “warning signs,” or “failure expected,” thus identifying clear high- and low-risk cases. Ambiguous cases enter the second analytical layer, where a parallel, multi-node, multi-feature subsurface monitoring network, integrated via DL and validated against geotechnical benchmarks, captures spatiotemporal dependencies, triggering uncertainties, and mitigates data loss. In the third focused evaluation, deterministic numerical methods and an automated parameterization tool refine I–D thresholds by incorporating spatial heterogeneity, and execute end-to-end automated stability analyses.

*Table 2.9. Research gaps and suggested recommendations.*

| Gaps | Recommendation |
|---|---|
| The literature lacks a complete review of subsurface monitoring techniques and prediction models. | Recommended investigation of current prediction methodologies and monitoring systems. |
| Limited examination of the real random effects of rainfall patterns: current empirical and physically based models struggle to deliver large-scale operational warnings, partly because they reduce rainfall to cumulative or hypothetical patterns and ignore its intrinsic randomness | Develop multi-AI spatiotemporal models with extensive datasets to capture trigger randomness and spatial variability. |
| Physically based causative models rely on surface displacement and neglect internal processes; data-transfer studies overlook the geotechnical perspective; most existing models are site-specific, relying primarily on point-based monitoring data. | Integrate subsurface monitoring data into AI-driven frameworks to mimic the geotechnical physical response and robustly capture spatiotemporal variability with operationality into account. |
| Harsh environments cause data loss, and most studies use statistical models that ignore physical slope properties. | Link a parallel, multi-node, multi-feature spatially variable monitoring system via DL to mitigate data loss. |
| High false-prediction rates of intensity–duration (I-D) thresholds due to neglect of spatial features | Combine data-driven models with deterministic analysis to integrate spatial variability and reduce false alarms. |
| Automation of stability-analysis workflows has received little attention | Implement an end-to-end automated stability-analysis system for real-time monitoring and forecasting |

## 2.6. SUMMARY

Figure 2.25 visualizes the integration among several models, with their input monitored parameters and predicted model accuracy. It examines how one approach leads to another from a macroscopic perspective, depending on input parameters and initial conditions, to a microscopic illustration of different analysis models for the same method. It first highlights the national scale models, then presents four main prediction techniques for local-scale models: I-D thresholds (grey), API thresholds (yellow), physically based causative thresholds (orange), and deterministic models (blue). It is illustrated as follows: **a)** At the national scale, factors including geology, geomorphology, soil properties, and triggering elements such as rainfall facilitate the generation of landslide susceptibility maps. These maps form the foundation for risk assessments employing multivariate models, AI techniques, and hybrid approaches, enabling precise identification of high-risk areas and their potential impacts on infrastructure and human safety; and **b)** Transitioning to local scales, the framework incorporates more

detailed models accounting for complex geometrical, hydrological, and soil conditions. Deterministic physical models simulate intricate landslide processes using site-specific conditions, while statistical regression, AI, and probabilistic models complement these analyses by predicting initiation based on slope responses, including displacement, groundwater levels, and moisture content.

A particularly innovative aspect of Figure 2.25 lies in its synthesis of fragmented knowledge into an effective visualization that systematically illustrates the interconnections between different modeling approaches. Complementing this, the color-coded accuracy assessment (ranging from reasonable to perfect) provides immediate evaluation of model reliability based on specific input parameters, establishing a multi-tiered framework that enables robust predictive capabilities adaptable from national-scale susceptibility mapping to site-specific risk assessment. While this chapter provides a comprehensive synthesis, several limitations should be acknowledged. The reliance on major databases, specific keywords, and English-language publications may have omitted some relevant studies. Furthermore, the vast number of studies precluded a manual, quantitative meta-analysis of accuracy by model type. This remains a significant gap and a critical recommendation for future work.

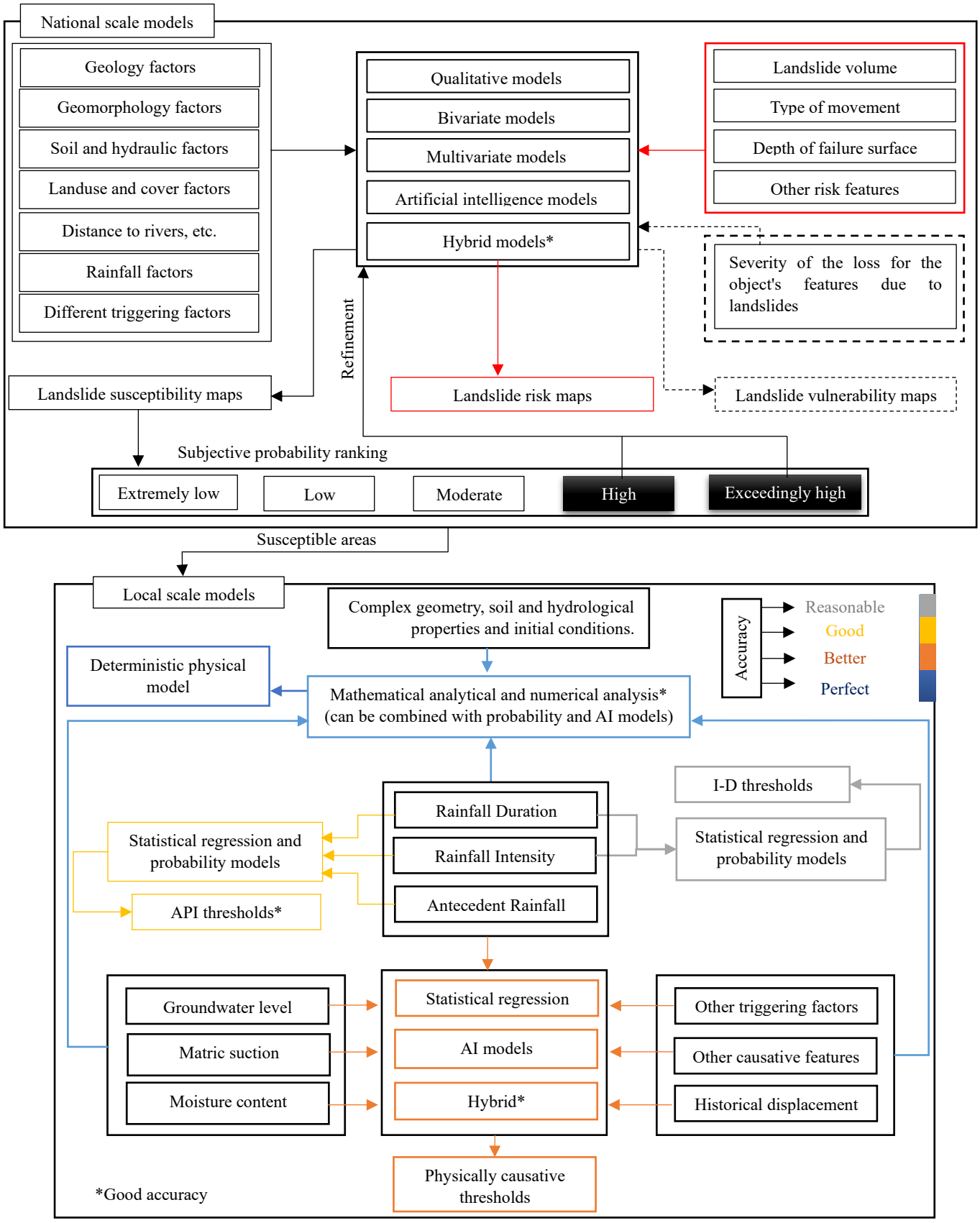



*Figure 2.25. Graphical summary for landslide prediction strategies*

# CHAPTER 3: METHODOLOGY

## 3.1. INTRODUCTION

This chapter describes the methodology used in this study. It begins with an overarching framework for achieving the research objectives and then details the specific methods and techniques applied, highlighting the study's novel contributions. The approach combines temporal DL predictors with high-fidelity numerical simulations and incorporates landslide subsurface monitoring data for dynamic model updating. This integrated, multi-model framework (see Figures 1.7 and 3.1) improves forecasting accuracy, supports operational scalability, and strengthens early-warning capabilities. The sections that follow explain how each component is implemented and how they interconnect to produce a multi-tiered AI-enhanced system.

## 3.2. LANDSLIDE PREDICTION MODEL (TIER 1)

### *3.2.1. Landslide Prediction Model Development*

#### a) Background and rationale

Existing landslide models often struggle with large-scale operational warnings because they oversimplify rainfall by treating it as a cumulative or hypothetical event, thereby ignoring its intrinsic randomness (Xiao et al., 2023; Qiu et al., 2024; Gao et al., 2018; Zhao et al., 2025). This limitation is compounded by the inherent spatial variability of rainfall effects, which differ across locations and create complex spatiotemporal interactions that are difficult to analyze. To address these challenges, a DL landslide-forecasting framework was developed. This approach is particularly well-suited because the sequential evolution of rainfall and its controlling features naturally lends itself to time series analysis, and DL models excel at capturing such dynamic dependencies. The framework was benchmarked against the LR model and validated against a deterministic slope-stability model. It utilizes a 1984–2022 Hong Kong dataset specifically designed to preserve rainfall randomness and spatiotemporal features, and it

integrates time-dependent modeling, tailored preprocessing, and a dual-threshold risk gradation strategy (implementation details are provided in Figure 3.2).

DL has emerged as a groundbreaking achievement in machine learning, capable of near-human or even super-human performance in diverse applications (Kiranyaz et al., 2021). For time-series tasks, Recurrent Neural Networks (RNNs) are a natural choice, but they suffer from vanishing and exploding gradients. While gradient clipping can mitigate exploding gradients, the problem of vanishing gradients persists. This led to the development of more advanced architectures like GRUs and LSTM networks (Cho et al., 2014; Hochreiter and Schmidhuber, 1997), which are designed to overcome these limitations. Alternatively, for detecting localized patterns in time-series data, such as short-term trends or spikes, a 1D CNN is highly effective, as it uses a sliding-window approach to capture nuanced variations without the need to maintain sequential hidden states (Kiranyaz et al., 2021).

The No Free Lunch theorem (Wolpert, 1996) reminds that no single algorithm is universally optimal, underscoring the empirical nature of DL where model architecture, input features, loss functions, and training strategies must be carefully optimized. Recent studies (Murray et al., 2023) reveal that while DL models each possess distinct strengths, their performance differences are often marginal. Consequently, this study introduces an interface that integrates a multi-model approach, incorporating LSTM, GRU, and CNN architectures to capitalize on the unique strengths of each model. At the same time, an LR model (Hosmer and Lemeshow, 2000) is employed as a benchmark, serving as a fundamental baseline for comparative analysis. The LSTM tackles vanishing gradients in long-term predictions by using memory cells equipped with three gates (input ($\boldsymbol{i}_t$), forget ($\boldsymbol{f}_t$), and output ($\boldsymbol{O}_t$)) to regulate data flow (refer to Equations 3.1 and 3.2) (Hochreiter and Schmidhuber, 1997). Each cell processes current input ($\boldsymbol{x}_t$), previous output ($\boldsymbol{h}_{t-1}$), and prior memory state ($\boldsymbol{C}_{t-1}$), as depicted in Figure 3.3.a.

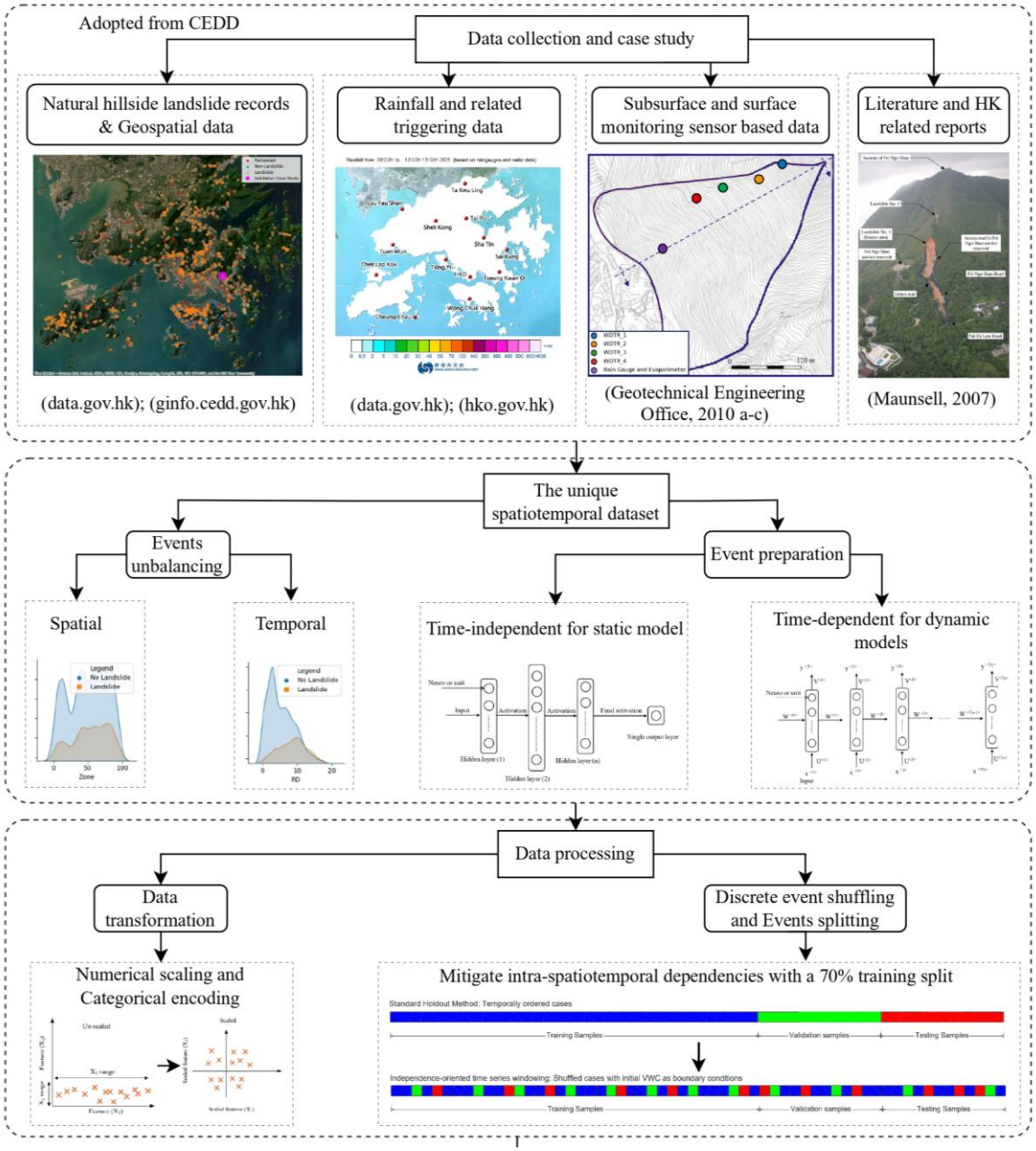



*Figure 3.1. Detailed methodology, deployments, and processing steps for all study objectives.*

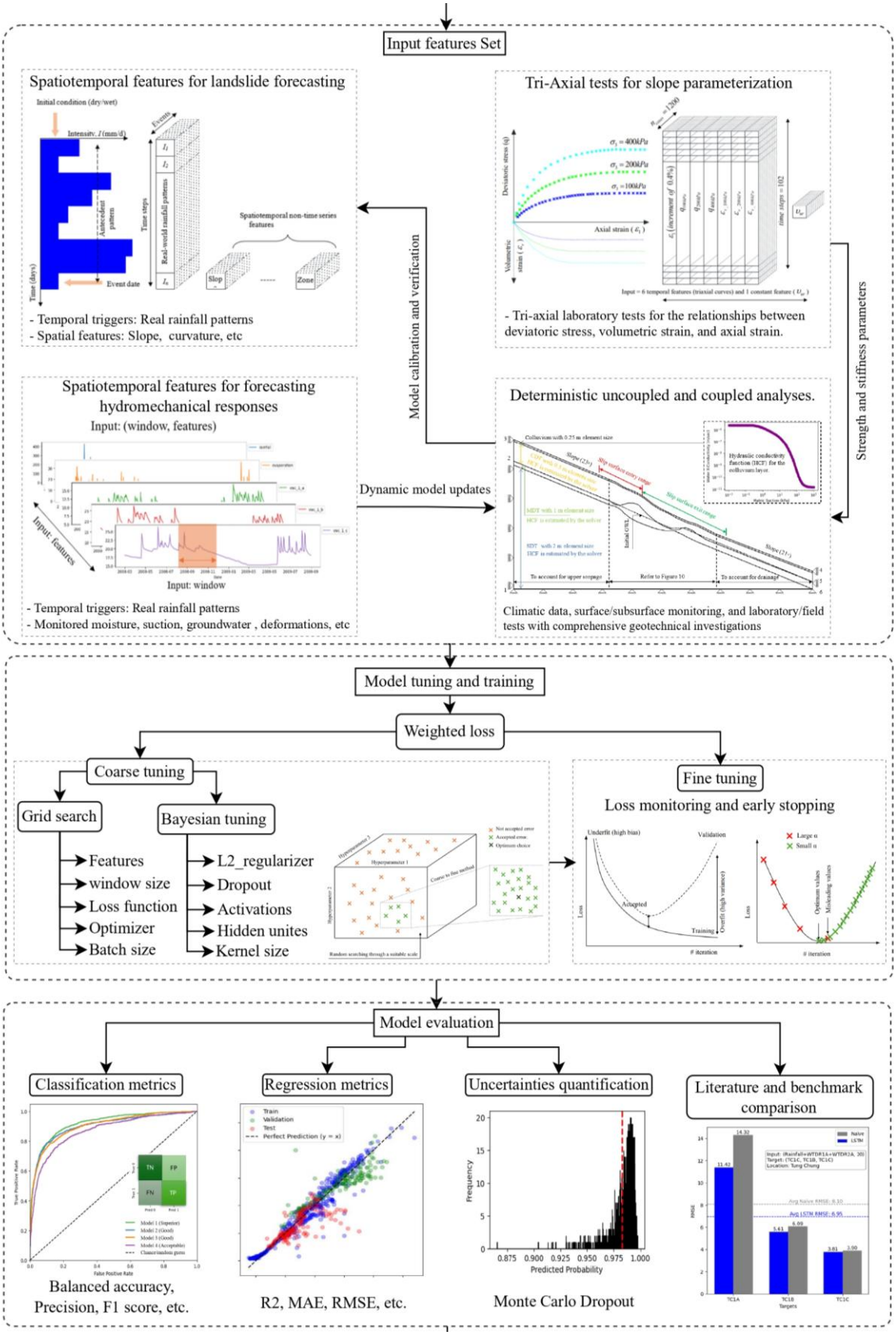



*Figure 3.1. Cont.*

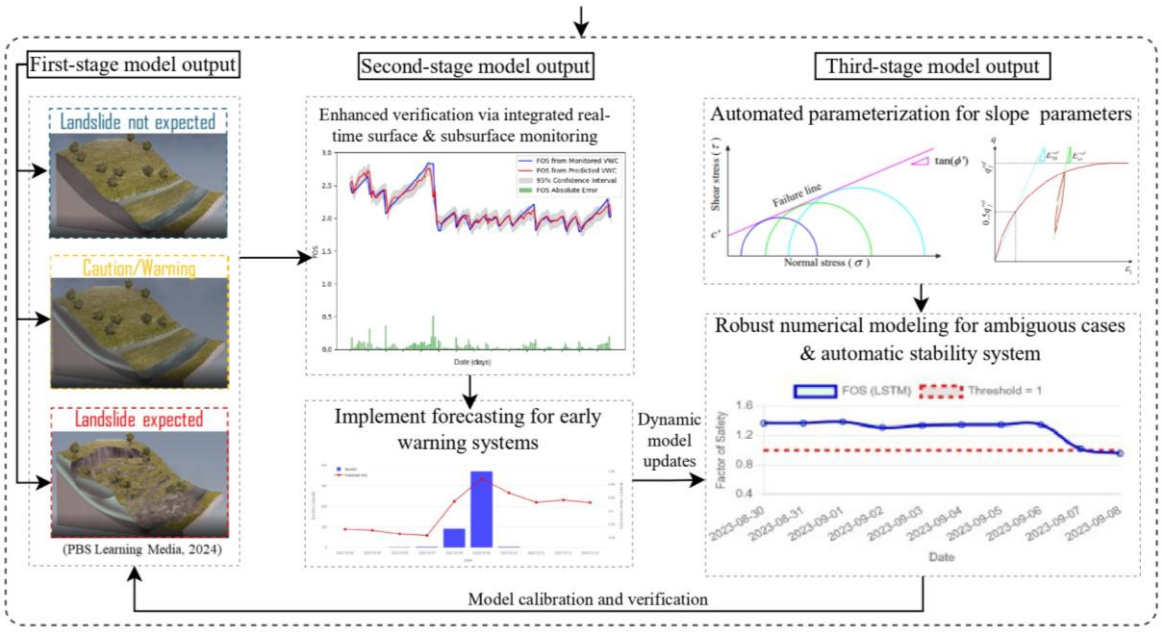



*Figure 3.1. Cont.*

The GRU is a streamlined version of RNNs that addresses vanishing gradient issues with fewer parameters (Cho et al., 2014). Merging the roles of the forget and input gates into a unified update gate ($\boldsymbol{Z}_t$) and introducing a reset gate ($\boldsymbol{R}_t$) to modulate the previous hidden state (refer to Equations 3.3 and 3.4) simplifies the architecture. Each cell processes the current input ($\boldsymbol{x}_t$) and previous hidden state ($\boldsymbol{h}_{t-1}$) through these gating mechanisms. Without maintaining a separate memory cell, the GRU directly computes the updated hidden state ($\boldsymbol{h}_t$) to efficiently incorporate both short- and long-term dependencies, as depicted in Figure 3.3.b. As for CNNs, the core operation involves applying a convolutional kernel over the input sequence to produce a feature map. The convolution operation at time step $t$ can be expressed as in Equation 3.5. This operation slides the filter across the input sequence, capturing localized features relevant to the task. The output feature maps are often downsampled using pooling layers, enhancing computational efficiency and providing a concise representation of the key temporal patterns in the data. This structure allows CNNs to process time-series data efficiently without relying on sequential dependencies like RNNs (Kiranyaz et al., 2021; Cacciari and Ranfagni, 2024). LR is a fundamental supervised learning algorithm for binary classification (Hosmer and

Lemeshow, 2000). It estimates the probability that a given input belongs to a particular class by applying the sigmoid function. The hypothesis function is expressed as in Equation 3.6.

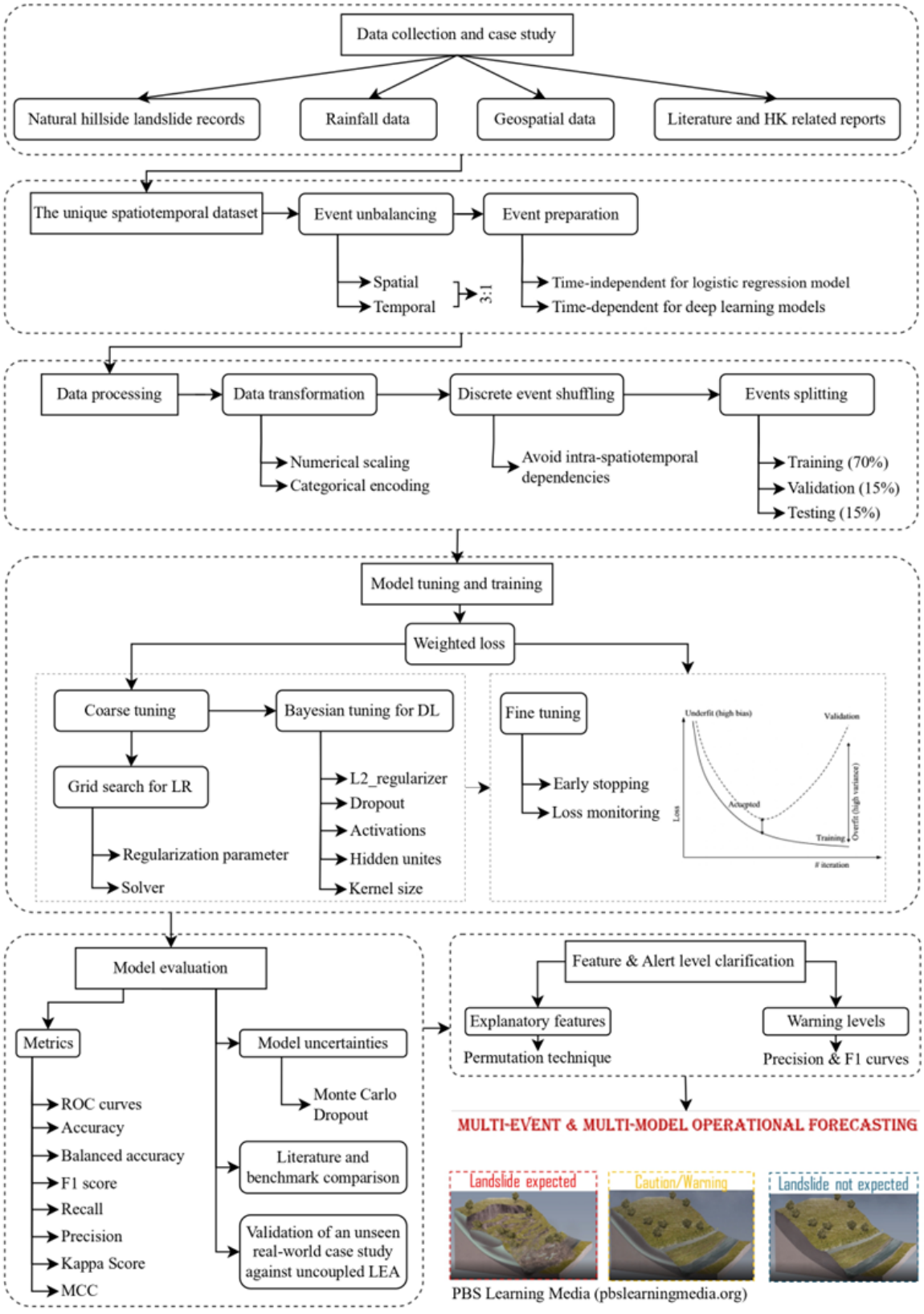



*Figure 3.2. Spatiotemporal modeling workflow*

$$\begin{cases} \boldsymbol{f}_t = \sigma(\boldsymbol{W}_f.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f) \\ \boldsymbol{i}_t = \sigma(\boldsymbol{W}_i.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i) \\ \boldsymbol{o}_t = \sigma(\boldsymbol{W}_o.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o) \end{cases} \quad \textit{(Equation 3.1)}$$

$$\begin{cases} \boldsymbol{C}_t = \boldsymbol{f}_t.\boldsymbol{C}_{t-1} + \boldsymbol{i}_t.\tilde{\boldsymbol{C}}_t \\ \tilde{\boldsymbol{C}}_t = \tanh\left(\boldsymbol{W}_c.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_c\right) \\ \boldsymbol{h}_t = \boldsymbol{O}_t.\tanh(\boldsymbol{C}_t) \end{cases} \quad \textit{(Equation 3.2)}$$

$$\begin{cases} \boldsymbol{Z}_t = \sigma(\boldsymbol{W}_Z.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_Z) \\ \boldsymbol{R}_t = \sigma(\boldsymbol{W}_R.[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_R) \end{cases} \quad \textit{(Equation 3.3)}$$

$$\begin{cases} \tilde{\boldsymbol{h}}_t = \tanh\left(\boldsymbol{W}_h.\boldsymbol{x}_t + \boldsymbol{U}_h\left(\boldsymbol{R}_t.\boldsymbol{h}_{t-1}\right) + \boldsymbol{b}_h\right) \\ \boldsymbol{h}_t = \boldsymbol{Z}_t.\boldsymbol{h}_{t-1} + \left(1 - \boldsymbol{Z}_t\right).\tilde{\boldsymbol{h}}_t \end{cases} \quad \textit{(Equation 3.4)}$$

$$\boldsymbol{z}_t^{(i)} = \sigma\left(\sum_{j=1}^{k} \boldsymbol{W}_j^{(i)}.\boldsymbol{x}_{t+j-1} + \boldsymbol{b}^{(i)}\right) \quad \textit{(Equation 3.5)}$$

$$\mathrm{f}\left(x\right) = \sigma\left(\boldsymbol{w}.\boldsymbol{x} + b\right) = \frac{1}{1 + \mathrm{e}^{-\left(\boldsymbol{w}.\boldsymbol{x}+b\right)}} \quad \textit{(Equation 3.6)}$$

σ is the activation function, and $\boldsymbol{W}$ and $\boldsymbol{b}$ represent the weight and bias, respectively. Once the output is computed, Equation 3.2 is used to obtain both the short-term memory state ($\boldsymbol{h}_t$) and the long-term memory state ($\boldsymbol{C}_t$). $\boldsymbol{W}$ and $\boldsymbol{U}$ are weight matrices associated with the input and previous hidden state, respectively. $\tilde{\boldsymbol{h}}_t$ is the candidate activation. $\boldsymbol{z}_t$ represents the feature map output for the $i^{th}$ filter, $\boldsymbol{W}_j$ is the $j^{th}$ weight of the $i^{th}$ filter, $\boldsymbol{x}$ is the input at position (t+j-1), $\boldsymbol{b}$ is the bias, and σ is the activation function.

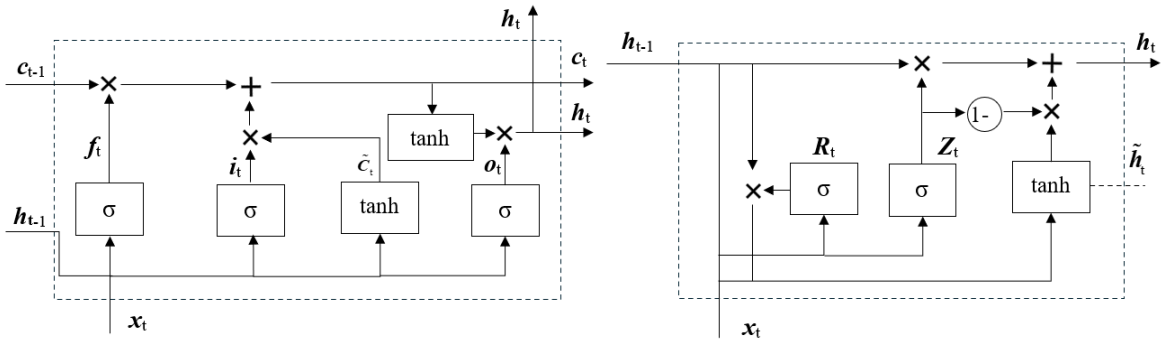



*Figure 3.3. a) LSTM cell structure* *b) GRU cell structure*

To validate the data-driven forecasting approach, it was compared with deterministic models. Specifically, a simplified Bishop method combined with a modified Mohr-Coulomb failure criterion was employed to incorporate shear strength variations due to matric suction in limit equilibrium analysis (Equation 3.7 (Fredlund et al., 1978)). The governing differential equation for saturated and unsaturated water flow in a two-dimensional soil is presented in Equation 3.8 (Lam et al., 1987): it relates the flow behavior to VWC and hydraulic conductivity functions (Fredlund and Xing, 1994). Specifically, variations in VWC lead to corresponding changes in pore water pressure. For the transient seepage analysis, it is assumed that there is no hysteresis between the drying and wetting paths of the soil-water retention curve.

$$\tau = c' + \sigma_{net} \tan\phi' + \sigma_s \tan\phi^b \qquad \textit{(Equation 3.7)}$$

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(k_y \frac{\partial h}{\partial y}\right) + q = m_v \left(\frac{\partial u_w}{\partial t}\right) \qquad \textit{(Equation 3.8)}$$

$k_x$ and $k_y$ denote hydraulic conductivity in the $x$ and $y$ directions. $q$ is the applied boundary flux. The total head is $h$, and $m_v$ is the slope of the water retention curve. $\tau$, $c'$, $\phi'$, $\sigma_{net}$, and $\sigma_s$ represent the shear strength, effective cohesion, effective friction angle, net normal strength, and matric suction, respectively. The angle that links the rate at which shear strength increases as matric suction increases is $\phi^b$.

### b) Data preparation

The development of the multi-model framework required careful data preprocessing, as outlined in Figure 3.2. Key steps included: (a) standardizing input parameters (e.g., dry patterns

versus high-intensity rainfall) to ensure consistent training scales; (b) randomizing data sequences within events to eliminate intra-event dependencies while preserving inter-event spatiotemporal relationships; (c) addressing class imbalance by implementing weighted loss functions to emphasize minority landslide cases; and (d) employing evaluation metrics beyond accuracy to reliably assess model performance given the imbalanced data distribution. The following subsections detail these critical development stages. It begins by standardizing input features (x) and target variables (y) to reduce sensitivity to varying scales. Numerical features are normalized to achieve a zero mean and unit variance (Equation 3.9), while categorical features are transformed into binary vectors via one-hot encoding (Pedregosa et al., 2011). For DL models, features are reshaped into a three-dimensional format (Equation 3.10) to effectively capture sequential and temporal dependencies. In contrast, LR, unable to inherently process time-series data, relies on engineered features that treat each case as temporally independent. To minimize bias from ordered data, case indices are shuffled, and the dataset is divided into training, validation, and testing subsets in a 70:15:15 ratio to ensure representativeness.

$$x_{(i,j)} = \left(x_{(i,j)} - \mu_j\right) / \left(\sigma_j\right) \qquad \textit{(Equation 3.9)}$$

$$\boldsymbol{x} = \text{Reshape}(n_i, t, n_f) \qquad \textit{(Equation 3.10)}$$

Here, $\mu_j$ is the mean value of feature j, $\sigma_j$ is the standard deviation, $t$ is the time steps, $n_f$ is the number of input features, and $n_i$ is the number of events.

### c) Explanatory features

Recognizing that similar rainfall patterns may exert varying impacts, this study aims to enhance model precision by integrating additional spatiotemporal features. These include: (a) the initial boundary condition for each triggering event ("Initial Month"); (b) the spatial susceptibility for radial flow accumulation ("Curvature"); (c) the spatial variability of soil and vegetation ("Zone," "Material Description," and "Nearby Structures"); (d) the spatial uncertainty of rainfall ("Zone Distance"); and (e) topographic characteristics ("Slope" and "Elevation"). For

DL models, rainfall data is treated as a time series, allowing the model to extract relevant features autonomously. In contrast, an LR benchmark was developed using engineered features. Each event is treated as an independent sample, with temporal dependencies maintained solely within each sample and not across different samples. Recognizing that DL inherently operates as a "black box", a sensitivity analysis comparing various combinations of features was implemented. Furthermore, a permutation technique is employed to assess feature importance quantitatively. This method involves permuting each feature's values and observing the change in model error (Fisher et al., 2019). The impact of each feature is effectively quantified by comparing the baseline error with the error after each permutation, measured using accuracy.

### d) Model structure, training, and optimization

Figure 3.4 illustrates a schematic representation of the DL model architecture. Temporal data are sampled using fixed antecedent windows for all instances, which are sensitively selected. The subsequent sections delineate the model's design, training, and optimization procedures: LSTM and GRU inherently capture temporal dependencies through their recurrent connections and gating mechanisms. In contrast, CNN models employ a convolutional layer, using a tuned kernel size, to effectively extract local sequential patterns. At the same time, the application of "same" padding preserves the input's original temporal dimensions and boundary information. This is followed by a GlobalMaxPooling1D layer that aggregates the feature maps into a fixed-length vector by selecting the most dominant activations across time (Pedregosa et al., 2011). The architecture was intensified with auxiliary layers to further enhance the DL models' capacity for capturing non-linear relationships. Specifically, each hidden layer employed the Leaky-ReLU activation function, which mitigates the "dying ReLU" problem by allowing a slight, non-zero gradient for negative inputs (Qi et al., 2024). Moreover, to reduce overfitting, dropout and L2 regularization techniques were integrated. Dropout prevents neurons from co-

adapting by randomly deactivating a subset of nodes during training, while L2 regularization penalizes large weight values, thereby constraining the model complexity (Srivastava et al., 2014). The binary cross-entropy loss function was utilized, while model training was conducted using the Adam optimizer due to its adaptability and efficiency. An early stopping criterion was employed to capture the optimal weights and mitigate overfitting or underfitting. Learning curves were scrutinized to assess model performance during the training and validation. The DL models' hyperparameters were optimized using Bayesian optimization (Snoek et al., 2012). An LR model was implemented as a benchmark classifier, incorporating seven non-time-series features, with temporal sensitivity assessed over cumulative windows of 1–4 weeks. L2 regularization was applied, with the regularization strength sampled from 0.01 to 10.01. Tuning was performed via a randomized grid search over 3000 iterations using 5-fold cross-validation, evaluating solvers ('saga', 'lbfgs', 'liblinear') with the F1 score as the selection metric. Class imbalance was addressed using balanced loss weights, and solver convergence was ensured by capping iterations at 5000. Table 3.1 presents the DL hyperparameter search space, while Table 3.2 details LR.

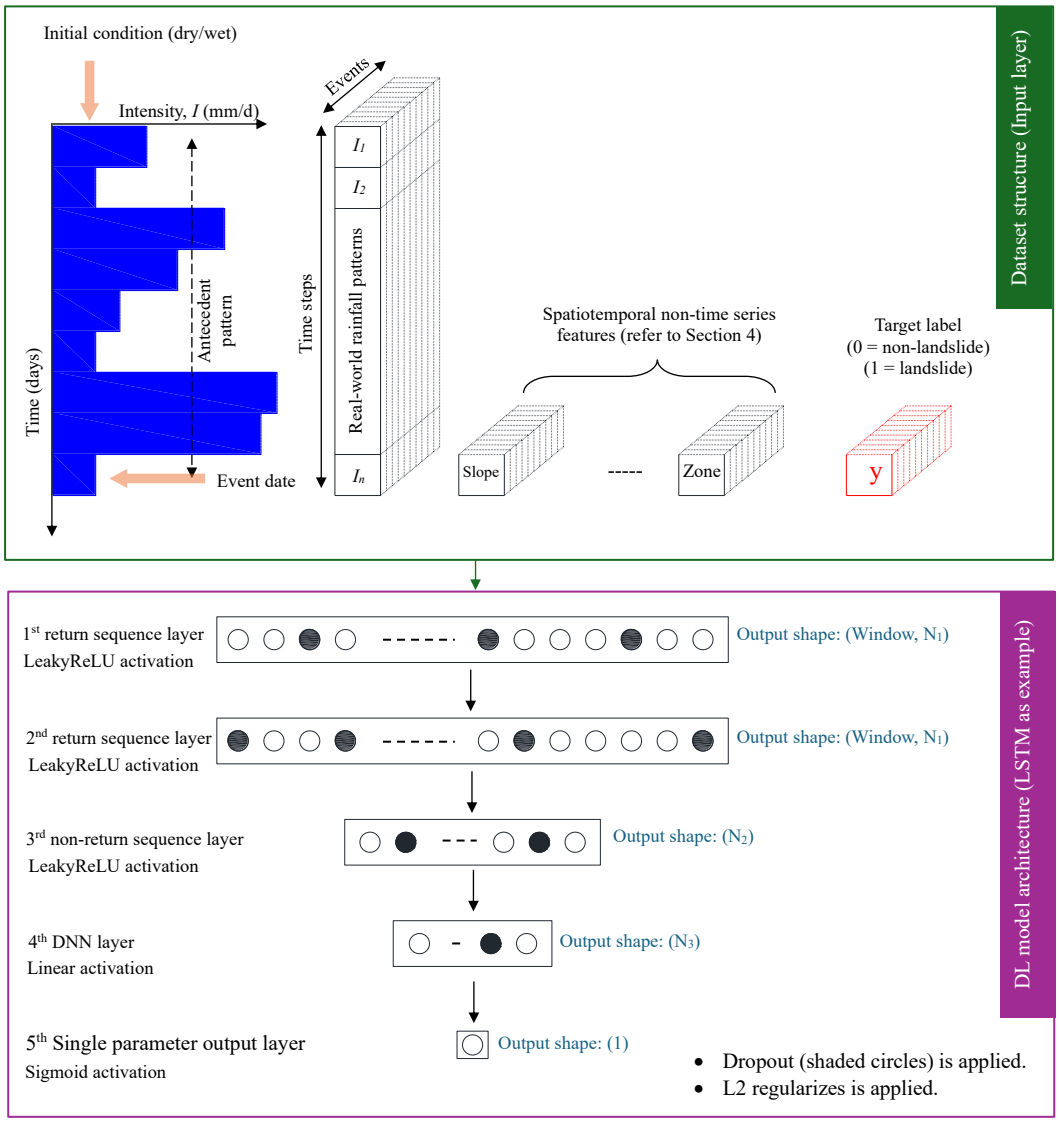



*Figure 3.4. Schematic of the spatiotemporal DL model architecture (LSTM as an example).*

*Table 3.1. DL models design parameters and search space.*

| Parameter | Search space/value | |
|---|---|---|
| | LSTM and GRU | CNN |
| Features | (1) time series features (rainfall) and eight non-time series features | |
| Window size | Sensitivity analysis was conducted across antecedent windows of 1 to 4 weeks. | |
| Kernel size | - | (K: 3, 5) |
| [Hidden layers] (Hidden units/filters) | [2] return sequence ($N_1$: 256, 512, 1024) | [2] Conv 1D ($N_1$: 256, 512, 1024) |
| | [1] non-return sequence ($N_2$: 128, 256, 512) | [1] Conv 1D ($N_2$: 128, 256, 512) |
| | [1] deep neural network ($N_3$: 64, 128, 256) | |
| Loss function | Binary cross-entropy $= -(1/n)\sum_{i=1}^{n}\left[y_i \log(\tilde{y}_i) + (1-y_i)\log(1-\tilde{y}_i)\right]$<br>Here, $y_i$ is the true label, $\tilde{y}_i$ is the predicted probability, and $n$ is the number of events. | |
| Loss weights | 1:2 ratio (non-landslide to landslide events, respectively) | |
| Optimizer | Adam | |
| Dropout/LeakyReLU | (D: 0.1, 0.3, 0.5, 0.7, 0.9)/(LR: 0.1, 0.3, 0.5, 0.7, 0.9) | |
| L2_regularizers | (L2: 0.0001, 0.001, 0.01, 0.1) | |

*Table 3.2. LR model design parameter and search space.*

| Parameter | Search space/value |
|---|---|
| Features | (7) engineered non-time series features |
| Window size (antecedent period) | Sensitivity analysis using cumulative windows of 1, 2, 3, and 4 weeks. |
| Regularization strength (C) | 0.01 to 10.01 |
| Optimization algorithm | 'saga', 'lbfgs', and 'liblinear' |
| Penalty | L2 |
| Grid search iteration | 3000 |
| Cross-validation | 5-fold |
| Scoring | F1 |
| Loss weights | Balanced |
| Solver iterations | 5000 |

### e) Model development and deployment

The system was implemented using TensorFlow, a Python-based framework for developing high-level neural networks (Abadi et al., 2015). Keras Tuner efficiently handled hyperparameter optimization (O'Malley et al., 2019). The network architecture was constructed with the Sequential API, allowing the integration of layers with customizable settings. Data manipulation and visualization were supported by several Python libraries: Pandas was used for data handling (Reback et al., 2020), NumPy for numerical computation (Harris et al., 2020), Matplotlib (Hunter, 2007), and Seaborn (Waskom, 2021) for graphical outputs, and scikit-learn for preprocessing tasks (Pedregosa et al., 2011). Geospatial mapping was enabled using GeoPandas (GeoPandas, 2023) and Shapely (Shapely, 2023), while Contextily (Contextily, 2023) supported detailed visualizations. The interface is developed using the Streamlit framework in Python (Khorasani et al., 2022). Model training was performed on the High-Performance Computing (HPC) platform, with a Jupyter Notebook environment. The hardware configuration included an NVIDIA RTX 3080 Ti GPU, 64 GB of RAM, six CPU cores, and one GPU card, running on Anaconda (2023-03-01) with Python 3.10.9 (The Hong Kong Polytechnic University HPC, 2025).

f) Model evaluation

Accuracy can be misleading in imbalanced datasets; thus, scikit-learn's suite of evaluation metrics, such as recall, F1 score, precision, ROC curve statistics, Cohen's kappa, Matthews correlation coefficient, and balanced accuracy, was leveraged to provide a detailed performance analysis of the classifiers (Pedregosa et al., 2011). The F1 score, in particular, was emphasized for its ability to effectively balance false positives and false negatives. Precision, which evaluates the correctness of predicted landslide occurrences, reflects the reliability of positive predictions. Additionally, uncertainty quantification was emphasized to assess model reliability for risk-sensitive applications. Detailed illustrations for the model evaluation metrics are presented in Appendix A (see Tables A1-2 and Figure A2)

Monte Carlo (MC) dropout was incorporated to determine predictive uncertainty (Gal & Ghahramani, 2016). Instead of deactivating dropout during inference, the dropout layers are maintained, performing T = 1000 stochastic forward passes for each input $x$. This strategy generates a set of predictions $\left\{f^{(t)}(\boldsymbol{x})\right\}_{t=1}^{T}$, the final predicted probability being determined by averaging these outputs. The model's predictive uncertainty is then approximated by calculating the standard deviation (Equation 3.11).

$$\sigma(\boldsymbol{x})=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(f^{(t)}(\boldsymbol{x})-f^{'}(\boldsymbol{x})\right)^{2}} \qquad \textit{(Equation 3.11)}$$

Here, $f^{(t)}(\boldsymbol{x})$ denotes the prediction during the $t^{\text{th}}$ stochastic forward pass, and $f^{'}(\boldsymbol{x})$ is the mean prediction across all passes.

*3.2.2. Landslide Prediction Model Assumptions*

Hong Kong provides an ideal, data-rich testbed for transferable landslide forecasting methods. About 60 % of its terrain is steep natural slope, mantled by 1–36 m of colluvium or completely decomposed granitic, volcanic, and sedimentary rocks. These gravelly, silty-clayey sands are highly susceptible to rainfall-induced failure. Coupled with intense monsoon storms, a four-

decade landslide inventory, and rapid urban expansion, this setting offers the complexity and documentation needed to develop models readily adapted to other mountainous regions worldwide (Ng et al., 2001; Ng and Shi, 1998; Xiao et al., 2023; Maunsell Geotechnical Services Ltd., 2007; Hencher and Lee, 2010). Environmentally, due to high humidity, rainfall can be effectively simplified into infiltration and runoff (Liu et al., 2022; Premchitt et al., 1992). Regarding hydraulic and mechanical properties, these soils exhibit saturated hydraulic conductivity values ranging from $5\times10^{-6}$ to $3.5\times10^{-4}$ m/s (GCO, 1982), cohesion, and internal friction angles with mean values of 5 kPa and 36°, respectively (Wong, 2020).

Vegetation influences slope stability through two primary mechanisms. The hydrological effect involves rainfall interception and increased surface runoff, which reduces the amount of water that infiltrates the soil. Premchitt et al. (1992) reported that runoff can range from 0% to 100% of rainfall depending on topographical and hydrological factors, including vegetation cover, slope geometry, soil properties, and rainfall characteristics. The mechanical effect involves root reinforcement that increases soil shear strength, primarily enhancing cohesion while leaving the friction angle largely unaffected (Salvatici et al., 2018) - Equations A1–A3 (Appendix A). However, as this study primarily employs data-driven models requiring extensive training datasets, reconstructing high-resolution temporal variations in land cover for Hong Kong's natural slopes, including bare soil, forest, shrub, grass, and engineered surfaces such as Chunam and Shotcrete, is challenging due to limited publicly available data.

With that in mind, this study primarily considers the inherent randomness of rainfall as the key landslide-triggering factor, with spatial features, particularly topographic characteristics affecting infiltration, integrated into the analysis. To address the challenge of spatial variability in soil and vegetation, a zonal framework was adopted. Incidents were grouped by location (i.e., Zone) and soil type into designated zones, within which soil and vegetation conditions are treated as spatially uniform. This simplification is necessary for a

regional-scale analysis spanning several decades. Furthermore, due to the difficulty of reconstructing high-resolution temporal variations in land cover (e.g., bare soil, forest, shrub, grass, and engineered surfaces like Chunam and Shotcrete) for each historical incident, and given its relatively minor influence on landslide initiation in Hong Kong (Qiu et al., 2024), vegetation was assumed to be temporally uniform. To further eliminate the dynamic effects of the vegetations or human activities, non-landslide events are identified as the two antecedent events preceding each recorded landslide. This creates a temporal sequence of three consecutive events: two non-landslide events followed by one landslide event. This short-term window mitigates the confounding effects of long-term temporal changes, such as vegetation dynamics or human activities. The complete dataset is detailed in Section 4.

## 3.3. SUBSURFACE SENSOR-BASED PREDICTION MODEL (TIER 2).

### *3.3.1. VWC Prediction Model Development*

#### a) Background and rationale

Rainfall-induced landslides result from complex hydrological and geotechnical interactions, with one of the key challenges being the accurate estimation of infiltrated rainfall. Uncoupled deterministic analyses have long been used for seepage modeling (Ng & Shi, 1998; Ng et al., 2001; Natalia & Yang, 2024), treating water infiltration as a time-dependent process governed by hydraulic conductivity, surface flux, and storage capacity. Although these methods provide precise analytical insights, their extensive data requirements and limited early warning capabilities call for alternative approaches. DL has emerged as a groundbreaking achievement in the machine-learning era, delivering near-human performance across various applications (Kiranyaz et al., 2021), as detailed in Table A3. However, while effective for site-specific predictions, the operational-scale application of DL remains largely unexplored. Furthermore, previous studies typically evaluated rainfall events individually or cumulatively using simplified temporal patterns (e.g., Rahimi et al., 2011), which may not capture the inherent

randomness of real-world conditions. These gaps underscore the necessity for a unique methodology that integrates spatiotemporal variability and rainfall stochasticity in a user-friendly and cost-effective manner, a challenge that lies at the core of this study.

With that in mind, this study introduces a unique operational model for predicting shallow VWC, a critical parameter for assessing rainwater infiltration. Using Hong Kong as a case study, the proposed approach overcomes spatiotemporal limitations in existing predictive models by accounting for the randomness of rainfall-triggering mechanisms. The methodology integrates a unique data preparation technique, independence-oriented time series windowing, with the predictive power of DL, specifically, LSTM networks, and deterministic seepage modeling via GeoStudio SEEP/W. The model is developed using field data from 15 sensors across three sites in Hong Kong (Pa Mei, Tung Chung, and Tsing Shan) at two depths (0.5 m and 1.5 m), complemented by a numerically validated case at Fei Ngo Shan Reservoir. 33 field samples were collected from eleven different locations to validate initial hypotheses and boundary assumptions. This study is among the first systematic evaluations comparing the accuracy of operational versus site-specific models. The outcome is OWHK (Operational VWC Forecasting Model for Shallow Layers in Hong Kong). The modeling framework is presented in Figure 3.5 and detailed as follows:

b) DL model structure

Figure 3.6 illustrates a single model architecture, delineating both sequential (LSTM) and non-sequential layers with their respective search space parameters. The sequence layers capture the temporal dependencies in the sequential data. The subsequent non-sequence layer compacts the representation of the features to process incoming signals efficiently. To further enhance the model's capacity to capture complex nonlinear relationships, auxiliary layers were incorporated. Leaky ReLU activation was used to mitigate the dying ReLU issue by preserving gradient flow for negative inputs. Dropout layers (denoted by shaded circles, D) are

incorporated to regularize the LSTM network and reduce overfitting. The model is trained using the Mean Squared Error (MSE) loss and optimized with the Adam optimizer, while an Early Stopping criterion is employed to secure the optimal weights. To address scaling discrepancies, normalization improves a model's predictive performance by standardizing input features to a consistent range. This process enhances convergence during training and reduces sensitivity to feature magnitude. In this study, Min-Max normalization was applied to rescale the data within predefined bounds, ensuring it aligns with the Leaky ReLU activation function used in the model. This approach suits the dataset's positive values, promoting faster convergence and better predictive accuracy.

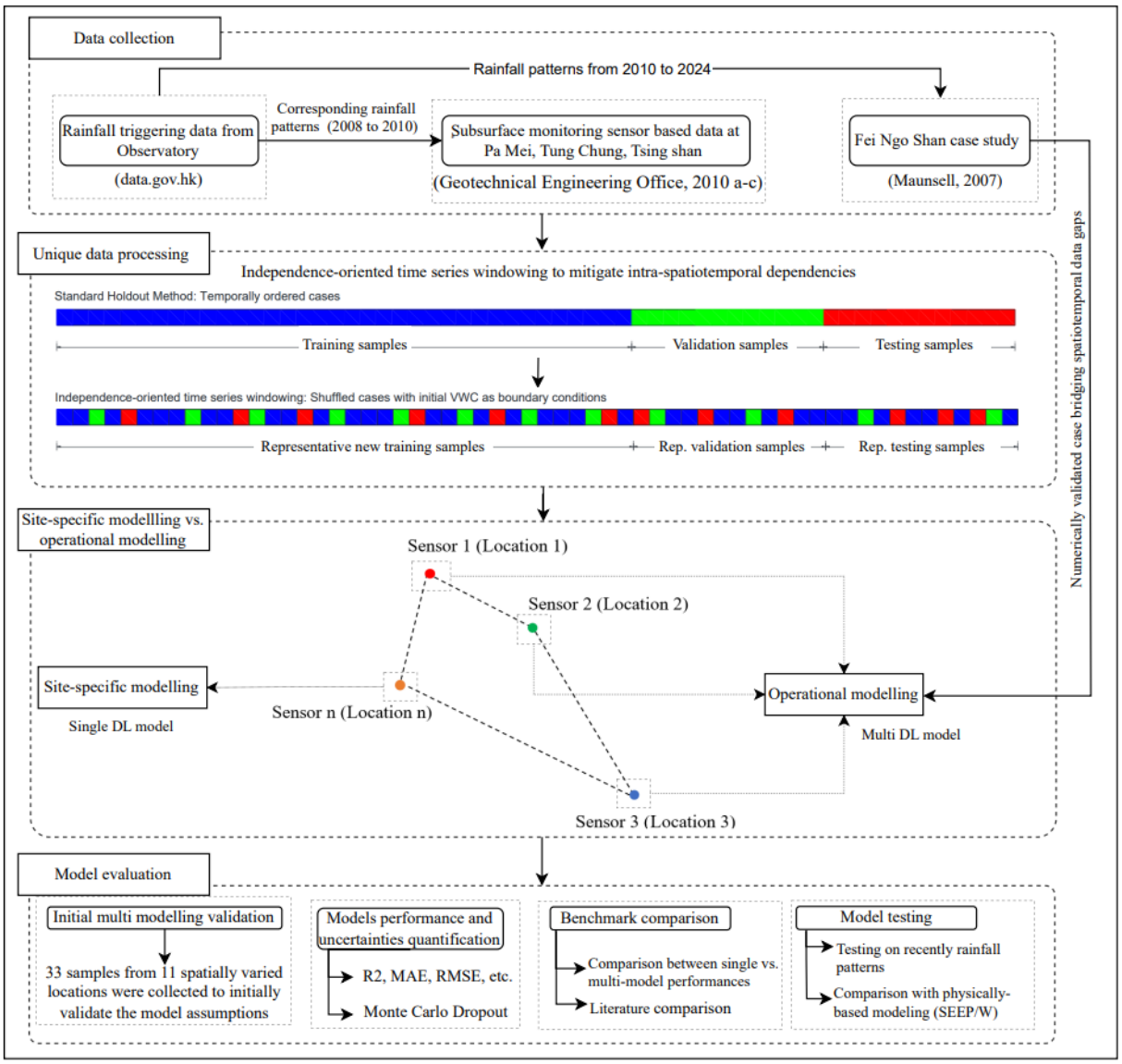



*Figure 3.5.Subsurface VWC operational modeling workflow*

Acknowledging that DL is inherently an empirical process that demands meticulous optimization and tuning. Traditional coarse tuning methods, such as grid search or random search, are widely used but can be computationally intensive and may not always yield the optimal solution. In contrast, Bayesian optimization (Snoek et al., 2012), employed in this study, efficiently investigates the sensitivity of diverse alternatives for model performance. However, it carries a risk of converging to local minima. In response, a two-stage coarse-to-fine-tuning approach was adopted. First, a coarse tuning phase is applied to identify promising model parameter ranges broadly. The parameters then undergo a second fine-tuning phase to refine the model settings. In addition, this study uniquely implements a tailored data handling strategy during fine-tuning to mitigate potential misleading metric outcomes.

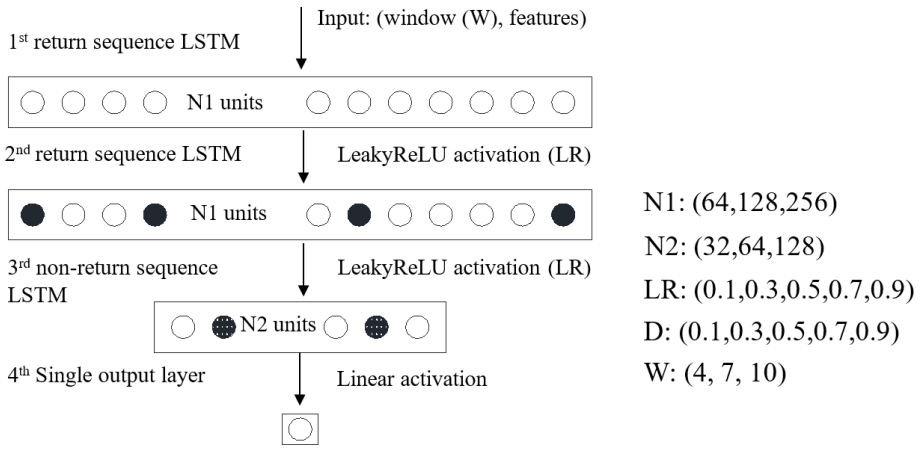



*Figure 3.6. Simplified operational VWC LSTM architecture with the parameter search space.*

### c) Data preparation and temporal ordering

Mostly, the time series problem is handled by keeping the temporal ordering, where the training set is chosen from the earlier time step and the validation and testing sets from the subsequent time steps. This methodology assumes that the response is consistent. However, this assumption is not generally valid when considering the complexity of landslide moisture dynamic responses. To support this, Figure 3.7 presents the VWC response to rainfall: blue indicates training samples, green denotes validation samples, and red boxes represent testing samples arranged via a conventional holdout method (i.e., temporally ordered). As illustrated

in Figure 3.7.a, this approach may fail to capture all relevant scenarios. For instance, if the testing set comprises only a dry period, the evaluation may be misleading. This could occur either because the model is under-trained for such conditions, resulting in high error rates, or because the intrinsically low VWC values during dry periods lead to deceptively low errors.

This study uses "Independence-oriented time series windowing" to ensure the training, validation, and testing sets represent the entire time series, accounting for temporal changes Figure 3.7.b (i.e., not temporally ordered). The dataset is divided into subsets using a sliding technique (Hindriks et al., 2016). Each subset is defined by the slice size (minimum time step of 1 day) and the window size, which accounts for temporal dependencies (antecedent effect). Each subset is considered an independent sample, where temporal dependencies are present only within the subset itself. There are no temporal dependencies between different subsets. Each subset is defined by the initial VWC as a boundary condition, with the target variable being the VWC at the end of the window size, given the rainfall characteristics as input features.

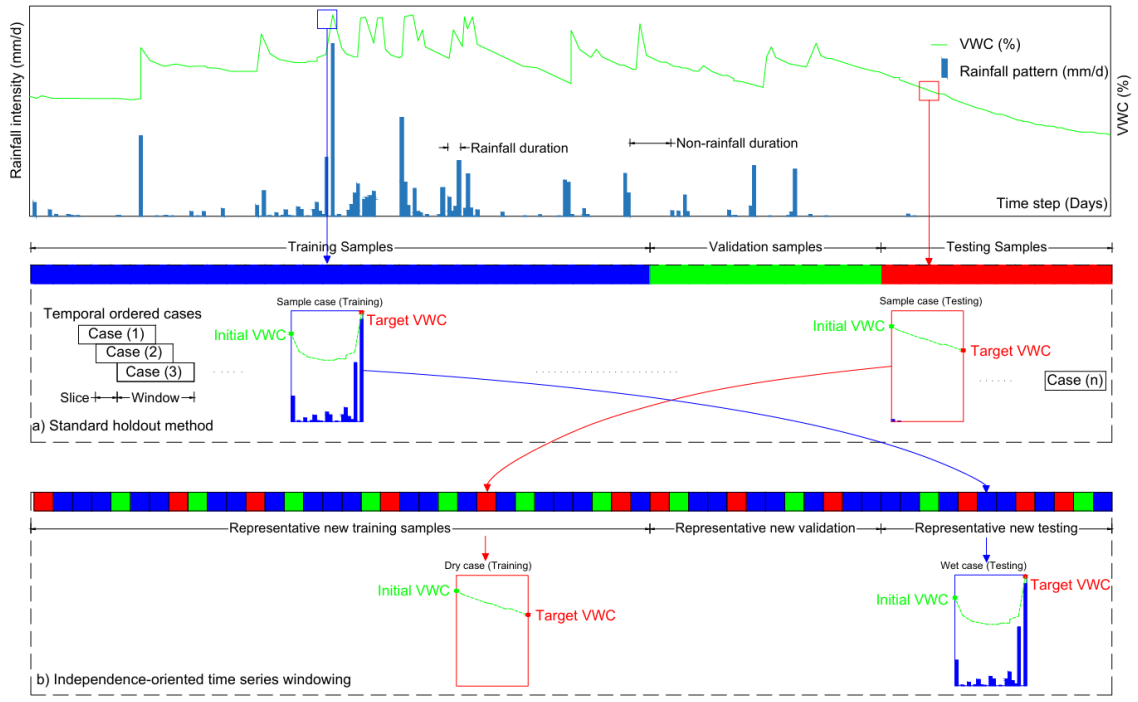



*Figure 3.7. Data sampling approach using independence-oriented time series windowing.*

### d) Site-specific models vs. operational models

This study aims to develop an operational model to predict VWC. However, firstly, a temporal challenge lies in ensuring the model accounts for the factors influencing the interaction between rainfall and VWC. This interaction can be described through infiltration, storage, and seepage processes. This process can be theoretically understood through the concepts of saturated and unsaturated water flow, as illustrated in Equation 3.8, assuming two-dimensional flow. The differential equation for water flow in unsaturated soil is derived by substituting ( $\partial\theta = m_v \partial u_w$ ) in Equation 3.8, where changes in pore water pressure are attributed to changes in VWC (Lam et al., 1987). For a site-specific model, the training data will be derived from sensor readings that capture unique soil characteristics. In other words, sensor data capture the local soil properties for a site-specific model, rendering spatial variability negligible and making temporal fluctuations the primary determinant of model behavior. At this stage, the LSTM model can indirectly capture the dynamic temporal responses by identifying key parameters such as rainfall features (intensity, duration, pattern, and non-rainfall periods) and the window size representing antecedent conditions. However, the model will still encounter challenges due to spatiotemporal variability. These dependencies arise from: a) potential unseen rainfall patterns (i.e., temporal) and b) varying material responses across different locations (i.e., spatial).

To minimize uncertainties arising from the temporal dependencies, this study employs the following techniques: a) shuffling field-monitored subsets to ensure the model is reliably trained for all possible real-world rainfall patterns within the range observed in the dataset (as illustrated by "Independence-oriented time series windowing"); b) incorporating numerically simulated real rainfall patterns to overcome the limitations of training the model on specific time-space data. To emphasize, standard LSTM implementations often employ a uniform weighting in their training losses, which may overlook the site-specific responses inherent to

different locations. This limitation highlights the need to extend the temporal modeling capabilities of LSTM to incorporate spatial heterogeneity. Accordingly, a unique multi-LSTM framework is proposed in Figure 3.8, wherein data subsets from various locations, each representing distinct soils and boundary conditions, are grouped. Figure 3.8 illustrates the assumed grouping method, where each group is defined by specific lower and upper VWC bounds. The advanced architecture of the proposed LSTM model is capable of capturing the complex nonlinear dynamics underlying these responses. For brevity, these dynamic interactions are collectively referred to as the SWRC slope $m_v$. This approach is designed to accommodate local variations and is expected to perform effectively within the specified study contexts and boundary conditions. The detailed modeling processes are similar to the illustrated steps in Section 3.2.1.

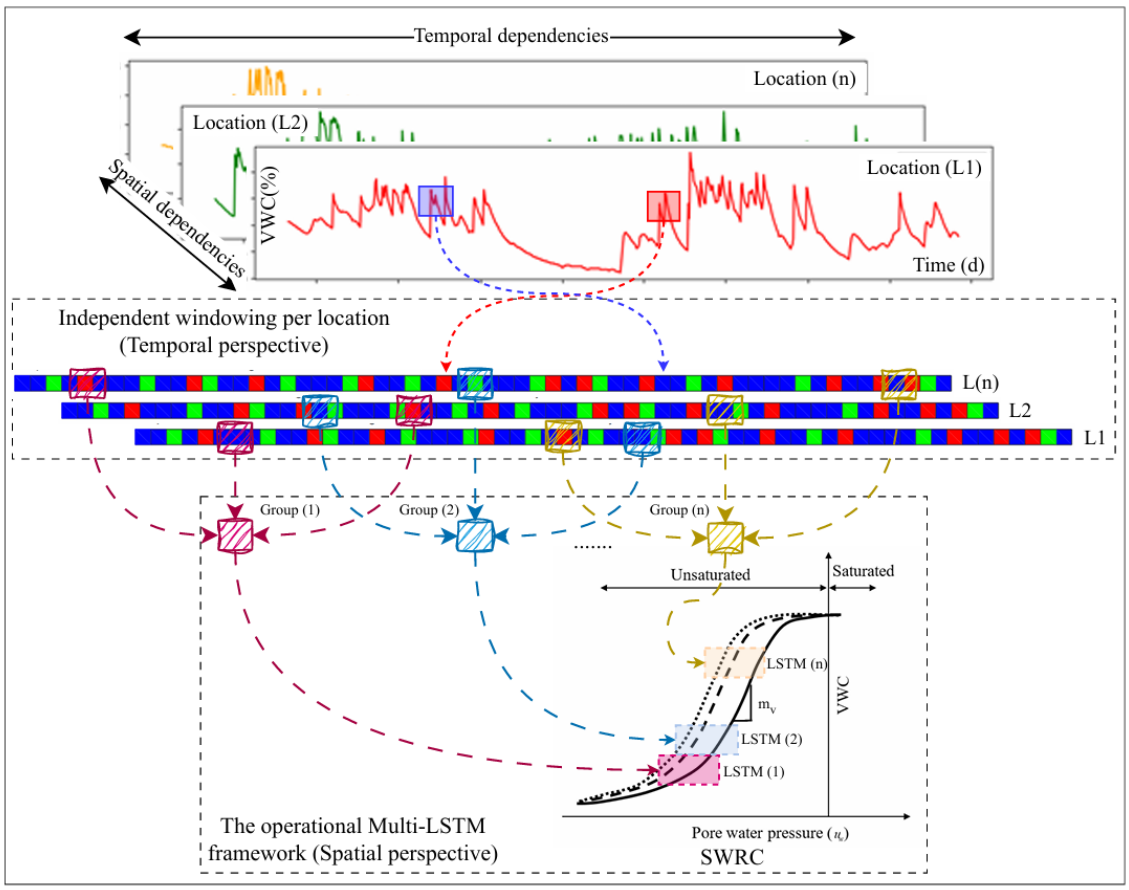



*Figure 3.8. The operational multi-LSTM framework*

### e) Feature selection

The chosen rainfall features (pattern, duration, intensity, and non-rainfall periods), visualized earlier in Figure 3.7, ensure accurate forecasting, as cumulative rainfall, commonly used in literature, may not reliably indicate soil saturation or recovery due to its neglect of random patterns and non-rainfall periods. To emphasize, different rainfall patterns can affect the landslide temporal response even with the same cumulative rainfall (Natalia & Yang, 2024). Additionally, short non-rainfall intervals may hinder soil recovery, while longer intervals could promote it (Bontemps et al., 2020). Given the availability of rainfall features and well-established correlations with the target parameters, an additional feature selection process is unnecessary. However, the primary challenge lies in determining the effective window size and incorporating these time-dependent characteristics into the modeling framework. Regarding the window size, the antecedent effect, influenced by rainfall history and hydrological conductivity, significantly impacts VWC's response to rainfall. The shallow colluvium layer in Hong Kong exhibits a relatively high permeability (on the order of $10^{-5}$ to $10^{-6}$ $m \cdot s^{-1}$), suggesting that a relatively short antecedent period is appropriate, as suggested by Au (1998). Furthermore, operational forecasting is constrained by the available forecast horizon from the Hong Kong Observatory (nine days plus the current day, totaling 10 days). Consequently, the antecedent window was coarse-tuned using durations of 4, 7, and 10 days to balance model performance and operational requirements. This rainfall window is modeled as a time-dependent variable within the LSTM framework.

### f) Numerical modelling

There is a limitation with field-monitored data, as it is restricted to specific time-space intervals. To address this issue, simulations can be used to consider various real-world rainfall patterns and their corresponding subsurface moisture response. Several numerical and analytical approaches, including widely used uncoupled analyses, have been conducted (Ng &

Shi, 1998; Ng et al., 2001). In this study, the seepage analysis is performed using SEEP/W (Geo-Studio, 2023a). In the transient seepage analysis, for simplicity, it is assumed that there is no hysteresis between the drying and wetting paths of the SWRC, as shown in Figure 3.8. The governing differential equation for saturated and unsaturated water flow through a two-dimensional unsaturated soil is illustrated in Equation 3.8 (Lam et al., 1987).

### g) Model implementation

The implementation was carried out using TensorFlow, a Python-based DL framework for developing advanced neural networks (Abadi et al., 2015). The proposed model structure was designed using the Sequential API. Various Python libraries supported different tasks: Pandas for data manipulation (Reback et al., 2020), NumPy for numerical computations (Harris et al., 2020), Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021) for graphical outputs, and scikit-learn for preprocessing (Pedregosa et al., 2011). Geospatial mapping was enabled using GeoPandas (Jordahl et al., 2020) and Shapely (Gillies et al., 2025), while Contextily (Arribas-Bel et al., 2020) supported visualizations. The interface is developed using the Streamlit framework in Python (Khorasani et al., 2023). The training was conducted on The Hong Kong Polytechnic University's High-Performance Computing (HPC) platform, equipped with an NVIDIA RTX 3080 Ti GPU, 64 GB RAM, six CPU cores, and Python 3.10.9 (The Hong Kong Polytechnic University HPC, 2025).

### h) Model evaluation

This study compared site-specific with operational models to benchmark the proposed model's ability to achieve the study's goals. The comparison metrics used were Mean Absolute Error (MAE), Normalized Mean Absolute Error (NMAE), and the coefficient of determination ($R^2$), as shown in Equations (3.12-14), respectively. MAE measures the average absolute error across all predictions. NMAE normalizes the MAE over the parameter's range, providing a balanced metric regardless of the parameter's magnitude. $R^2$ is a statistical measure indicating how well

the predicted values align with the actual values. To quantify predictive uncertainty and to simulate the possible inherent variability and heterogeneity of soil properties in the sensor-based dataset, Monte Carlo (MC) dropout was integrated (Gal & Ghahramani, 2016) (Equation 3.11).

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \tilde{y}_i\right| \qquad \textit{(Equation 3.12)}$$

$$\mathrm{NMAE} = \mathrm{MAE} / \left(y_{\max} - y_{\min}\right) \qquad \textit{(Equation 3.13)}$$

$$\mathrm{R}^2 = 1 - \left(\sum_{i=1}^{n}\left(y_i - \tilde{y}_i\right)^2 / \sum_{i=1}^{n}\left(y_i - \bar{y}_i\right)^2\right) \qquad \textit{(Equation 3.14)}$$

Where $y_i$ represents the true value, $\tilde{y}_i$ is the predicted value, $n$ is the total number of predictions, $y_{\max}$ and $y_{\min}$ represents the maximum and minimum values of the parameter range, respectively, and $\bar{y}$ is the mean of the observed data.

### *3.3.2. VWC Prediction Model Assumptions*

The key assumptions that underpin the robustness and applicability of this methodology are detailed as follows: the selection of this colluvium layer is to ensure that the developed model is significant to be used with real-world applications since this layer covers about 15% of the territory, and most rainfall-induced landslides in Hong Kong are shallow, occurring in the colluvium layer. Moreover, the relative spatial variability of the colluvium across different locations supports the implementation of simplified, yet reasonable, assumptions in the modeling framework. This layer, formed from past mass movements, is weakly consolidated, texturally uniform, and typically only a few meters thick (Geotechnical Engineering Office, 1993). This layer could be classified as loam or silt loam (Maunsell Geotechnical Services Ltd., 2007) according to the USDA. The collected samples (33 samples) across 11 geographically dispersed sites were predominantly characterized as moist, brown, silty-clayey, and gravelly sands. Effective cohesion averaged 3.67 kPa (range: 0.4-9.6 kPa) and the internal friction angle averaged 30.5° (range: 20.4–35.8°) (ETS-Testconsult Ltd., 2023), in agreement with the literature (Maunsell Geotechnical Services Ltd., 2007; Wong, 2020). The observed standard

deviations were 2.5 kPa for cohesion and 5.04° for the friction angle. The colluvium layer is expected to have a hydraulic conductivity of $10^{-5}$ to $10^{-6}$ $m.s^{-1}$ (Ng & Shi, 1998; Maunsell Geotechnical Services Ltd., 2007), indicating relative variability. Environmentally, rainfall could be one of the primary drivers of subsurface soil behavior. Because of Hong Kong's high humidity, rainfall may be divided into infiltration and runoff. With that in mind, runoff is influenced by factors such as vegetation area, rainfall duration, rainfall volume, and slope geometry (which ranges from 15° to 40° (Liu et al., 2022), among other factors. However, some of these parameters are not readily monitored.

Based on the insights above, several unavoidable assumptions were adopted. First, it was assumed that the initial VWC serves as a proxy for antecedent hydrological behavior, capturing inherent spatial features such as slope geometry and vegetation cover. In essence, because VWC quantifies the actual infiltrated rainfall, it naturally reflects the runoff ratio. Furthermore, the unique data preparation method and multi-approach modeling framework were designed to effectively accommodate the natural variability. These assumptions were initially validated through sample collection (33 samples) across 11 locations and subsequently reinforced via comprehensive modeling and testing using data from 15 distinct sensors deployed throughout the region (refer to Chapter 4). To clarify, initial and soaked moisture contents for 33 samples were measured, ranging from 0.063 to 0.219, with the standard deviation (SD) of the response (soaked - initial) being 0.038. Partitioning the data using non-overlapping subgroups based on the initial value led to reduced variability. For instance, a single cutoff produced groups with SDs of 0.017 and 0.030; two cutoffs yielded SDs of 0.017, 0.017, and 0.024; and three cutoffs resulted in 0.015, 0.018, 0.003, and 0.009. It is important to note that the optimal number of cutoffs depends on the dataset size, which in turn influences model performance. Ultimately, the low intra-group variability across categorized samples

validates this assumption, confirming that these groupings reasonably capture spatiotemporal variability.

### *3.3.3. Hypertuning Strategy for Enhanced Model Performance*

#### a) Background and rationale

One of the key features that directly indicates how much infiltrated rainfall is the VWC. From a physical perspective, VWC, defined as the ratio of water volume to soil volume and typically expressed in $[cm^3 \cdot cm^{-3}]w$ or percentage, serves as a critical variable governing Earth surface-atmosphere interactions with significant meteorological and climatic implications (Seneviratne et al., 2010). For shallow rainfall-induced landslides, research demonstrates that VWC gradient features provide reliable predictive capability (Suk et al., 2022). Specifically, changes in matric suction and VWC within the surface soil layer (0-5m depth) represent fundamental triggering mechanisms for shallow slope failures (Tsuchida et al., 2014; Johari et al., 2018). The physical process unfolds as rainfall infiltration increases pore-water pressure and bulk density, thereby elevating driving forces while simultaneously reducing shear strength and slip resistance. These effects intensify following prolonged or intense rainfall events, substantially compromising slope stability. Mechanistically, rainfall infiltration propagates a wetting front downward until encountering an impermeable layer, subsequently raising the water table (Figure 3.9.a). VWC evolution occurs in two distinct phases: initial response as the wetting front passes the sensor location ($t_1$), followed by water table rise to the surface ($t_2$). This progression defines the critical temporal window for shallow slide initiation, as displacement accelerates with increasing VWC, culminating in failure near saturation conditions (Fukuhara, 2022; Tohari et al., 2007) (Figure 3.9.b–c). Consequently, continuous VWC monitoring provides an effective mechanism for early warning of impending landslides (Suk et al., 2022).

With that in mind, current research in landslide prediction reveals a significant gap: while substantial work has focused on surface parameters, subsurface forecasting remains

comparatively underdeveloped (Section 2.4.3; Table 3.3). Although the operational VWC model presented in Section 3.3.1 shows promise, it requires enhancement to achieve robust performance across heterogeneous sites and multiple soil depths. This section addresses these limitations by developing a systematic hypertuning approach that synergizes empirical DL optimization with the physical principles of infiltration dynamics, explicitly targeting the complex interactions of unmonitored rainfall-related features.

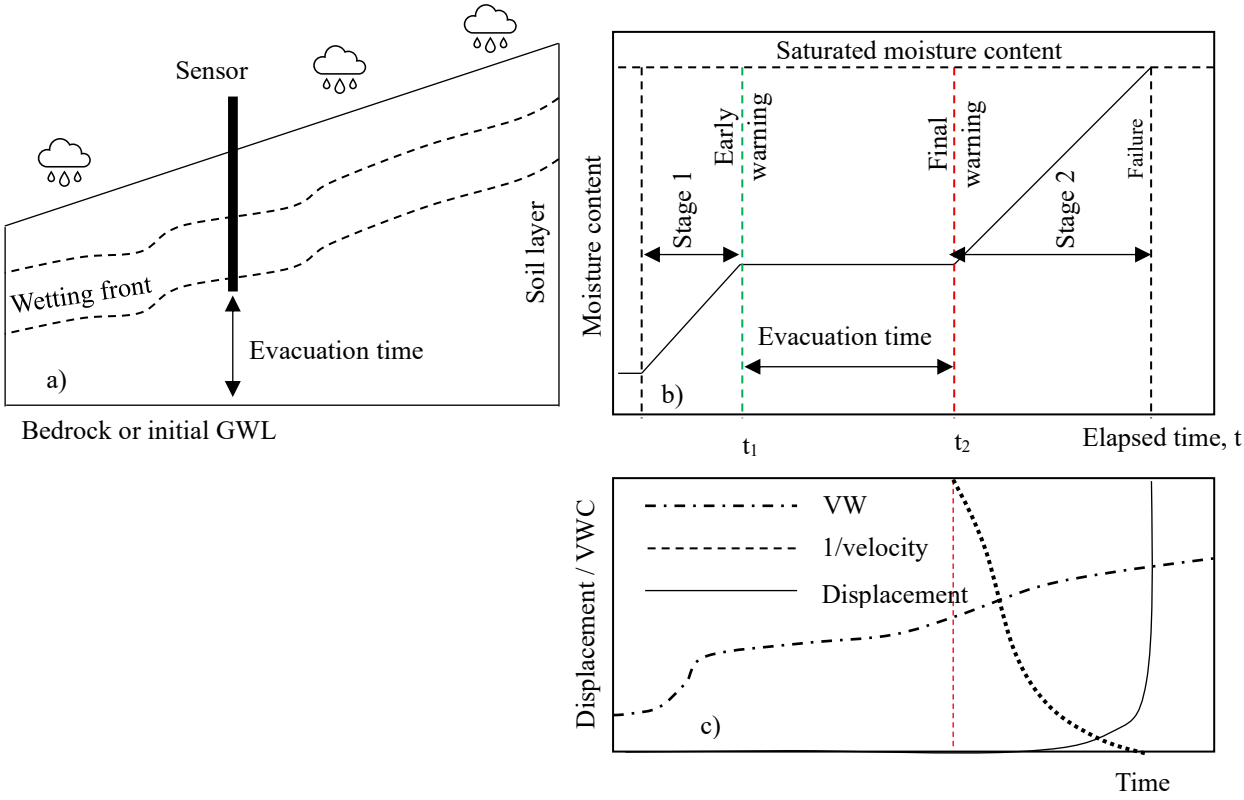



*Figure 3.9. a) Seepage under rainfall; b) Moisture content-based prediction; c) VWC versus displacement and velocity inverse (modified from Fukuhara, 2022; Tohari et al., 2007)*

*Table 3.3. Comparison with available literature*

| Comparison | Existing literature | This study |
|---|---|---|
| Predicted target | DL models are commonly used to forecast landslides on reservoir slopes, with limited attention to time-varying components. | This study predicts and correlates the time-dependent VWC response of shallow-induced landslides in spatially varied soils. |
| Physical response | Focused on surface displacement predictions where the cumulative displacement was decomposed into trend and periodic terms. | This study leverages subsurface VWC, which, unlike surface displacement, exhibits no trend or creep component. Figure 2.18 illustrates the trend and periodic components of a time series problem. |
| Input sequence | The research has not given enough consideration to the window size (i.e., antecedent value). | The dynamic response of VWC was examined using a Hypertuning approach. |
| Spatiotemporal prediction | Single-point prediction is widely considered. | The uncertainty of VWC predictions is assessed through interval analysis. |
| Time intervals | Typically, one month; it may run up to ten years with reasonably smooth curves. | A minimum time interval of one day, accounting for the variation in the VWC, is considered. |

### b) Model selection and development

Figure 3.10 outlines the systematic time series modeling workflow, with detailed methodological specifications provided in Table 2.8. For this specific problem, several critical considerations, often overlooked in the literature, should be emphasized. These include determining the optimal antecedent period (window size) and establishing a rigorous, systematic framework for model evaluation and tuning, among other considerations (refer to Table 3.3). The modeling process, specifically tailored to achieve the study's objectives, is outlined in Figure 3.11. The methodology proceeded in two primary stages. First, the target VWC values were used as both input and output to establish a naïve benchmark. Subsequently, a comprehensive suite of statistical and DL models was evaluated to select the best performer and set a more robust "model benchmark," acknowledging that no single model is universally optimal. The models compared included: Simple Average, Single Exponential Average (Holt, 1957), Triple Exponential Average (Winters, 1960), SARIMA (Box et al., 1994), Rolling

Moving Average (Zivot & Wang, 2006), Multivariate Regression (Draper & Smith, 1998), DNN (LeCun et al., 2015), RNN (Rumelhart et al., 1986), LSTM (Hochreiter & Schmidhuber, 1997), and CNN (LeCun et al., 1998), as summarized in Table A4.

To better capture nonlinearity, activation layers are incorporated. TensorFlow provides several activation functions, including the rectified linear unit (ReLU), Leaky ReLU, hyperbolic tangent (tanh), and sigmoid. ReLU reduces computational load by using simple gradient values, but suffers from the "dying ReLU" issue. Leaky ReLU addresses this by allowing small gradients for negative inputs (Maas et al., 2013). This study applies Leaky ReLU in hidden layers and a linear function in the output layer. Effective input pre-processing enhances ML performance (Varangaonkar & Rode, 2023), while regularization techniques like dropout mitigate overfitting. In this study, two dropout layers were applied within the LSTM. Additionally, it was demonstrated that normalizing enhances the model's performance. This study employed Maximum and Minimum normalization techniques (Equation 3.15).

$$X_{j-scaled}^{(i)} = \frac{X_j^{(i)} - X_{j-\min}}{X_{j-\max} - X_{j-\min}}, 0 \le X_{j-scaled}^{(i)} \le 1 \qquad \textit{(Equation 3.15)}$$

Where $X_j^{(i)}$ is the input feature or variable, $X_{j\text{-}max}$ is the maximum value of $X_j^{(i)}$, and $X_{j\text{-}min}$ is the minimum value of $X_j^{(i)}$.

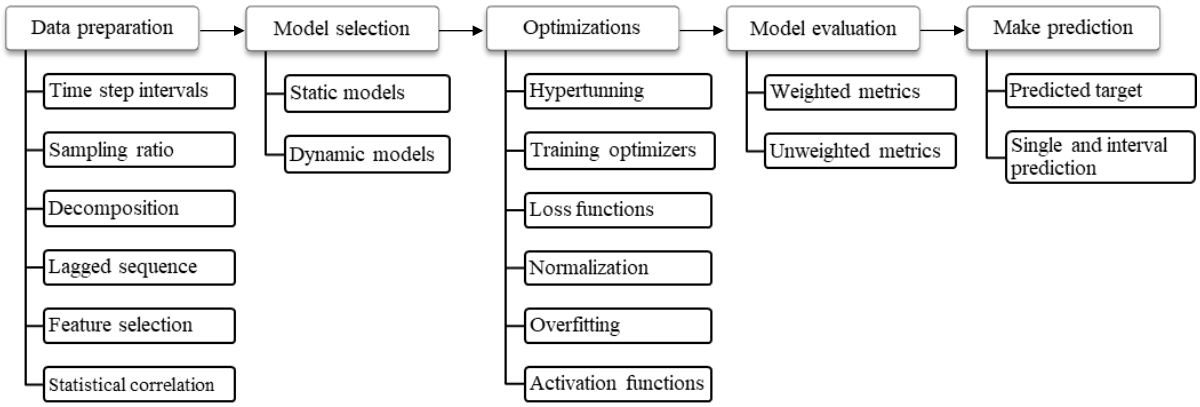



*Figure 3.10. Time series modeling process.*

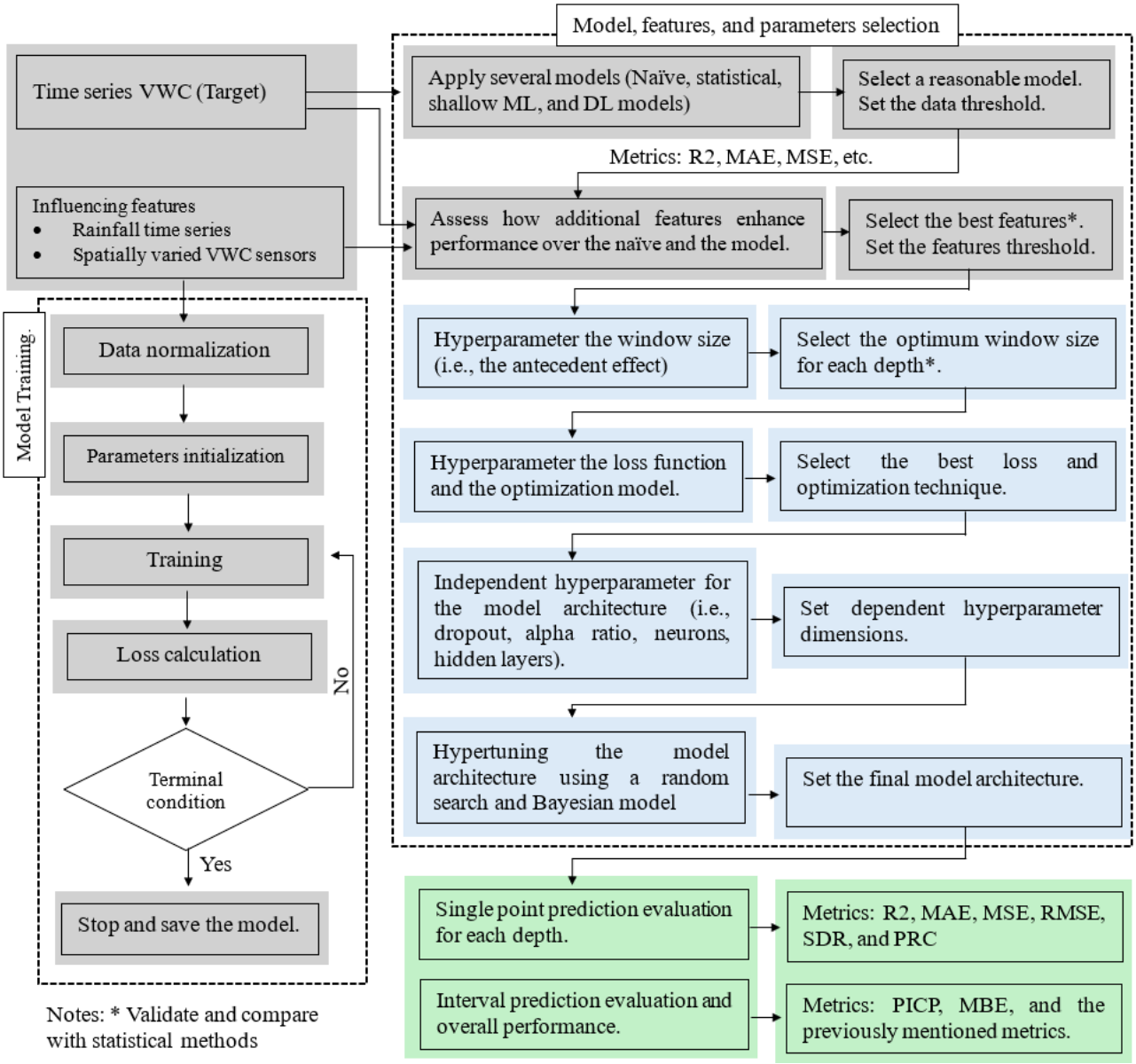



*Figure 3.11. Time series model hypertuning flow chart*

### b) Data preparation and feature selection

The augmented Dickey-Fuller (ADF) test was used to ascertain whether the data were stationary. The ADF test is carried out by first assuming the presence of a unit root in the time series and then computing its significance statistic (p). The time-series data are said to be stationary if the value of p is smaller than 0.1 (Hyndman & Athanasopoulos, 2018). Two methodologies can be followed for feature selection: knowledge-based and statistical methods. Considering the knowledge-based method, first and foremost, rainfall, affected by other factors

(e.g., evapotranspiration), is the primary element influencing the VWC response. Rainfall is a complicated phenomenon with significant geographical and temporal variability across various scales (Folino et al., 2023). Thus, neighboring sensors were employed to account for unmonitored data and increase VWC prediction accuracy. This method is employed in several relevant pieces of literature, where characteristics such as displacement, velocity of displacement, and displacement increment are used to enhance the displacement prediction model's performance (Huang et al., 2022b; Meng et al., 2024; Nava et al., 2023; Xi et al., 2023; Xing et al., 2020). Statistical methods utilize models such as the Pearson correlation (Pearson, 1895), the maximal information coefficient (MIC) (Reshef et al., 2011), the information gain ratio approach (Quinlan, 1986), the Shapley Additive exPlanations (SHAP) algorithm (Lundberg & Lee, 2017), Kendall correlation coefficient (Kendall, 1938), and the DL model itself to extract the relevant features. These approaches capture the connection between two variables, calculate linear and non-linear correlations, and discover essential and difficult-to-recognize correlations (Li et al., 2020). This study employs Kendell, Pearson, and MIC methods to investigate the link between the two variables, as well as provide a sensitivity analysis to look at how DL can successfully extract features.

### c) Model benchmarks and evaluation

As illustrated in Figure 3.11, first, in this process, the target VWC values were used as input and output to calculate a naïve benchmark (the best prediction accuracy could be achieved for the given time series). Then, the aforementioned models were compared to select a reasonable model and set another benchmark: model benchmark (the best prediction accuracy the model could achieve). Second, the features were analyzed to determine the optimal set and set another benchmark: features benchmark (the best accuracy a given input feature can provide). For this objective, six metrics were used: coefficient of determination ($R^2$), mean absolute error (MAE), root mean squared error (RMSE), and mean squared error (MSE). These parameters can be

obtained as presented in Equations 3.12, 3.14, 3.16, and 3.17 (Willmott & Matsuura, 2005). The greater the value of these parameters, the worse the prediction of the model, when values closer to 0 indicate higher accuracy. Additionally, the standard deviation ratio (SDR) and Pearson-R Correlation (PRC) are used to evaluate the model. Where the closest SDR to 0 highlights high accuracy (Bland & Altman, 1996) and Pearson-R Correlation (PRC) assumes that the values should fall between -1 and +1, where -1 denotes the strongest possible agreement, and 0 is the largest possible disagreement (Cohen, 1998). Interval predictions were evaluated using Prediction Interval Coverage Probability (PICP) (Ge et al., 2023). Overall performance was evaluated using the aforementioned metrics, supplemented by Mean Bias Error (MBE). Positive MBE values indicate overestimation, while negative values denote underestimation, as described in Equation 3.18 (Khalil & Shaffie, 2013).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2} \qquad \textit{(Equation 3.16)}$$

$$MSE = (RMSE)^2 \qquad \textit{(Equation 3.17)}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i) \qquad \textit{(Equation 3.18)}$$

Where $Y_i$ is the specific value of the $i^{th}$ real data, $\overline{Y}$ is the average value of the real data, and $X_i$ is the specific value of the $i^{th}$ predicted data.

### c) Data splitting and temporal ordering

One important factor influencing the predicted value is the ratio of the training set length. An excessively high percentage of the training set in a particular situation will lead to an inadequate test sample size. In this instance, the test prediction values' uncertainty and variability will rise with the model's random variable. On the other hand, a small training set will result in poor prediction performance for the model. As of right now, there is no consistent standard that specifies the length of the training set or how long the training set is in terms of the duration of the entire time series. Typically, a ratio of 70% to 30% is adopted. This research used a training-

to-testing ratio of 400:149 (73% to 27%), which will be sensitively investigated. To keep temporal ordering and, as much as possible, have the same pattern for the triggering features represented in each subset, the standard holdout was chosen as the validation strategy, ensuring that the validation and test data are at the end of the series (Togneri et al., 2022). The data is divided using the sliding technique (Hindriks et al., 2016) with a 15-day window, which will be sensitively investigated.

### d) Model optimization, training process, and implementation

For model architecture, training the model could be done effectively by choosing a) a suitable loss function to monitor the training and validation states, and b) an acceptable optimizer to control the initial weights, epochs, stopping criteria, fast convergence, and so on. Sensitivity analysis was performed to select the loss function and the optimizer as follows: a) for the loss functions, Welsch (Barron, 2019), Mean Squared Error (MSE), Huber (Gupta et al., 2020), Quantile (Chen et al., 2020), Log-cosh (De Backer et al., 2019), Tukey's Biweight (Egrioglu et al., 2023), and Cauchy (Li & Li, 2023) functions were investigated; b) for the optimizer, SGD, LAMB (You et al., 2019), AdamW (Loshchilov & Hutter, 2017), Ranger (Yong et al., 2020), Yogi (Zaheer et al., 2018), Adamax, Nadam, Adadelta (Zeiler, 2012), Adagrad, RMSProp, Adam (TensorFlow) were investigated. The current study used both grid search and random search: the grid search was selected to find the trend for each parameter independently of other parameters such as window size, number of features, return sequence layer, loss function, and optimizer; while random search with KerasTuner (O'Malley et al., 2019) to determine the parameters of the model, such as the neurons, the dropout rate, the Leaky ReLU parameter, and the stepsize of the Optimizer. Visual tools like learning curves (Matplotlib package) are utilized to understand how the model performs during validation and training. Table 3.4 contains a list of the model's hyperparameters. Because this procedure takes time, it was only done for the best-chosen model (Figure 3.12).

*Table 3.4. LSTM hyperparameter search space*

| Parameter | Search space | Hypertuning |
|---|---|---|
| Features | [(R), (R+E), (R+E+$V_1$), (R+E+$V_1$+$V_2$+$V_3$)] | Independent tuning using a grid search |
| Window size | [1, 5, 10, 15, 20, 25, 30] | |
| Return hidden sequence layer. | [1,2,4,8,16] | |
| Loss function | [Welsch, MSE, Huber, Quantile, Log-cosh, Tukey's Biweight, Cauchy] | |
| Optimizer | [SGD, LAMB, AdamW, Ranger, Yogi, Adamax, Nadam, Adadelta, Adagrad, RMSProp, Adam] | |
| Batch size | [4,8,16,32,64,128] | |
| Dropout | [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9] | Dependent tuning using random search |
| Leaky ReLU | [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9] | |
| Hidden units | [64,96,128,160,192,224,256] | |

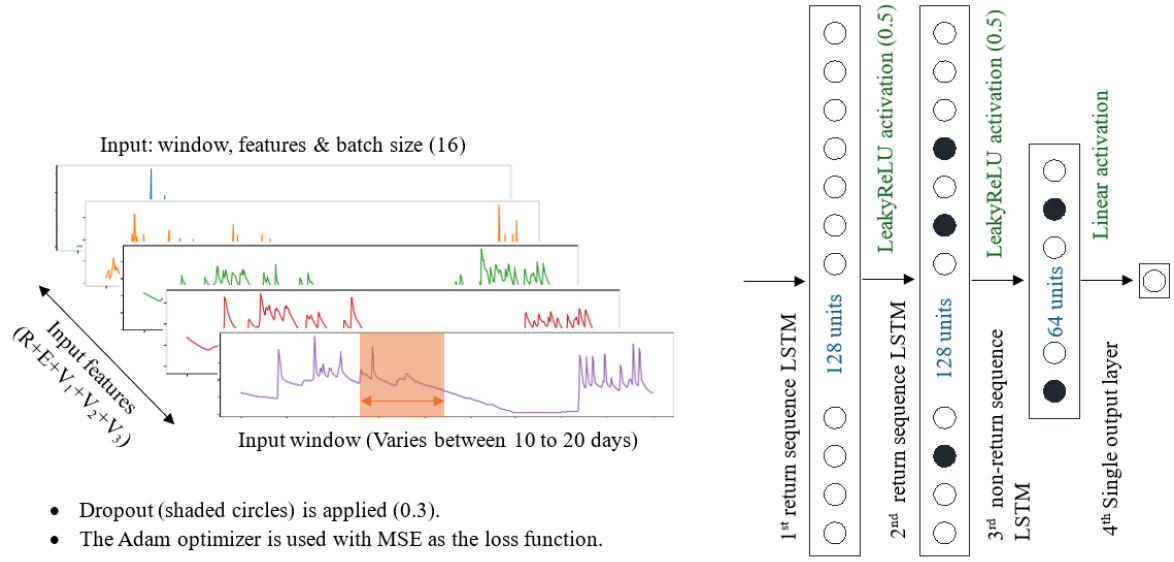



*Figure 3.12. Schematic representation of the shared network model structure*

The implementation was carried out using TensorFlow (Abadi et al., 2015). Keras Tuner was employed for hyperparameter optimization. The Sequential API was used to design the model, incorporating LSTM, dropout, and dense layers with Leaky ReLU activation. Python libraries supported various tasks: Pandas for data manipulation (Reback et al., 2020), NumPy for numerical computations (Harris et al., 2020), Matplotlib for visualization (Hunter, 2007), Seaborn for statistical plotting (Waskom, 2021), scikit-learn for preprocessing (Pedregosa et al., 2011), and Scikit-Optimize for fitting a Gaussian distribution (Head et al., 2018). For feature analysis, additional libraries were employed, including the statsmodels library (Seabold & Perktold, 2010), the MINE package (Albanese et al., 2013), and scipy.stats (Virtanen et al.,

2020). Taylor diagrams were generated using OriginPro 2025b (OriginLab Corporation, 2024). The training was conducted on Google Colaboratory (Google Research, 2024), while hyperparameter tuning, requiring intensive computation, utilized The Hong Kong Polytechnic University's High-Performance Computing (HPC) platform with an NVIDIA RTX 3080 Ti GPU, 64 GB RAM, six CPU cores, and Python 3.10.9 (The Hong Kong Polytechnic University HPC, 2025).

### e) Interval predictions

Estimating the VWC confidence interval is essential for quantifying the uncertainty in predicted values. This methodology assumes that the residuals (ζ) follow a Gaussian distribution with zero mean and are independent of the input variables. The standard deviation (σ) of the residuals is computed as a measure of their spread, enabling the construction of confidence intervals at a given confidence level (95%). Using the Z-score (z=1.96), the interval width is calculated as interval = Z.σ. The upper and lower bounds of the interval are defined as $\left[y-(\sigma.z), y+(\sigma.z)\right]$, respectively, where $y$ is the output predicted variable. The reliability of the interval is quantified using the PICP, calculated as $PICP = N_C / N_T$ where $N_C$ is the number of actual values within the interval and $N_T$ is the total number of validation samples.

### f) Shallow rainfall-induced landslides FOS

The FOS during the progression of the wetting front in the shallow layer can be calculated as presented in Equation 3.19 (Fredlund & Rahardjo, 1993; Tsuchida et al., 2014). This equation accounts only for the variation of the VWC and suction with time. This assumption is valid within the context of this study, as it is widely reported in the literature that rain-induced landslide failures are generally shallow, with slip surfaces parallel to the slope surface (Das et al., 2022; Saadatkhah et al., 2015; Thang et al., 2022). This equation simplifies the FOS calculation by accounting for the temporal variation in VWC and suction. Despite its simplified form, this approach is well-suited for evaluating the performance of the proposed model and

its predicted VWC values. Specifically, the FOS is calculated using both monitored and predicted VWC values, with suction derived from the VWC using the Fredlund equation (Fredlund & Xing, 1994). For the transient seepage analysis, it is assumed no hysteresis between the drying and wetting paths of the Soil-Water Retention Curve (SWRC). Therefore, variations in VWC can be correlated with changes in pore water pressure $\partial\theta_w = m_w \partial u_w$. This approach offers a robust, approximate framework for quantifying the impact of VWC prediction deviations.

$$FOS = \frac{c_o + \gamma h \cos^2 \beta \ \tan\phi_o + \psi \tan\phi^b}{\gamma h \cos\beta \ \sin\beta} \qquad \textit{(Equation 3.19)}$$

Where $c_o$ is the initial state of soil cohesion, $\gamma$ is the bulk unit weight of soil, $h$ donates the soil depth, $\beta$ is the soil slope, $\phi_o$ is the initial state of the soil internal friction angle, $\psi$ is the matric suction, which reflects the soil's ability to retain water due to capillary forces, $\phi^b$ is the angle that kinks the rate at which shear strength grows as matric suction increases, σ refer to the normal stress acting on the soil, $W_t$ is the weight of soil, $\tau$ is the shear stress acting on the soil, and $m_v$ is coefficient of volumetric change.

As for FOS estimation using VWC and suction dynamics, Table A5 lists the slope characteristics for the Pa Mei area. These parameters are based on relevant literature. In the current study, the soil in the sensor response zone is identified as colluvium. Referring to the relevant literature in Hong Kong, shallow layers typically comprise colluvium, completely decomposed granitic, volcanic, and sedimentary rocks, which the USDA classifies as loam or silt loam (GCO, 1982; Maunsell Geotechnical Services Ltd., 2007). Consequently, the Fredlund-Xing parameters were derived, assuming loam soil characteristics. It is important to emphasize that this study primarily focuses on enhancing the prediction accuracy of VWC. The parameters listed in Table A5 serve to provide reasonable and contextually appropriate estimations for the FOS, without affecting the core objective.

### *3.3.4. Distributed Multi-Sensor (Subsurface) Prediction Model*

#### a) Background and rationale

Complementary to the approaches in Section 3.3, this objective also aims to predict the earth's pressure, suction, and moisture dynamics. To clarify, this objective aims to achieve better predictions using DL. Then, in the case of data-related issues, it can help solve the following issues: a) enhanced the level of uncertainty due to the unavailability of rainfall controlling features by learning from best correlated hydrological and mechanical characteristics; b) missing hydrological or mechanical responses and how DL can outperform the superficial and shallow statistical techniques; c) subsurface hydrological and mechanical response prediction of future scenarios, supporting decision making and early warning; and d) virtual sensing and understanding the spatiotemporal correlations among different features, aiming to reduce the number of active sensors.

#### b) Modeling process

The modeling process is similar to the above-illustrated procedure (Figures 3.10 and 3.11) except for some minor changes (Figure 3.13) as follows: the process starts with data preparation by handling the missed and distorted data. The missing data is handled by first assuming linear filling (Huang, 2021). The augmented Dickey-Fuller (ADF) test was used to ascertain whether the data were stationary. The time-series data are said to be stationary if the value of the significance statistic (p) is smaller than 0.1 (Hyndman & Athanasopoulos, 2018). This step is followed by using the assumed linear fitted dataset with the DL model to update the filled data (for non-stationary cases). The DL model is proposed to be trained with the best-related features to update better and improve the accuracy of the filled data. The distorted data is handled by dropping these values. Then, the validated values and the corresponding related features are used to validate the theoretical deterministic model (Equation 3.20), which will estimate the distorted values by fiddling with the equation with the related features.

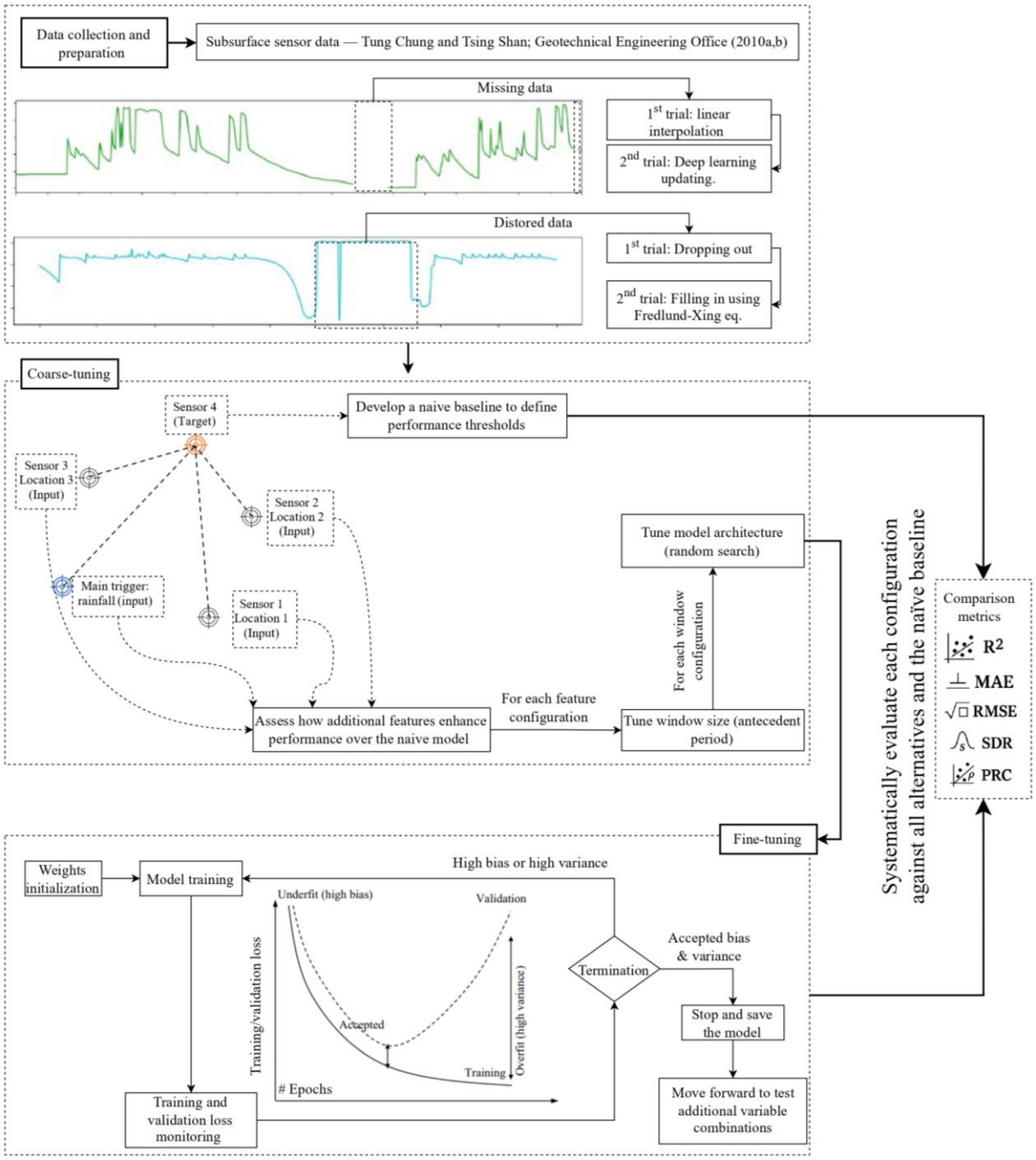



*Figure 3.13. Multi-sensor prediction modeling workflow.*

To clarify, distorted suction values were identified as sensor dry-out when these conditions were met: (1) readings near zero occurred during extended dry periods with no recorded rainfall, (2) VWC remained at dry values, and the suction–VWC pair deviated from the site SWCC. Flagged points were removed from the calibration dataset; remaining valid suction–VWC pairs were used to fit the Fredlund–Xing model (Equation 3.20) (Fredlund,

1998; Fredlund & Xing, 1994). Since the distorted data primarily occurred during dry periods, a simplified dry approach was adopted, assuming no hysteresis between the wetting and drying Soil Water Characteristic Curve (SWCC). This study employs the LSTM model (Hochreiter & Schmidhuber, 1997). The feature selection matrix is listed in Table 3.5, and the model Hypertuning is listed in Table 3.6. The study validated the proposed model using field-monitored data from Tung Chung and assessed its generalization using the Tsing Shan landslide case study.

$$\theta_w = C(\psi)\left[\frac{\theta_s}{\ln\left\{\left(e+\left(\frac{\psi}{a_f}\right)^{n_f}\right)\right\}^{m_f}}\right], C(y) = \left[1-\frac{\ln\left(1+\frac{\psi}{\psi_r}\right)}{\ln\left(1+\frac{1000000}{\psi_r}\right)}\right] \qquad \textit{(Equation 3.20)}$$

Where, $\theta_w$ is the VWC, $\theta_s$ refers to the saturated water content, $a_f$ is a suction value related to the air entry of the soil, $n_f$ is a soil parameter that reflects the slope of the soil-water characteristic curve, $\psi$ refers to the soil suction, $m_f$ a parameter reflects the residual water content, e equals 2.71828, $C(\psi)$ is a correction function, and $\psi_r$ is the suction value corresponding to the residual water content and can be assumed to be 1500 kPa.

*Table 3.5. Multi-sensor model feature selection matrix*

| Features | Target | | | |
|---|---|---|---|---|
| | Rainfall | Suction | Earth pressure | VWC |
| Rainfall | - | √ | √ | √ |
| VWC | - | √ | √ | - |
| Suction | - | - | - | √ |
| Earth pressure | - | - | - | √ |
| Groundwater level | - | - | - | √ |
| Spatially varied data | - | √ | √ | √ |
| Amended features * | - | √ | √ | √ |
| *Amended (high-level) features are integrated by investigating the adequate input sequence length. | | | | |

*Table 3.6. Multi-sensor model Hypertuning search space*

| Parameter | Search space | Hypertuning |
|---|---|---|
| Features | Refer to Table 3.5 | Independent Hypertuning using a grid search |
| Window size (days) | [10, 20, 30, 40, 50] | |
| Return hidden sequence layer. | 1 | |
| Loss function | MSE | |
| Optimizer | Adam | |
| Batch size | 16 | |
| Dropout | [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9] | Dependent tuning using random search |
| Leaky ReLU | [0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9] | |
| Hidden units | [64,96,128,160,192,224,256] | |

### c) Feature selection

The selection of input features is a critical step in building a predictive model for complex systems like landslides. This study employs a hybrid feature selection strategy, integrating both domain-knowledge-based and statistical-data-driven methods to ensure the model captures both the physical drivers and the complex, non-linear interdependencies of the subsurface environment. The primary driver of landslides in the study area is rainfall; however, its effect is mediated by the subsurface hydrological and mechanical processes. Therefore, the knowledge-based selection was guided by the following principles: a) Rainfall data was included as a fundamental input feature for all target variables, acknowledging its role as the initial catalyst; b) To account for the spatial variability of the slope's response, readings from sensors measuring different variables at different, nearby locations were incorporated. This allows the model to learn from the spatial context and effectively treats the state of the surrounding slope as a dynamic boundary condition; c) Key subsurface variables known to interact mechanically and hydrologically were included as inputs for predicting other variables. For instance, VWC directly influences matric suction and earth pressure and was therefore used as an input feature for predicting those targets.

While the knowledge-based approach provides a physically sound foundation, it does not quantify the strength or nature of the relationships. To refine the feature set and capture complex temporal correlations, a two-step data-driven approach was implemented: a) The

Pearson correlation coefficient (Pearson, 1895) was first employed to identify strong linear relationships between potential input features and the target variables. This helped to validate the knowledge-based selections and identify the most directly influential contemporaneous features; b) Recognizing that Pearson correlation cannot account for time-lagged dependencies, a critical aspect of landslide triggering, a sensitivity analysis was conducted using the LSTM model itself. For each target variable, the model was trained using the knowledge-based feature set. The importance of each input feature was then evaluated by analyzing the change in model performance when that feature's historical sequence was changed. This process identifies which features, and over what time window (sequence length), are most critical for accurate prediction, thereby capturing non-linear and time-dependent relationships that are "difficult-to-recognize" with static methods.

The final input feature set for each target variable was determined by synthesizing the knowledge-based and data-driven results. This involved creating "amended" or high-level features: a) The optimal input sequence length for each feature was determined through sensitivity analysis during model optimization, ensuring the model was provided with sufficient historical context to learn the antecedent conditions leading to a change in the target variable; b) The final models predict a target variable (e.g., Suction) using a multivariate time series input that includes contemporaneous and antecedent values of rainfall and the most influential nearby subsurface responses (e.g., VWC at multiple locations), as identified by the above strategy. The initial feature matrix based on this knowledge-based approach is summarized in Table 3.5. The exact, detailed list of input features and their sequence lengths for each predictive model is provided in Appendix Tables A9-A15.

## 3.4. SLOPE STABILITY ANALYSIS AND AUTOMATION FRAMEWORK (TIER 3)

### *3.4.1. Automated Parametrization Model*

#### a) Background and rationale

A prerequisite for stability analysis is obtaining accurate soil parameters. This section presents a methodological framework for parameterizing the hardening soil (HS) model, offered as a clear example that can be adapted to other constitutive models. The traditional conceptual parameterization workflow, provided by GEOSLOPE International Ltd. (2021), which has widespread use and an established methodology, consists of three main steps: i) the experimental approach, ii) the numerical approach, and iii) the result verification. In the experimental approach, the soil sample is prepared, triaxial tests are performed, and stress-strain data are collected (Figure 3.14. (a-c)). Following that step, manual parameterization begins. The numerical approach (i.e., manual parametrization) involves parametrization of model constraints, including: a) strength parameters ($c'$, $\phi'$, $R_f$) (Figure 3.14. (d-e)); and b) stiffness parameters ($E_{oed}^{ref}$, $E_{50}^{ref}$, $E_{ur}^{ref}$, $K_0^{NC}$, $m$) (Figure 3.14. (f-j)). Since this study emphasizes the parameterization process, a concise overview of how these parameters are derived manually is provided in the following lines. For a detailed explanation, readers are encouraged to refer to GEOSLOPE International Ltd. (2021) for comprehensive insights: a) Concerning the strength parameters, in Figure 3.14. (d), the $c'$ is derived from the shear-normal stress space as the intersection of the failure line with the vertical axis, and the $\phi'$ corresponds to the slope of the failure line. In Figure 4(e), the $R_f$ is calculated based on the slope of parallel lines in the relationship between $(\varepsilon_1 q_f / q)$ and axial strain $(\varepsilon_1)$; b) For stiffness parameters, the reference stress ($\sigma^{ref}$) is used to determine the failure deviatoric stress ($q_f^{ref}$) following the Mohr-Coulomb strength parameters (Figure 3.14. (g)). The $E_{50}^{ref}$ is calculated as the slope of a line passing through the point corresponding to 50% of the failure deviatoric stress derived from a specific triaxial test (Figure 3.14. (f)). The $E_{ur}^{ref}$ is determined using the least squares method, as depicted in Figure 3.14. (f). The $E_{oed}^{ref}$ is defined as the tangent stiffness at the point where

vertical stress equals the reference stress, as shown in Figure 3.14. (h). $E_{oed}^{ref}$ requires results from an oedometer test but can be approximated as a ratio of the half-failure reference stiffness, typically ranging between 50% and 80%, depending on the soil type (GEOSLOPE International Ltd., 2020). Similarly, the coefficient of lateral earth pressure, the ratio of lateral stress ( $\sigma_3$ ) to vertical stress ( $\sigma_1$ ), can be computed by the solver ( $K_0^{NC} = 1 - \sin\phi'$ ) (Figure 3.14. (i)). The $m$ exponent is determined as the slope of a trend line derived from the relationship between $\ln(E_{50} / E_{50}^{ref})$. Typically, this exponent is assumed consistent for both loading and unloading-reloading phases (Figure 3.14. (j)).

Figure 3.15 illustrates the automated conceptual parameterization workflow. The proposed tool streamlines the traditional parameterization process shown in Figure 3.14. (d-j) by automatically processing laboratory data (Figures 3.14. (a-c)) using the user-friendly APHs tool. This tool integrates a novel multi-LSTM model focusing exclusively on triaxial test data. In contrast to conventional methods, which depend on oedometer tests or empirical correlations for determining parameters like $E_{oed}^{ref}$, and require unloading-reloading stress-strain data to estimate $E_{ur}^{ref}$, this approach capitalizes on the proposed multi-LSTM's ability to detect intricate patterns solely from triaxial loading test data.

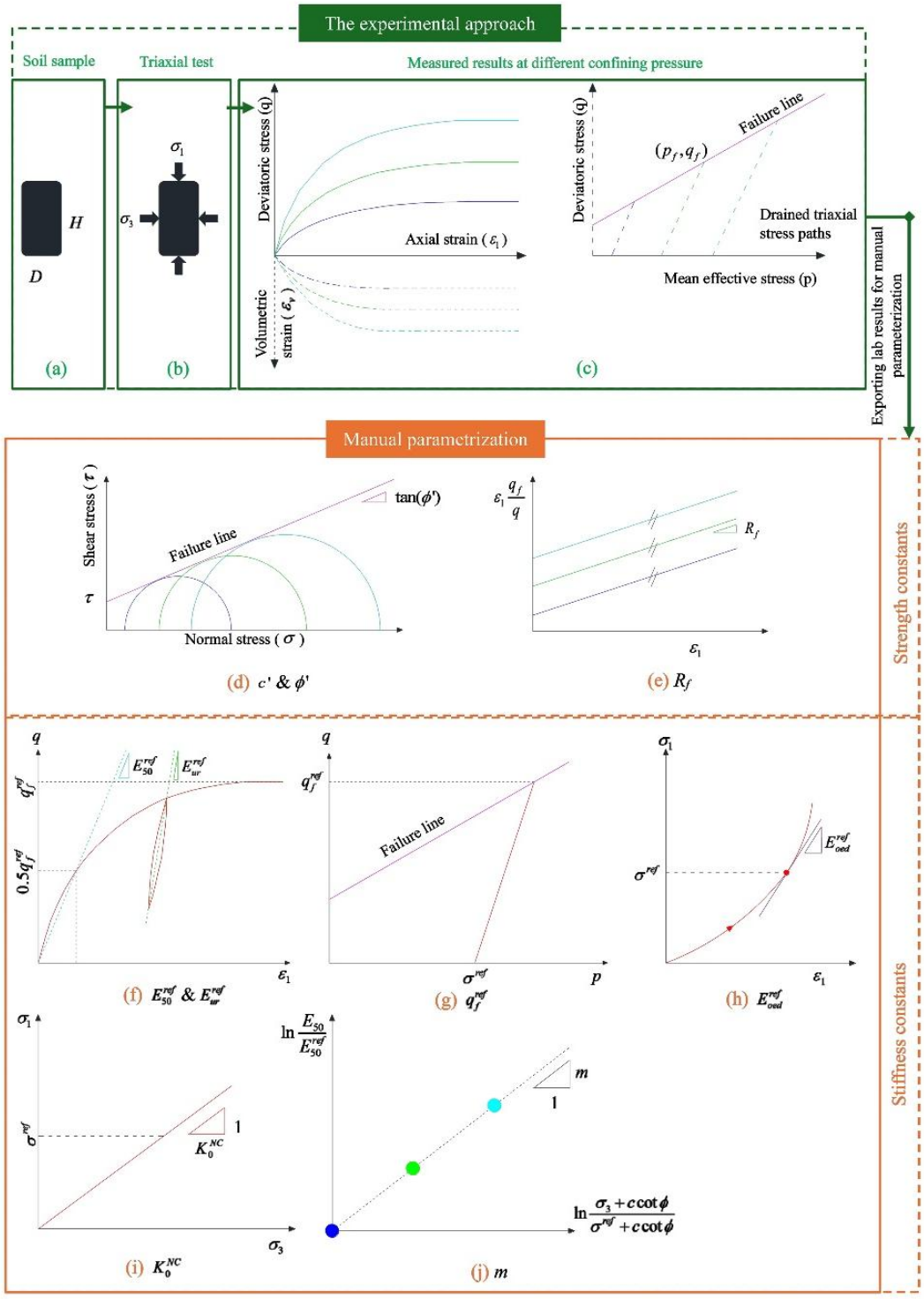



*Figure 3.14. Traditional parameterization workflow: (a–c) the experimental approach; (d–e) strength constants parameterization; and (f–j) stiffness constants parameterization.*

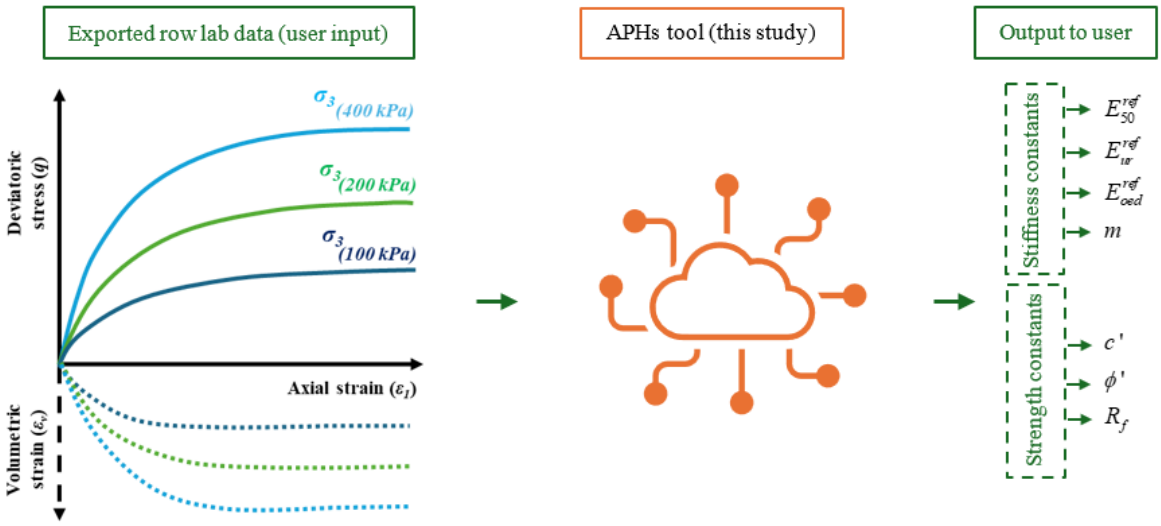



*Figure 3.15. The workflow of the proposed APHs tool.*

### b) APHs model development

This study aims to ensure the tool's practical applicability and accurate predictions by establishing an appropriate parameter range. This approach is essential for optimizing the dataset size and ensuring adequate search space coverage, which is critical given that even minor prediction errors can lead to significant simulation deviations. With this in mind, Table A6 summarizes the parameter ranges applied in this study (strength parameters: $c'$, $\phi'$, $R_f$, $\psi$, and stiffness parameters: $E_{oed}^{ref}$, $E_{50}^{ref}$, $E_{ur}^{ref}$, $m$, $\upsilon_{ur}$) which are determined as follows: Field data, laboratory tests, and relevant literature defined the boundaries for strength parameters: $c'$, $\phi'$, and $\psi$. In May 2023, field sampling was conducted at multiple locations across Hong Kong to determine the basic soil properties of the shallow layers. A total of 33 samples were collected from 11 different locations. Laboratory tests conducted by ETS-Testconsult Ltd. (2023) indicate that the samples primarily consist of moist, brown, silty clayey gravelly sand. These samples were subsequently tested using small shear box tests under normal 20, 30, and 50 kPa stresses to determine their cohesion and internal friction angle. The median values were 3.67 kPa for effective cohesion and 30.5° for the internal friction angle, as shown in Figure A3. The range of these parameters is consistent with findings from the literature. Maunsell Geotechnical

Services Ltd. (2007) reported cohesion values ranging from 0 to 8 kPa and internal friction angles between 31° and 36° for shallow colluvium and completely decomposed tuff layers. Zhou (2008) noted that typical completely decomposed granite (CDG) soils in Hong Kong exhibit an effective cohesion of approximately 0 kPa and an internal friction angle of around 35°. Wong (2020) recommended values of 5 kPa for cohesion and 36° for the internal friction angle for fill/colluvium soils based on extensive testing. Observations from the small shear box tests indicate that dilation can reasonably be assumed to be negligible (Zhou, 2008).

Concerning stiffness parameters (i.e., $E_{oed}^{ref}$, $E_{50}^{ref}$, $E_{ur}^{ref}$, $\upsilon_{ur}$), relevant literature guided the establishment of the parameter boundaries (refer to Table A6). Local studies, such as those by Lee et al. (2008), Lam (2018), and Lu et al. (2018), reported typical ranges for soil parameters suitable for modeling. Specifically, they found that $E_{oed}^{ref}$ = 0.7 to 1 $E_{50}^{ref}$, while $E_{ur}^{ref}$ = 2 to 3 $E_{50}^{ref}$. For example, Lu et al. (2018) reported values of 20.78 MPa for $E_{ur}^{ref}$, 7.573 MPa for $E_{oed}^{ref}$, and 10.39 MPa for $E_{50}^{ref}$ for a CDG layer, with $\upsilon_{ur}$ of 0.2 and $m$ = 0.517. Lam (2018) considered values of 10.5, 45, and 15 MPa for $E_{oed}^{ref}$, $E_{ur}^{ref}$, $E_{50}^{ref}$, respectively, while Chan et al. (2021) used 5 MPa for both $E_{oed}^{ref}$ and $E_{50}^{ref}$, with 15 MPa for $E_{ur}^{ref}$ in a fill layer. The minimum and maximum values listed in Table A6 were determined to ensure reasonable coverage of the parameter values discussed above. For $c'$, the range of 0 to 8 kPa was adopted. Similarly, $\phi'$, since all literature suggests values greater than 30°, which aligns with the laboratory tests, lower values (i.e., below 28°) were not considered. This provides a comprehensive range between 29° and 36°. $E_{ur}^{ref}$ was chosen to cover a broader range (10 to 50 MPa) than typically reported. For parameters that lack a clearly identified range, such as $R_f$, m, and $\upsilon_{ur}$, a wider range have been adopted to consider all possible scenarios. Most importantly, to ensure that the generated search space aligns with the theoretical assumptions, the maximum values of $E_{50}^{ref}$ are capped to $0.5E_{ur}^{ref}$ and, similarly, $E_{oed}^{ref}$ assumed to fall within the range of 0.5 to 0.8 $E_{50}^{ref}$.

These soils are assumed to be fully saturated, normally consolidated, and non-dilative (Ng et al., 2001; Zhou, 2008). These conditions establish a solid framework, with future work expanding it to a broader range of scenarios.

### c) Latin hypercube sampling (LHS)

Modeling all scenarios within the parameter range specified in Table A6 requires an extensive number of samples, making manual sampling impractical. To address this, the study utilizes LHS (McKay et al., 1979), a statistical method that ensures efficient and comprehensive parameter space coverage. LHS is a stratified sampling technique that ensures each parameter's range is divided into ($n$) intervals of equal probability, with one sample taken from each interval. For a given parameter ($P_i$) with a range $[P_{i,\min}, P_{i,\max}]$, the sampling values are defined as illustrated in Equation 3.21. This process is repeated for all parameters, ensuring an evenly distributed and non-overlapping sampling across the multi-dimensional parameter space. LHS achieves efficient coverage, reducing the number of simulations needed compared to simple random and Monte Carlo (MC) sampling methods (McKay et al., 1979; Wang, 2022).

$$P_i^j = P_{i,\min} + \left(P_{i,\max} - P_{i,\min}\right).\frac{j - r_i}{n}, j \in \{1,2,...,n\} \qquad \textit{(Equation 3.21)}$$

Where $j$ is the sample index, $r_i$ is a random value in [0,1), ensuring uniform randomness within each interval, and $n$ is the number of samples

### d) Numerical model for triaxial test

The samples generated using LHS within the parameter ranges specified in Table A6 were simulated using an axisymmetric finite element model in SIGMA/W (Geostudio, 2023c). Key considerations include a) Modeling strain up to 40% to cover a wide range of behavior; b) Ensuring all parameter sets reached apparent failure before 40% strain; sets not meeting this condition were adjusted manually; c) Conducting triaxial simulations under confining pressures ($\sigma_3$) of 100, 200, and 400 kPa; d) Extracting $q$ and volumetric strain ($\varepsilon_v$) at $\varepsilon_1$

increments of 0.4%. Extracting $\varepsilon_1$ in small increments helps represent each curve with enough points, minimizing the impact of any potentially noisy data. Figure 3.16 illustrates the model's boundary conditions and dimensions to simulate a conventional drained triaxial test. The model is built as an axisymmetric 2D simulation (in an *r–z* coordinate) that leverages the assumption of rotational symmetry about the axis. This allows it to capture the essential 3D behavior without the full computational cost of a complete 3D simulation.

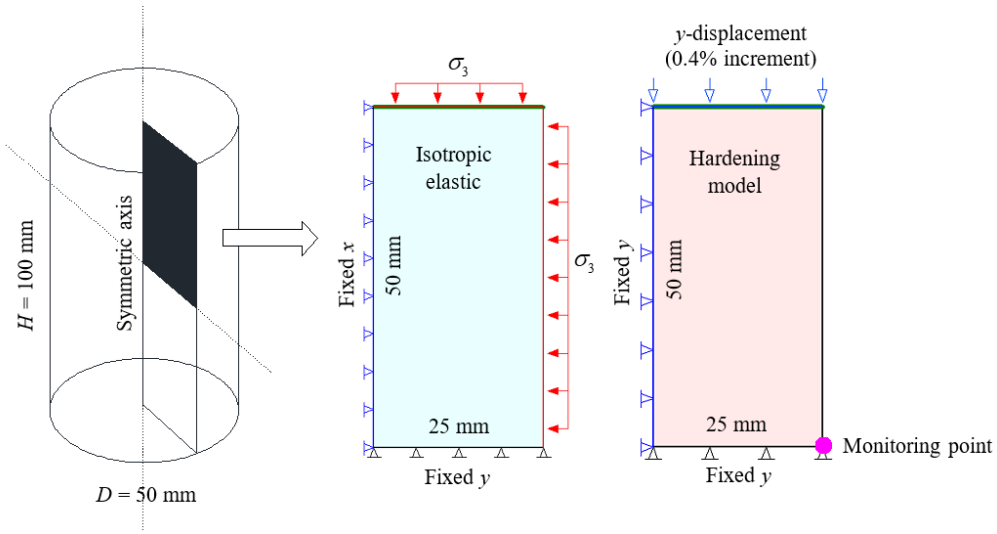



*Figure 3.16. Axisymmetric finite element model (SIGMA/W)*

### e) LSTM model development

The LSTM model was developed and trained using an augmented dataset generated within the parameter ranges outlined in Table A6. Each dataset sample, representing a simulated triaxial test, comprises six curves: $q$ versus $\varepsilon_1$ and $\varepsilon_v$ versus $\varepsilon_1$, each under $\sigma_3 = 100, 200, 400$ kPa . The temporal dependencies within these curves make the LSTM architecture a suitable choice. The model aims to predict strength and stiffness parameters for the HS model, but several challenges necessitate tailored pre-processing and design considerations: a) The varying scales of parameters (e.g., $E_{50}^{ref}$ vs. $c'$) require normalization to ensure consistent training; b) The distinct nature of predicted parameters necessitates a multi-model approach for optimal accuracy; c) Some parameters, like $c'$, may include zero values, requiring metrics that account

for sparsity; d) Triaxial test measurements used as model inputs (no additional feature selection process is necessary), with no inter-sample temporal dependencies (i.e., no overlap between samples); e) Each sample has inherent temporal dependencies within its features. The following subsections detail critical steps in developing the LSTM model:

### f) Data preparation

The data preparation process is structured into key steps to ensure compatibility with the LSTM model, optimizing its performance through proper pre-processing. These steps are described as follows: a) Normalization: input features $\boldsymbol{X}$ and target variables $\boldsymbol{y}$ are standardized to ensure consistency and mitigate sensitivity to varying scales of the features. Standardization transforms values into a standard normal distribution with zero mean and unit variance (LeCun et al., 2015); b) Reshaping and temporal ordering: input features are reshaped into a three-dimensional format, to capture the sequential and temporal dependencies, represented by axial strain increments, required by the LSTM model; c) Multi-model approach: seven distinct LSTM models are trained, each focusing on predicting a single HS model parameter per case. This approach emphasizes the unique characteristics of each target and applies weighted losses to account for their differing scales and importance (Bengio, 2012); d) Data splitting and randomization: case indices are shuffled to ensure a random distribution, minimizing biases from ordered data. The dataset is then split, with the validation set comprising 20% of the training set, maintaining representativeness.

### g) Feature selection

This study utilizes triaxial test results as input features, as these measurements directly capture the physical behavior of soil under varying confining pressures. Given these features' inherent physical meaning and well-established relationships with the target parameters, no additional feature selection process is necessary. The input dataset includes stress-strain curves: $q$ versus $\varepsilon_1$ and $\varepsilon_v$ versus $\varepsilon_1$ under $\sigma_3 = 100, 200, 400$ kPa, and $\upsilon_{ur}$. Each triaxial test is considered

an independent sample, where temporal dependencies are present only within the sample itself. There are no temporal dependencies between different samples. To quantify the correlation between each feature and the target parameters, feature importance was assessed using the permutation method. This approach evaluates the effect of each feature on model performance through the permutation of its values and observing the resulting change in error (Fisher et al., 2019). By comparing the baseline error with the error after permuting each feature, the importance of each feature was quantified. The error was evaluated using mean absolute error (MAE). This approach is ideal for time-series models like LSTM, as it preserves the temporal structure of the data, making it more suitable than model-agnostic methods like SHAP, which are typically used for static datasets.

### h) Model structure, training, and optimization

Figure 3.17 presents a schematic view of the model architecture designed for predicting a single output parameter. The input, representing a triaxial test, comprises numerical variables (6 temporal triaxial curves and 1 constant feature $\upsilon_{ur}$). Temporal variables are sampled at regular axial strain intervals to maintain a consistent window size across all samples. The pre-processed data was reshaped into a three-dimensional tensor. This study employs seven parallel LSTM networks, each following the architecture depicted in Figure 3.17, to accurately predict the seven HS model parameters. The design, training, and optimization processes of the model are detailed in the subsequent lines: To enhance the performance of the LSTM model, each hidden layer employed the Leaky Rectified Linear Unit (Leaky-ReLU) activation function to improve the model's nonlinear characteristics. The Leaky-ReLU addresses the "dying ReLU" problem by allowing a slight, non-zero gradient for negative input values (Qi et al., 2024). This study incorporated dropout and *L2* regularization techniques to mitigate overfitting in the LSTM model. Applying both methods can improve the generalization performance on unseen data. Dropout minimizes the risk of neurons becoming overly reliant on one another, while *L2*

regularization ensures that the weights remain small (Srivastava et al., 2014). The Mean Squared Error (MSE) was utilized as the loss function, while the Adam optimizer was employed for model training, as it is more adaptable and efficient. An Early Stopping criterion was applied to select the best-performing model weights and prevent overfitting or underfitting issues. Learning curves were analyzed to monitor the model's performance during training and validation. The hyperparameters of the LSTM model, such as the number of neurons, dropout rate, the *L2* parameter, the batch normalization technique, and the Leaky-ReLU parameter, were determined through the Bayesian optimization technique (Snoek et al., 2012) with a search space outlined in Table 3.7.

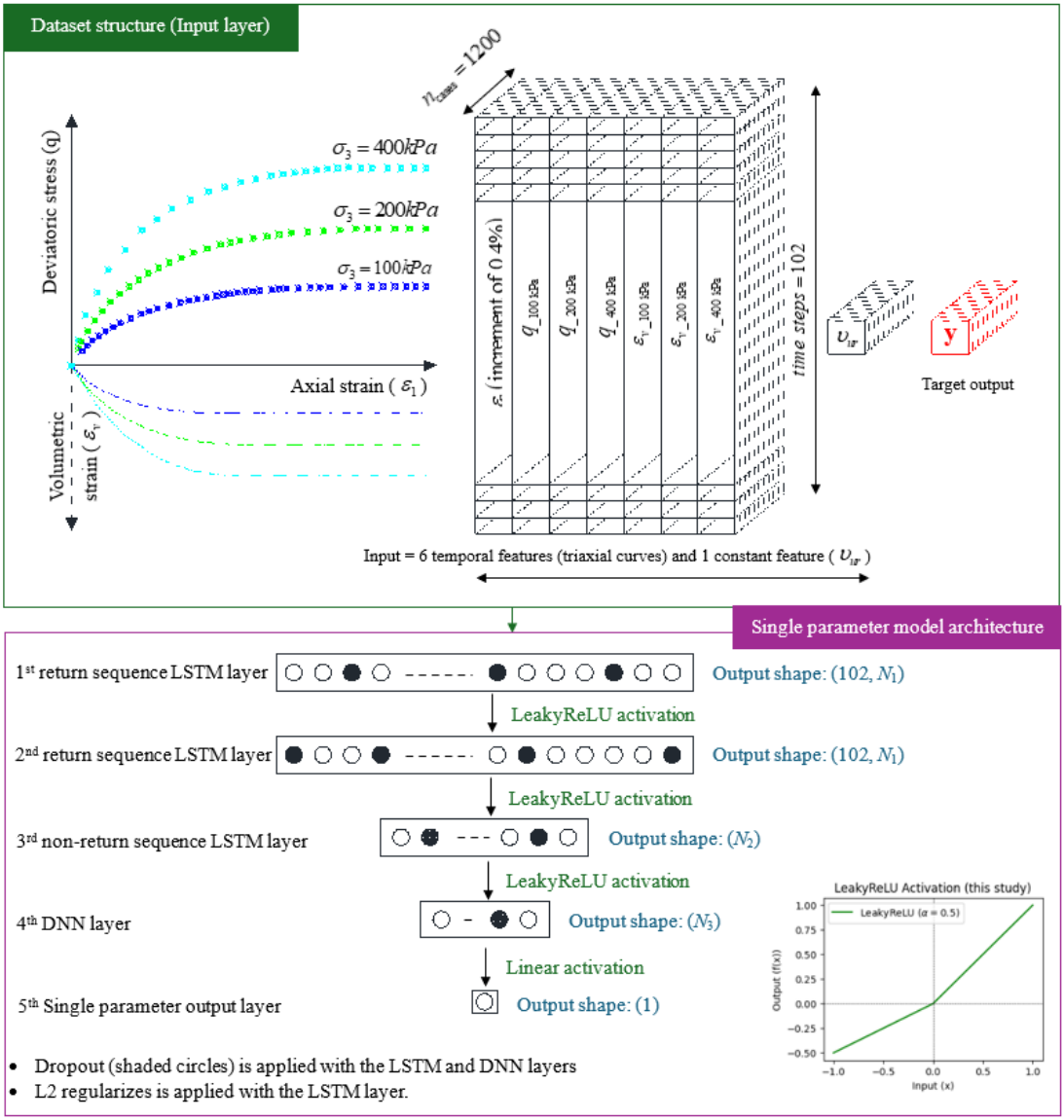



*Figure 3.17. APHs model architecture for a single parameter.*

*Table 3.7. APHs model design parameter search space*

| Parameter | Search space/value |
|---|---|
| Features | (6) temporal features (triaxial curves) and (1) constant feature ( $\upsilon_{ur}$ ) |
| Window size ( $\varepsilon_1$ _range) | (102) |
| Hidden layers (Hidden units) | (2) LSTM return sequence ($N_1$: 256, 512, 1024 units) |
| | (1) LSTM non-return sequence ($N_2$: 128, 256, 512 units) |
| | (1) Deep neural network (DNN) ($N_3$: 64, 128, 256 units) |
| Loss function | MSE=$\frac{1}{n}\sum_{i=1}^{n}(y_i-\tilde{y}_i)$ Where $y_i$ represents the actual value, $\tilde{y}_i$ is the predicted value, and $n$ is the number of data points. |
| Optimizer | Adam |
| Dropout | ($D$: 0.1, 0.3, 0.5, 0.7, 0.9) |
| LeakyReLU | ($LR$: 0.1, 0.3, 0.5, 0.7, 0.9) |
| $L2$ regularizers | ($L2$: 0.0001, 0.001, 0.01, 0.1) |
| Batch normalization | ($B$: True, False) |

### i) Model evaluation

For this objective, three primary metrics were employed to evaluate model performance: MAE, Mean Absolute Percentage Error (MAPE), and Normalized Mean Absolute Error (NMAE). These metrics assessed the model's accuracy in predicting each parameter. MAPE measures the error percentage relative to the actual values, as described in Equation 3.22. However, MAPE can be sensitive to small or large values of ( $y_i$ ), leading to disproportionate errors. To mitigate this issue, the NMAE is proposed. NMAE normalizes the MAE over the parameter range, providing a more balanced metric regardless of the parameter's magnitude (Equation 3.23). Lower values of MAE, MAPE, and NMAE indicate higher accuracy, with values approaching zero being ideal. Additionally, the coefficient of determination ($R^2$) was used to quantify prediction accuracy further. $R^2$ is a statistical measure of how well the predicted values align with the actual values. A higher $R^2$ value (closer to 1) indicates better predictive performance, while a lower value signifies worse accuracy. The Monte Carlo (MC) dropout is integrated to quantify predictive uncertainty (Gal & Ghahramani, 2016) (Equation 3.11). It is important to note that while all these metrics offer valuable insights into model reliability, MAE

is beneficial because it quantifies errors in the same units as the parameter, thereby facilitating direct deterministic geotechnical interpretation.

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \tilde{y}_i}{y_i}\right| x\ 100 \qquad \textit{(Equation 3.22)}$$

$$\text{NMAE} = \text{MAE} / \left(y_{\max} - y_{\min}\right) \qquad \textit{(Equation 3.23)}$$

Where $y_i$ represents the actual value, $\tilde{y}_i$ is the predicted value, $n$ is the total number of predictions, $y_{\max}$ and $y_{\min}$ represent the maximum and minimum values of the parameter range, respectively, and $\bar{y}$ is the mean of the observed data.

### *3.4.2. Numerical Slope Stability Modeling (Focused Site Specific Analysis)*

#### a) Background and rationale

The main objective of this research is to develop an accurate, practical, and cost-efficient approach to reducing landslide risk. One way to achieve this aim is an exact simulation utilizing numerical analysis in which the following questions, which have rarely been addressed earlier, are covered, underlining the uniqueness and relevance of this study: Why do specific locations report landslides in particular areas? Where should the monitoring locations be distributed? How does the data frequency affect the temporal prediction? This objective addresses the earlier questions by developing a rigorous analysis framework to predict pre- and post-failure mechanisms. The framework employs non-, partially-, and fully coupled analyses using the commercial packages Geo-Studio (2023) and MPM (Anura3D, 2023).

#### b) Numerical modelling process

Figure 1.7 shows the analysis framework where the modeling process began with: **a)** seepage analysis using SEEP/W (Geo-Studio, 2023a). The soil water characteristic curve and the hydraulic conductivity functions are back-calculated to match the field-monitored data. Then, the pore water pressure will be exported to the SLOPE/W using Bishop analysis with the entry-exit method boundary condition to calculate the deterministic FOS (Geo-Studio, 2023b). This

step validates and back-calculates the hydraulic characteristics of the landslide owing to the convergence advantages of uncoupled analysis, considering the actual real-time random variation of the external triggering. The hydraulic characteristics calculated from the uncoupled analysis are exported to the coupled analysis, which employs an implicit integration approach known as SIGMA analysis (Geo-Studio, 2023c); **b)** Then, the constitutive model is parameterized from laboratory tests and field deformation monitoring to back-calculate the mechanical characteristics. This process is automated as illustrated in Section 3.4.1 by coupling the APHs tool with numerical analyses, while real-time sensor measurements (Section 3.3) will dynamically update the model. For identified historical landslides, the fieldwork, such as the slip surface determined by the trial pits, is used to validate the results. Next, stresses and water pressure are exported to SIGMA to analyze the local and global FOS, considering the coupled mechanism of the water pressure and slope deformation. This step validates and back-calculates the mechanical characteristics of the landslide, providing accurate predictions. This process provides novel insights into the spatial distribution of landslides, determining optimal monitoring locations and evaluating the impact of data frequency on temporal prediction; and **c)** Unsaturated material with two phases with suction effect MPM is utilized in this study as a large deformation analysis to simulate the travel distance and debris velocities (Ceccato et al., 2021). The mechanical and hydraulic characteristics from the small deformable analysis are exported to the MPM. The initial conditions (groundwater level and hourly infiltrated rainfall) are imported from the seepage non-deformable and small-deformable analysis before failure (Soga et al., 2016). Field experiments, such as the debris flow velocity and travel distance, were used to validate the model.

c) Parametric analysis

Additionally, a parametric study is proposed to account for the regional geometric effect aiming to identify the vulnerable locations for effective monitoring with the following reasonable

assumptions (refer to Figures 3.18 and 3.19): a) Local single slope to 10 $km^2$; b) has similar hydromechanical characteristics; c) the spatial variation of the triggering feature is neglected; d) the surface geometric effect is neglected (i.e., surface natural drainage channels); e) the geometry is simplified to be axisymmetric converging or diverging neglecting the complex field geometry, investigating how the ratio R/r controls the regional geometric effect. Furthermore, knowing that rainfall dynamics are controlled by intensity, duration, non-rainfall durations, and patterns (the pattern is controlled by data frequency (i.e., the pattern of hourly time steps is different from daily or monthly ones), different hourly and daily rainfall patterns, and several patterns, durations, intensities, and non-rainfall periods were numerically modeled (refer to Figure 3.20). This will aid in selecting the appropriate frequency rate that achieves optimum power usage with reasonable accuracy. Additionally, this study will help define the optimum antecedent period that will improve the I-D threshold accuracy.

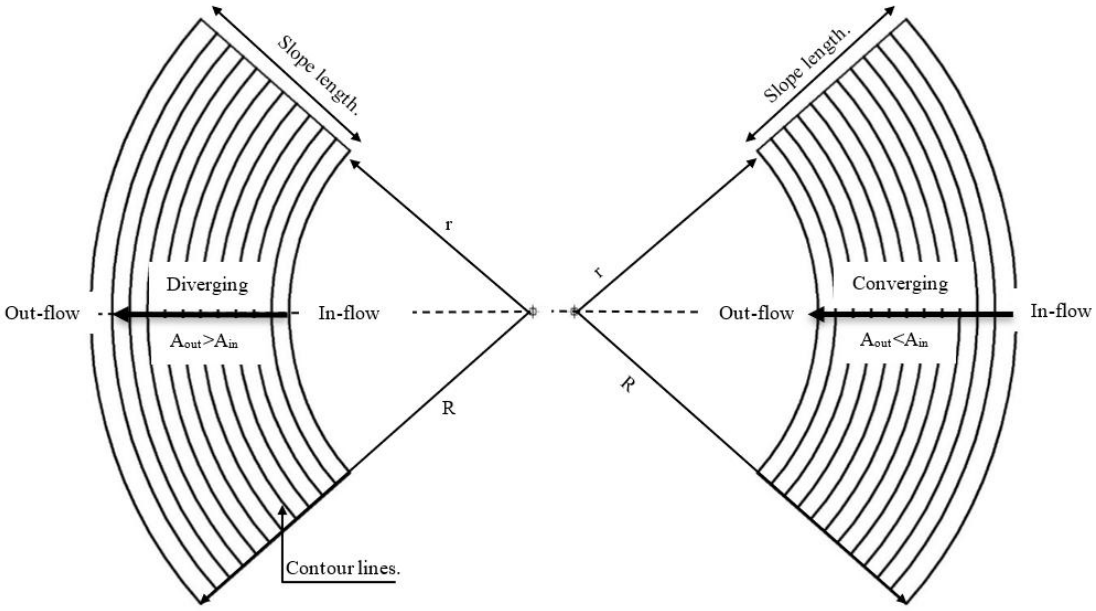



*Figure 3.18. Schematic view of the parametric study highlighting the converging and diverging boundaries.*

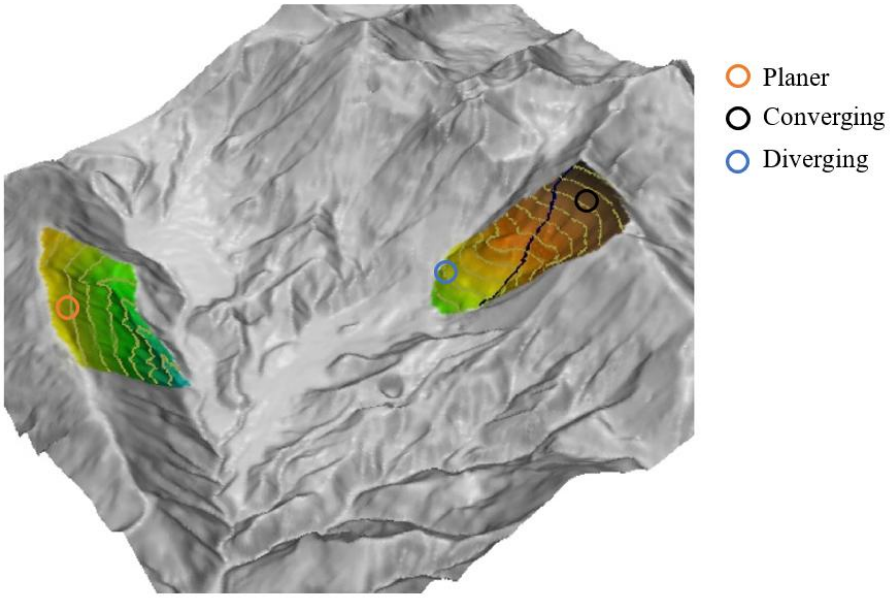



*Figure 3.19. 3D schematic view of curvature features (modified from Giuseppe et al., 2016).*

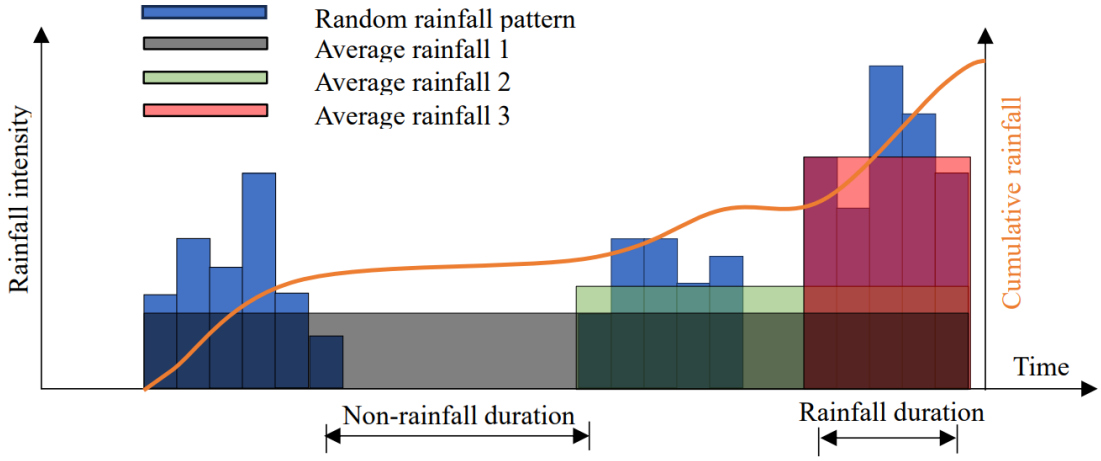



*Figure 3.20. Schematic view of the rainfall controlling features.*

### e) Non-coupled analyses

This analysis utilizes the simplified Bishop approach with a modified Mohr-Coulomb failure criterion that allows for shear strength change owing to matric suction in the limit equilibrium analysis to achieve the framework mentioned above. The governing differential equation for saturated and unsaturated water flow through a two-dimensional unsaturated soil is illustrated throughout Equation 3.8 (Lam et al., 1987). This process can be well represented by the VWC and hydraulic conductivity functions (Fredlund & Xing, 1994). The change in pore water

pressure can be attributed to the change in VWC ( $\partial\theta_w = m_v \partial u_w$ ) (Lam et al., 1987). To predict shear strength on unsaturated soil, Fredlund et al. (1978) developed a linear relationship from the Mohr-Coulomb failure criteria (Equation 3.7). Vanapalli et al. (1996) developed a non-linear shear strength equation using the normalization of the VWC function (Equation 3.24).

f) Partially-coupled numerical analyses

Changes in climatic conditions or infiltration-related effective stress lead to volume changes in the coupled stress and pore water pressure analysis. Equations 3.25 (load and deformation equation) and 3.26 (pore pressure and flow equation) are the two equilibrium equations. Equation 3.27 results from a coupled analysis where the seepage and deformation analyses are solved by a numerical combination of the two equations (Geo-Studio, 2023c). The VWC of the saturation zone in the uncoupled analysis is constant and equal to the porosity of the soil. In the coupled analysis, the volume compressibility $m_v$ (Equation 3.28) equals the slope of the saturation zone curve.

$$\tau = c' + \sigma_{net} \tan\phi' + \sigma_s \, S_e \tan\phi' \qquad \textit{(Equation 3.24)}$$

$$[K_s]\{\Delta d\} = \{\Delta F\} \qquad \textit{(Equation 3.25)}$$

$$\left[[K_s][L]/[L]^T[K_w]\right]\{\Delta d / \Delta u\} = \{\Delta F / \Delta Q\} \qquad \textit{(Equation 3.26)}$$

$$[K_w]\{\Delta u\} = \{\Delta Q\} \qquad \textit{(Equation 3.27)}$$

$$m_v = \frac{3(1-2\upsilon)}{E} \qquad \textit{(Equation 3.28)}$$

$\tau$, $c'$, $\phi'$, $\sigma_{net}$, and $\sigma_s$ represent the shear strength, effective cohesion, effective friction angle, net normal strength, and matric suction, respectively. $S_e$ represents the effective degree of saturation. $K_s$ and $K_w$ denote the material and element characteristics of the solid and water states. F means the nodal load or force, d represents the nodal deformation, Q is the nodal flow, u is the nodal pore water pressure, and L is a coupling matrix. E is the modulus of elasticity, and $\upsilon$ is the Poisson ratio.

g) fully-coupled analyses

The particle-based MPM considers the material as a collection of material points. Newton's laws of motion define how those points deform (Sulsky et al., 1995). The medium is classified as solid, liquid, or gas in the material point technique (Anura3D, 2023). Just one material across two phases (solid (s) and liquid (L)) is simulated, with the two unknowns being the acceleration ($a_S$ and $a_L$) (Ceccato et al., 2019). The bishop concept determines effective stress, which considers partial saturation. The equation of Van Genuchten (1980) is used to simulate the soil water characteristic curve (SWCC), whereas the equation of Mualem (1976) is used to model the hydraulic conductivity function. Fully coupled hydromechanical behavior is captured in Equations 3.29 and 3.30 using dual momentum balance equations, detailed in Anura3D (2023).

$$(1-n)\rho_S a_S = \nabla.(\sigma - nSuI) + (\rho_m - \mathrm{nSu})\,\mathrm{b} + \frac{nS\mu_L}{k.k_i}(\mathrm{V_L} - V_S) \qquad \textit{(Equation 3.29)}$$

$$\rho_L a_L = \nabla u + \mathrm{u}.\mathrm{b} - \frac{nS\mu_L}{k.k_i}(\mathrm{V_L} - V_S) \qquad \textit{(Equation 3.30)}$$

Here, n represents porosity, $I$ stands for identity matrix, $b$ for body force vector, $\rho_m$ for mixture density, $k_i$ for intrinsic permeability, $\mu_L$ for liquid dynamic viscosity, $V_L$ for liquid phase volume, and $V_S$ for solid phase volume.

*3.4.3. Automated Slope Stability Analysis Model (A Hybrid DL–Physics Model)*

a) Background and rationale

As illustrated earlier in Section 2, regional landslide forecasting has traditionally relied on physically based deterministic models like TRIGRS (Baum et al., 2008) and SINMAP (Pack et al., 1998). While theoretically robust, their operational application for real-time early warning is hindered by extensive data requirements. As a result, empirical rainfall thresholds are often used as a practical alternative, but these typically oversimplify the process by specifying only a minimum rainfall amount and fail to account for dynamic factors like antecedent soil moisture (Guzzetti et al., 2025). Although more advanced "causative

thresholds" that integrate variables such as soil suction and groundwater levels have been proposed (e.g., Bednarczyk, 2018; Cao et al., 2020), they remain challenging to apply regionally. Collectively, a clear gap exists for a methodology that can capture the physical, time-dependent response of slopes at a regional scale without the prohibitive data demands of deterministic models. This objective bridges this gap by developing an integrated approach that leverages both data-driven and physical models. The core of this novel framework is the use of DL to simulate spatially distributed saturation levels, learning the complex relationship between historical rainfall patterns and slope saturations from past landslide incidents. These predicted saturation levels are then integrated into an infinite slope stability model (Equation 3.31; Huang et al., 2025), thereby incorporating fundamental soil characteristics and mechanical principles. This technique combines the predictive power of DL with the physical rigor of stability equations, mitigating spatial uncertainties.

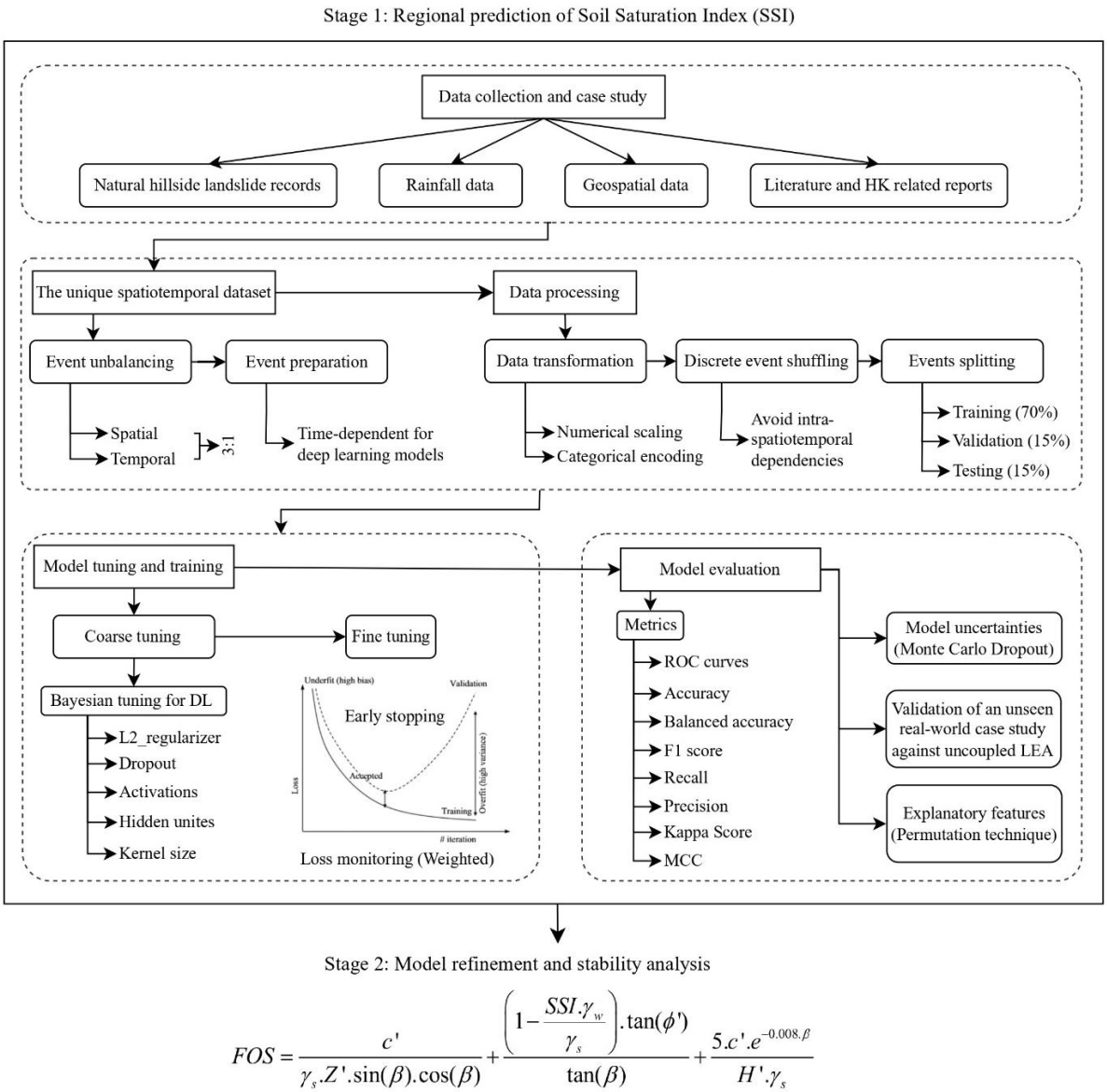
$$FOS = \frac{c'}{\gamma_s.Z'.\sin(\beta).\cos(\beta)} + \frac{\left(1 - \frac{SSI.\gamma_w}{\gamma_s}\right).\tan(\phi')}{\tan(\beta)} + \frac{5.c'.e^{-0.008.\beta}}{H'.\gamma_s} \qquad \textit{(Equation 3.31)}$$

Here, *SSI* is the soil saturation index automated through DL algorithms, *Z'* is the Vertical sliding depth, and *H'* is the vertical height from slope toe to crest.

b) Modelling process

This framework is operationalized into an interactive, user-friendly interface for early warning, using Hong Kong as a case study. The multi-data-driven model was trained on natural hillside landslide data from 1984 to 2022 (data.gov.hk) using LSTM (Hochreiter & Schmidhuber, 1997) and 1D CNN (Kiranyaz et al., 2021). Acknowledging the inherent uncertainties in data-driven forecasts, this study validates the model's probabilities against a deterministic seepage stability analysis (Geo-Studio, 2023a, b) for a real-world case: the 2005 Fei Ngo Shan Service Reservoir in Hong Kong (Maunsell Geotechnical Services Ltd., 2007). Although tailored to Hong Kong's natural slopes, the framework's methodology is adaptable to other regions with

appropriate data and parameter tuning. Figure 3.21 illustrates the overall study design. The modeling process and deployment follow the steps illustrated in Section 3.2. Table 3.8 lists the model parameter search space, and Figure 3.22 shows the model architecture for the CNN as an example. For the interface, geospatial functionality was implemented using coordinate transformations with Proj4js (PROJ contributors, 2023) in the web interface, spatial queries with Scipy's KDTree (Virtanen et al., 2020), and interactive mapping with Leaflet (Agafonkin, 2023), while the decision-support system was developed as a custom web application with an HTML frontend and Python backend.




$$FOS = \frac{c'}{\gamma_s.Z'.\sin(\beta).\cos(\beta)} + \frac{\left(1 - \frac{SSI.\gamma_w}{\gamma_s}\right).\tan(\phi')}{\tan(\beta)} + \frac{5.c'.e^{-0.008.\beta}}{H'.\gamma_s}$$

*Figure 3.21. Schematic view of the hybrid model design and workflow.*

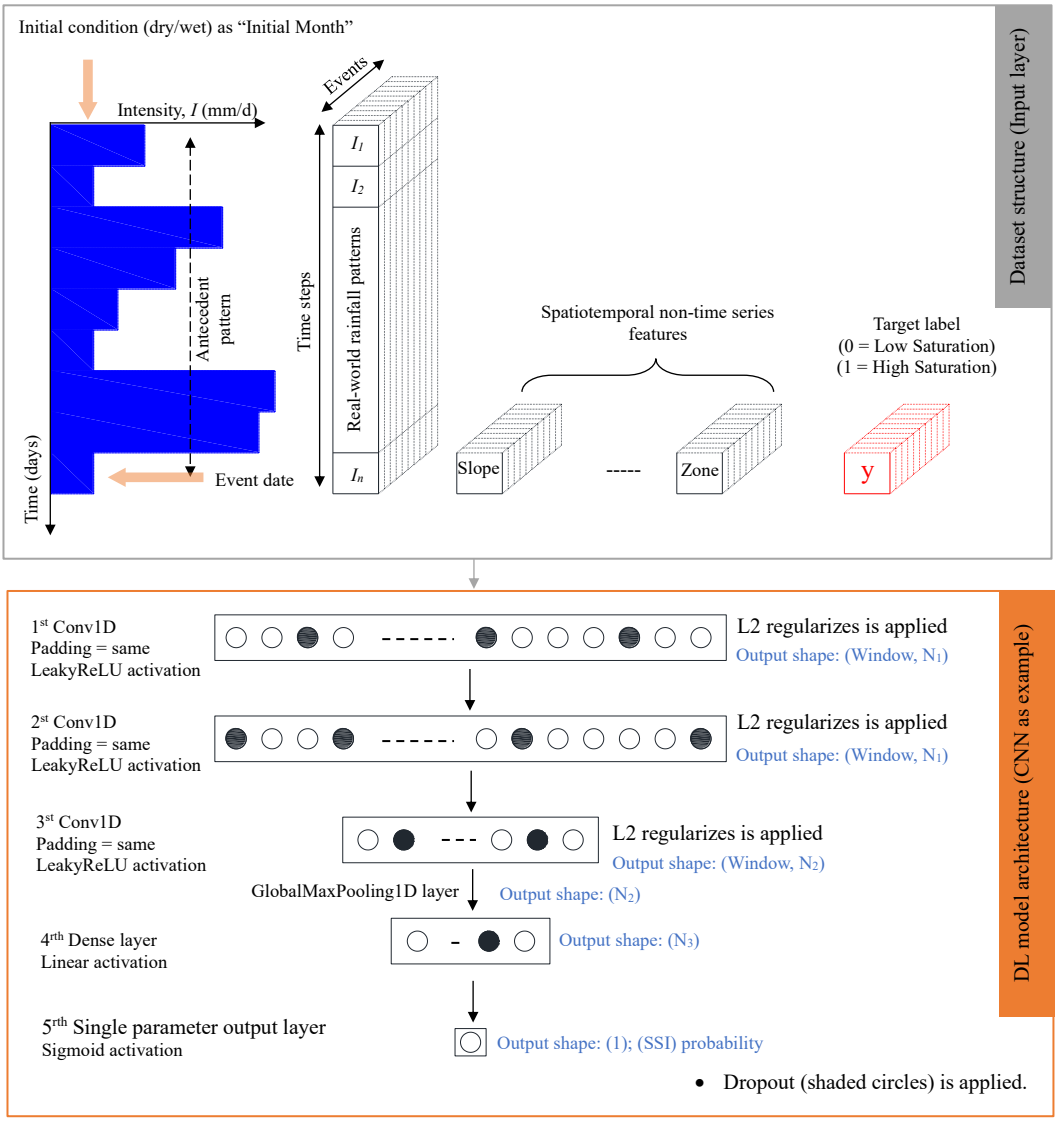



*Figure 3.22. Schematic diagram of the DL model (CNN as an example) architecture.*

*Table 3.8. DL models design parameters and search space*

| Parameter | Search space/value | |
|---|---|---|
| | LSTM | CNN |
| Features | (1) time series features (rainfall) and (9) non-time series features | |
| Window size (antecedent period) | 4 weeks (i.e., 28 days). | |
| Padding/Kernel size | - | Same/(K: 3, 5) |
| [Hidden layers] (Hidden units/filters) | [2] return sequence ($N_1$: 256, 512, 1024) | [2] Conv 1D ($N_1$: 256, 512, 1024) |
| | [1] non-return sequence ($N_2$: 128, 256, 512) | [1] Conv 1D ($N_2$: 128, 256, 512) |
| | [1] deep neural network ($N_3$: 64, 128, 256) | |
| Loss function | Binary cross-entropy $= -(1/n)\sum_{i=1}^{n}\left[y_i \log(\tilde{y}_i) + (1-y_i)\log(1-\tilde{y}_i)\right]$<br>Here, $y_i$ is the true label, $\tilde{y}_i$ is the predicted probability, and $n$ is the number of events. | |
| Loss weights | 1:2 ratio (non-landslide to landslide events, respectively) | |
| Optimizer | Adam | |
| Dropout/LeakyReLU | (D: 0.1, 0.3, 0.5, 0.7, 0.9)/(LR: 0.1, 0.3, 0.5, 0.7, 0.9) | |
| L2 regularizers | (L2: 0.0001, 0.001, 0.01, 0.1) | |

### c) Model assumptions

To ensure the model's regional applicability across Hong Kong, key assumptions and boundary conditions were defined based on the area's distinct geoenvironmental context. The region is dominated by natural hillslopes underlain by colluvium and completely decomposed granitic, volcanic, and sedimentary rocks. These materials form a shallow mantle (1–36 m) of silty, clayey, gravelly sands that are highly susceptible to rainfall-induced failure, a vulnerability exacerbated by urban development in a territory where such slopes constitute 60% of the land area (ETS-Testconsult Ltd., 2023; Hencher & Lee, 2010; Lam, 2018; Maunsell Geotechnical Services Ltd., 2007; Ng et al., 2001; Ng & Shi, 1998). A foundational assumption for model development is that historical landslide incidents occurred under conditions of high saturation. This premise allows the data-driven model to learn the critical hydrological signatures associated with failure from the historical record. The primary triggering factor is the inherent randomness of rainfall, while other spatially and temporally variable factors are treated as follows: a) Hydrological process: In Hong Kong's humid climate, the rainfall-runoff-infiltration process is simplified to infiltration and runoff components. Recognizing that each rainfall event exhibits a unique runoff ratio, this variability is accounted for by grouping data by zone and incorporating corresponding topographic features, which implicitly reflect the event-specific runoff ratio through their influence on actual infiltrated rainfall; b) Soil properties: the soils exhibit a saturated hydraulic conductivity ranging from $5\times10^{-6}$ to $3.5\times10^{-4}$ m/s (GCO, 1982), with representative shear strength parameters (cohesion ~5 kPa, friction angle ~36°) (Wong, 2020). To accommodate spatial variability, the predicted saturation levels are integrated into an infinite slope stability model (Huang et al., 2025), which assigns unique soil designations to each incident; c) Vegetation: given the difficulty in obtaining high-resolution temporal data on vegetation for each landslide incident and its relatively minor influence on triggering in this context, vegetation cover is treated as temporally uniform (Qiu

et al., 2024).To accommodate the spatial variability, cases by location were grouped, assigning each incident a designated zone where similar spatial attributes are assumed; and d) Topography: key topographic features that control infiltration and subsurface flow are explicitly incorporated into the spatial analysis.

## 3.5. SUMMARY

This chapter details the methodologies implemented to achieve the research objectives. It first presents the DL algorithms and the modeling framework developed to enable operational, first-tier landslide prediction. The framework is then adapted for sensor-based prediction, where a novel time-series modeling technique is introduced to improve monitoring-system performance. Finally, the chapter describes a methodology for automating slope-stability assessment: a standalone tool for automatic parameterization and an integrated automated system that couples DL with deterministic stability equations. Collectively, these methodologies aim to establish a robust and effective monitoring system.

# CHAPTER 4: DATA COLLECTION AND CASE STUDY

## 4.1. INTRODUCTION

This chapter describes the data collection employed in this study. It compiles a comprehensive dataset that integrates three complementary sources: historical landslide records, sensor-derived measurements, and synthetic data generated by finite element modeling. These datasets were used to train, validate, and test a predictive system designed to mitigate limitations in data availability and to capture spatial and temporal variability in landslide behavior. Below are the dataset components and their roles in the study.

## 4.2. HISTORICAL LANDSLIDE RECORDS DATASET

Focusing on natural landslides in Hong Kong, the dataset comprises events from 1984 to 2022 (data.gov.hk). Sensor-based rainfall measurements were retrieved from data.gov.hk and integrated with supplementary geospatial data, including slope, elevation, curvature, and soil type, sourced from the interactive Geotechnical Information Infrastructure platform (ginfo.cedd.gov.hk). The dataset was constructed as follows: a) For each landslide incident, the date was determined using the “Fail Date” when available; if not, the “Call Date” was used. Incidents for which spatial characteristics could not be extracted from the “ginfo” platform were excluded; b) Non-landslide events are identified as the two antecedent events preceding each recorded landslide. This creates a temporal sequence of three consecutive events: two non-landslide events followed by one landslide event. For example, if a landslide occurred on January 1, 2020, the non-landslide events are defined by the rainfall periods from October to November and November to December, while the December to January period is the landslide event. This short-term window mitigates the confounding effects of long-term temporal changes, such as vegetation dynamics or human activities;

c) Landslides are explicitly defined as events with a minimum volume of 1 $m^3$. Reported incidents falling below this threshold are classified as non-landslide cases, a distinction that

enhances the model's capacity to differentiate between susceptible and stable zones. This classification approach yields a spatially imbalanced dataset that more accurately reflects real-world landslide distribution patterns; and d) The aggregation of these sequences across all incidents resulted in a final dataset with an overall ratio of approximately 3:1 (3,733 non-landslide to 1,307 landslide cases). While a balanced 1:1 ratio is common (Wang et al., 2020b), it fails to represent the real-world imbalance where stable conditions significantly outnumber failures. Although extreme class imbalance can impair model performance (Xiao et al., 2023), the chosen 3:1 ratio represents a pragmatic compromise to better reflect real-world conditions. A systematic investigation of the optimal class ratio is designated for future work.

Figure 4.1 presents the Hong Kong map with the marked landslide incidents. In addition, it displays the non-landslide cases and the rain gauge locations. The statistical characteristics of each feature for landslide and non-landslide cases are provided in Tables A7-A8, with some selected features illustrated in Figures 4.2 and 4.3. These tables list the distribution and frequency of each feature for both landslide and non-landslide events, while Table 4.1 provides a detailed description of these features. However, it should be emphasized that the features above require manual extraction and are not automatically derived. Therefore, key topographic attributes, including slope, elevation, curvature, Topographic Roughness Index (TRI), aspect, and Topographic Wetness Index (TWI), were extracted from a digital elevation model to describe the geomorphological setting, as summarized in Table 4.1 and expanded in Table 4.2. Figure 4.4 presents: (a) a pair plot of the training dataset, and (b) key topographic attributes derived from DEM analysis, including Elevation, Slope, Aspect, Curvature, TRI, and TWI, characterizing the study area's geomorphology. The significant overlap between landslide and non-landslide features underscores the dataset's inherent complexity, necessitating advanced DL models capable of discerning subtle, non-linear patterns that conventional methods may miss.

*Table 4.1. Overview of key features and their justification*

| Feature | Condition | Description/ Justification | | Units |
|---|---|---|---|---|
| Rainfall | Dynamic | CR(1-4)W | The amount of rainfall for the antecedent (1-4) weeks | Numerical (mm) |
| | | MRI | The maximum intensity during the antecedent 4 weeks | Numerical (mm) |
| | | RD | The number of rainy days that recorded > 10 mm/day for the antecedent 4 weeks | Numerical (day) |
| | | Rainfall | The antecedent window, sensitivity selected, of daily rainfall with real patterns (Figure 4.2) | Numerical (mm/day) |
| | | The antecedent rainfall window is defined as a period of continuous daily rainfall, incorporating observed real-world patterns (see Figure 4.3). Three distinct consequence events are modeled for each incident, each maintaining its internal temporal dependencies while remaining independent. This approach represents real-world data imbalances. | | |
| Initial Month | Dynamic | This feature quantifies the antecedent moisture state preceding the rainfall event, thereby enabling the differentiation of similar rainfall patterns that stem from distinct initial conditions. | | Categorical: Jan. (1), Feb. (2), Mar. (3),... Dec. (12) |
| | | As an example, Figure 4.3 shows that some events could be initiated under high suction during dry months (i.e., April), a scenario that would be markedly different from a comparable event initiated under nearly zero suction during wet months (i.e., August). This justifies the consideration of the initial month as a boundary condition. | | |
| Slope | Static | This feature represents the average gradient measured across the area corresponding to each incident. | | Numerical (degrees) |
| | | Steeper slopes often correlate with higher landslide susceptibility. | | |
| Curvature | Static | This feature captures the localized regional geometric effect, including lateral flow, to differentiate spatial configurations under similar triggering conditions. Despite comparable triggering and initial conditions, it may address why some areas record shallow landslides while others do not. | | Categorical: Planner (0), Converging (1), Diverging (-1) |
| | | In diverging topographies, water flows from a confined area to a broader region, limiting lateral flow and restricting the rise in perched groundwater level (by mass conservation). In contrast, converging areas exhibit the opposite (Figure 3.19). | | |
| Elevation | Static | This feature denotes the vertical height of the incident's location above mean sea level. | | Numerical (meters) |
| | | Elevation correlates with variations in geology. Moreover, it affects water flow and groundwater accumulation (Guzzetti et al., 1999). | | |
| Material description | Static | The landslide material: Categorical: Soil (1), Soil & Rock (2), and rock (3) | | |
| Nearby structures | Static | The nearest structure to the slope: Categorical: Roads (1), Paths (2), Drainages (3), Residential Areas (4), Construction & Industrial Areas (5), Recreational Areas (6), Public Facility (7), Open Spaces (8), and Hillside (9). | | |
| Zone | Static | The Hong Kong map is segmented into 95 distinct zones, each represented by a rain gauge location. | | Categorical: H01 to H30 (1-30), K01 to K10 (31-40), N01 to N55 (41-95) |
| | | It is essential to factor in the spatial variability of soil properties and vegetation cover, as these elements significantly influence hydrological processes (Wong and Ho, 1996). | | |
| Zone distance | Static | The distance between the nearest rain gauge and the landslide | | Numerical (meters) |
| | | Soil variability in each zone should be accounted for. Additionally, it is essential to account for the spatial uncertainty in rainfall, as rain gauge data may not accurately represent localized rainfall intensities at the incident location (Wicki et al., 2021). | | |

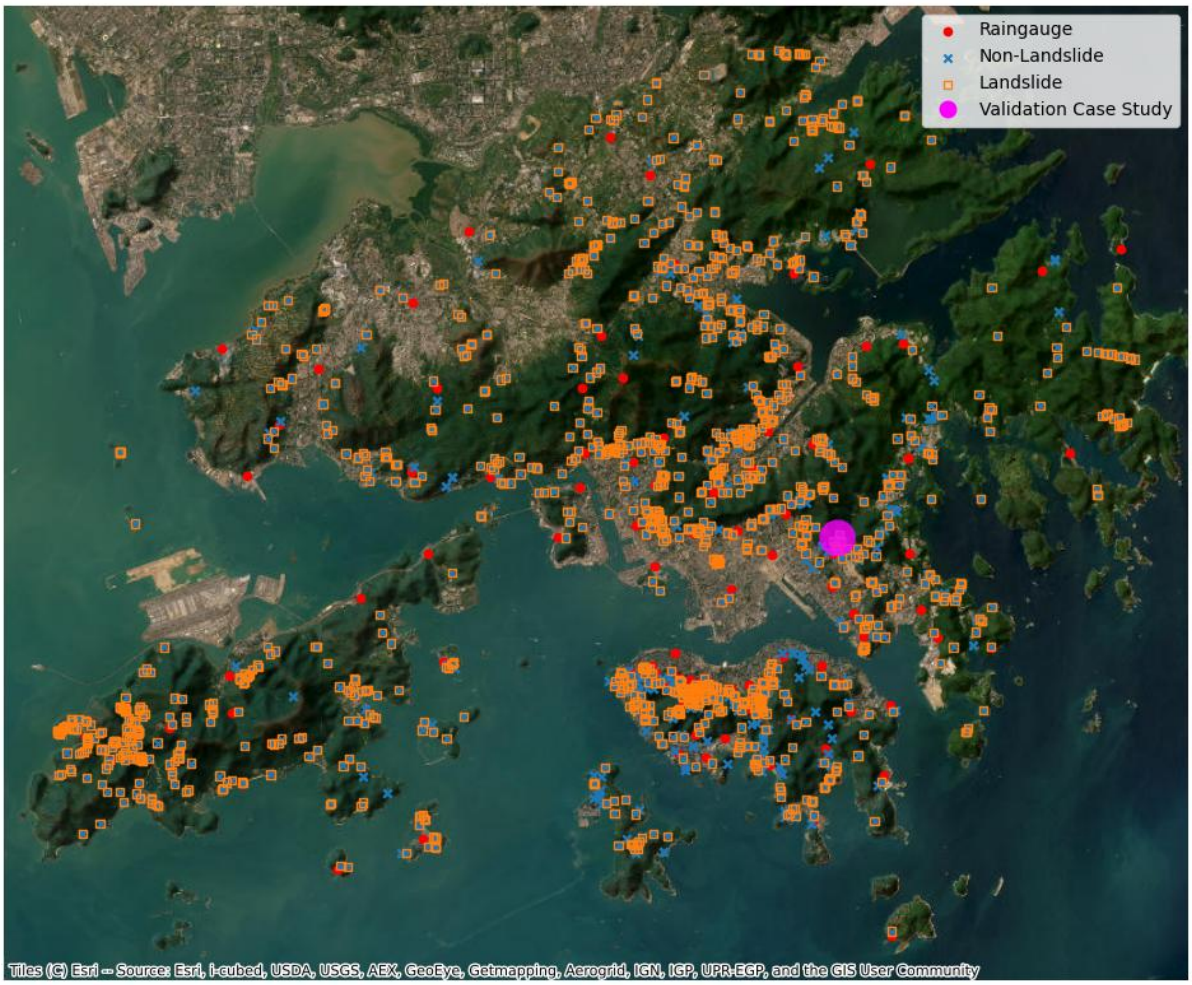



*Figure 4.1. Map of Hong Kong showing locations of the validation case study, raingauges, landslides, and non-landslide incidents. Basemap imagery provided by Esri and its partners.*

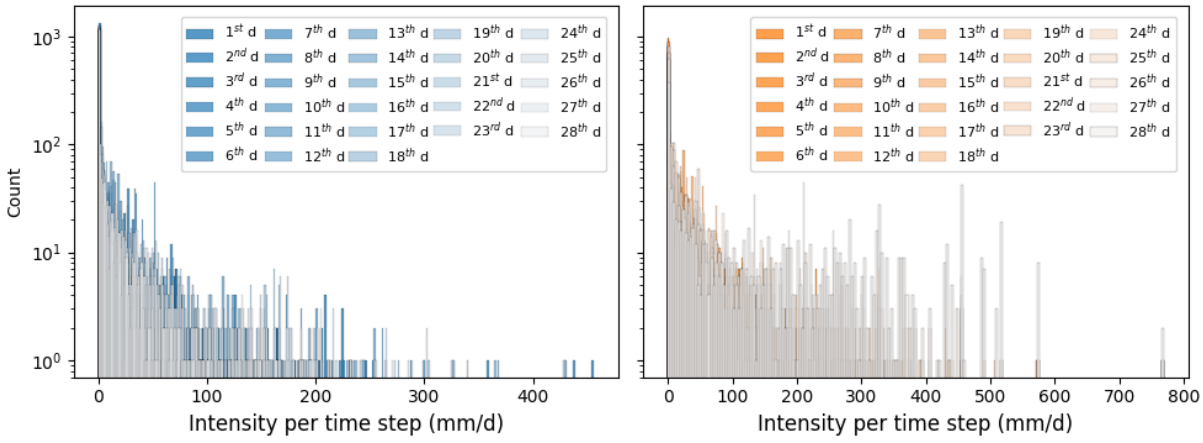



*Figure 4.2. Frequency distribution of rainfall intensity per time step. Blue indicates non-landslide cases, and orange indicates landslide events.*

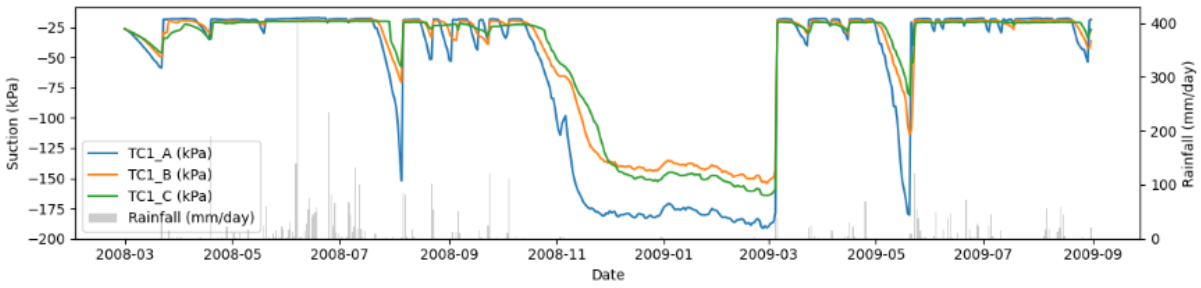



*Figure 4.3. Suction dynamics at a Hong Kong landslide (Tung Chung as an example), illustrating the Initial Month concept (Geotechnical Engineering Office, 2010b)*

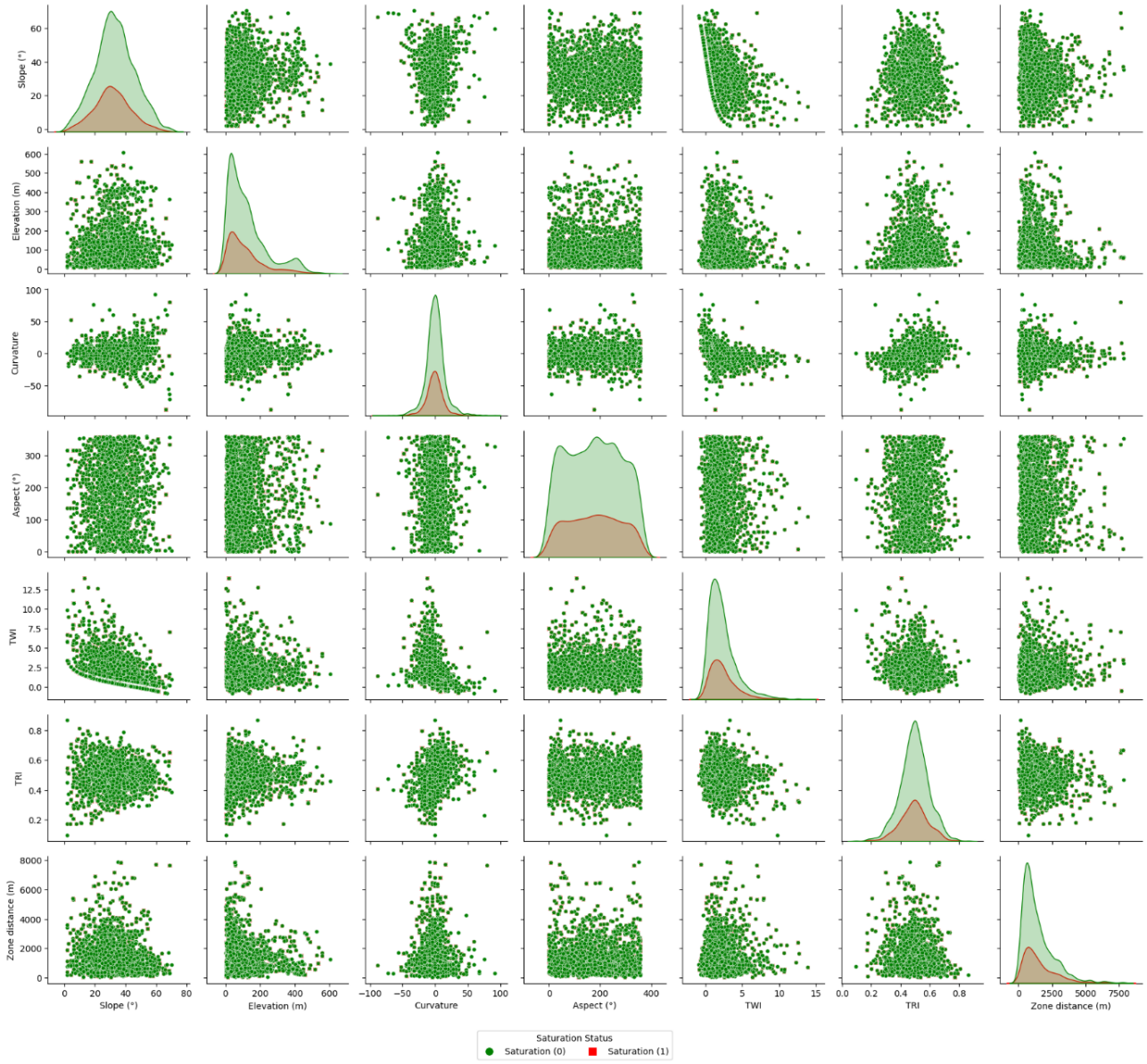



*Figure 4.4. a) Pair and frequency plot of the features extracted from the DEM.*

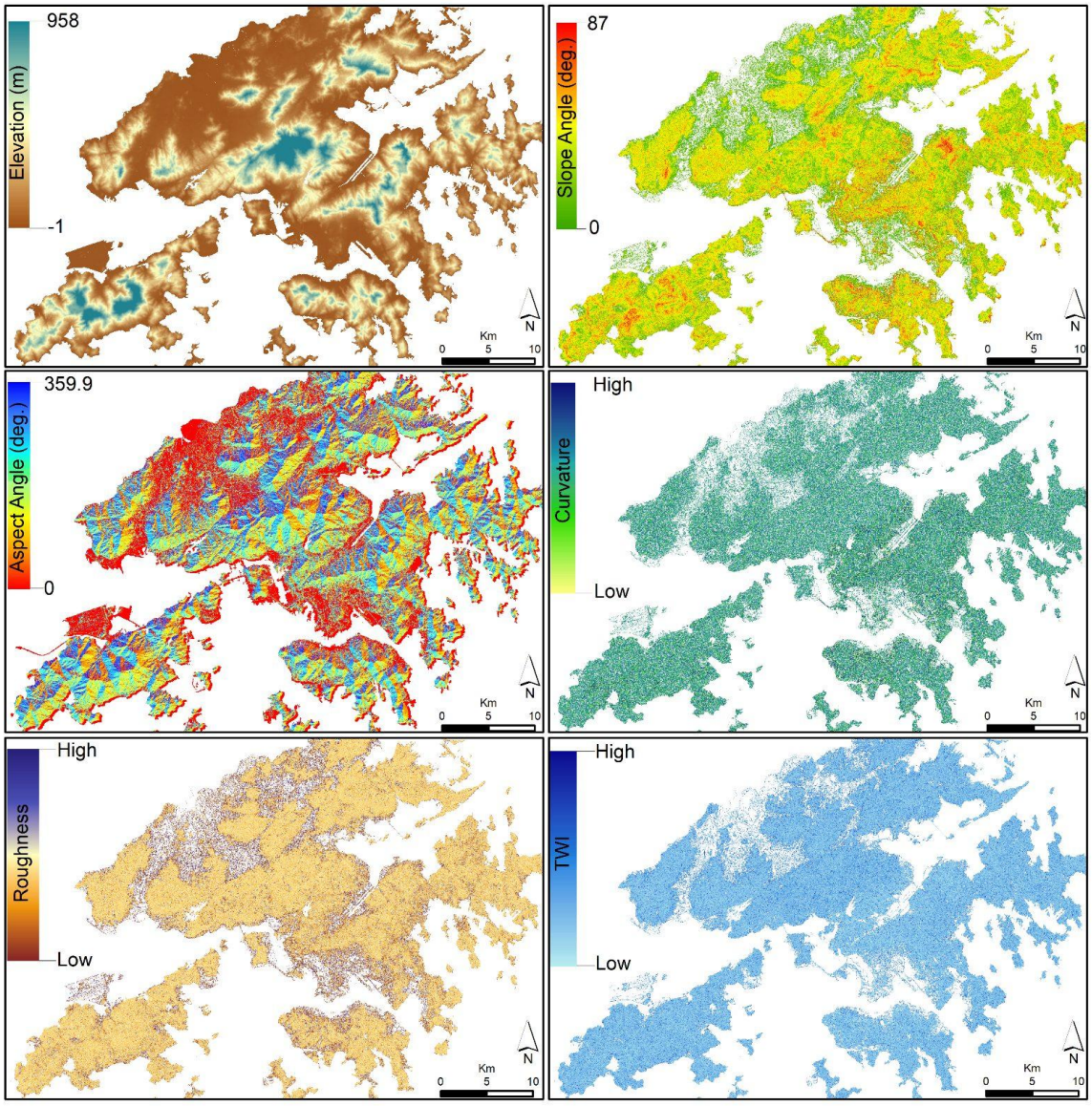



*Figure 4.4. Cont. b) Key topographic attributes from the DEM characterizing the study area's geomorphology: (a) Elevation; (b) Slope; (c) Aspect; (d) Curvature; (e) TRI; (f) TWI*

*Table 4.2. Overview of key features extracted from the DEM*

| Feature | Condition | Description/ Justification | Units |
|---|---|---|---|
| Curvature | Static | This feature quantifies the concavity or convexity of the terrain surface.<br>Concave areas (negative curvature) promote water concentration, while convex areas (positive curvature) encourage water dispersion. | Numerical (1/m) |
| TWI | Static | This feature is the Topographic Wetness Index, which models the propensity of an area to accumulate water.<br>Higher TWI values indicate a higher likelihood of saturated conditions, which can decrease slope stability. | Numerical (unitless) |
| TRI | Static | This feature is the Terrain Ruggedness Index, measuring the elevation difference between a cell and its surroundings.<br>Areas with high TRI values, indicating rough and steep terrain, are often more susceptible to erosion and shallow landslides. | Numerical (meters) |
| Aspect | Static | This feature describes the compass direction that a slope faces.<br>Aspect influences soil moisture and vegetation through differential solar exposure and wind effects. | Numerical (degrees) |

## 4.3. GEOTECHNICAL INVESTIGATION AND LABORATORY DATASET

Figure 4.5 shows the validation case study used to assess the model: The 21 August 2005 landslide at the Fai Ngo Shan Service Reservoir (Maunsell Geotechnical Services Ltd., 2007). On 21 August 2005, a landslide triggered a debris flow of approximately 3350 $m^3$. The landslide is roughly 100 m NW of the reservoir, 210 m west of Fai Ngo Shan Road, and 280 m west of Pak Fa Lam Road. Extensive field investigations and laboratory tests conducted provided the data necessary for constructing the physical model: Ground investigations (trial pits, drill holes, trenches, and slope strips) were carried out from March to May 2006, followed by monitoring installations (standpipes, piezometers) from July to September 2006 and tensiometers from March to September 2006. Laboratory tests assessing shear strength and mechanical and hydrological properties further informed the numerical modeling process, detailed in (Maunsell Geotechnical Services Ltd., 2007). Figure 4.5 shows the tensimeter locations (T1, T2, T3, and T4) at depths of 0.50 m (A), 1.00 m (B), and 1.50 m (C). Section A corresponds to tensimeters T3 and T4, while Section B designates T1 and T2. Groundwater levels, the lowest and the highest, are shown at four locations (BH1, BH2, BH3, and QP4). The soil profile is characterized by four layers: colluvium, completely decomposed tuff (CDT), medium to highly decomposed tuff (M/HDT), and medium to slightly decomposed tuff (M/SDT). These soil properties are listed in Table 4.3.

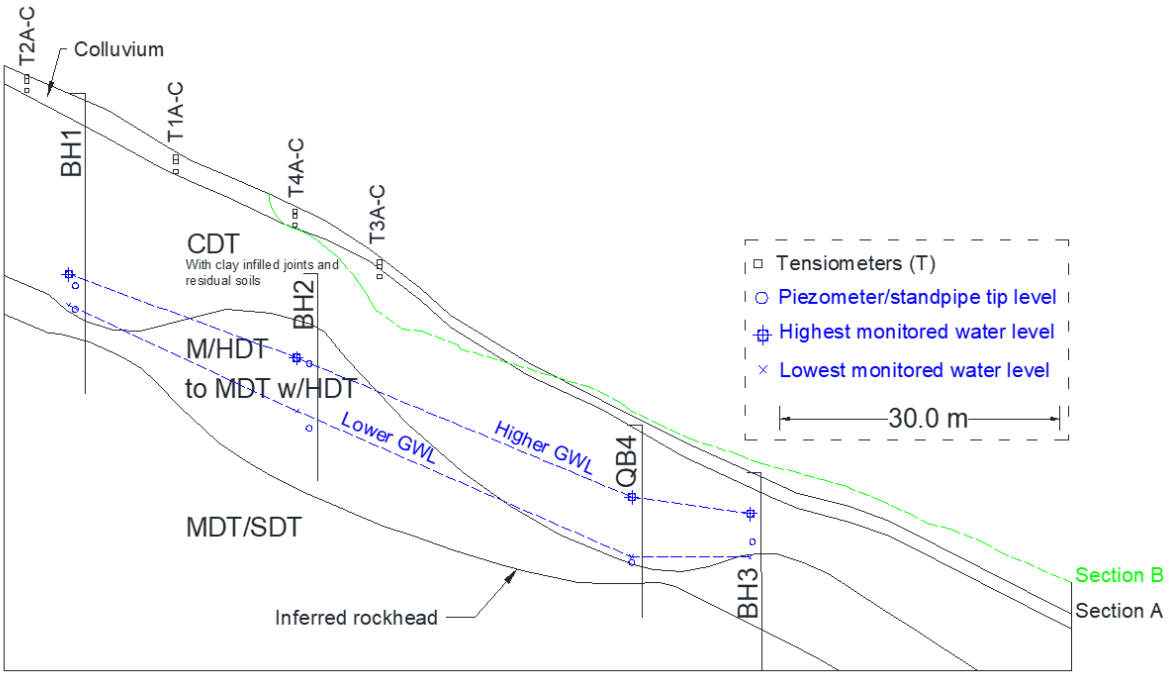



*Figure 4.5. Fei Ngo Shan Landslide soil profile cross-sectional elevation (modified from Maunsell Geotechnical Services Ltd., 2007)*

*Table 4.3. Soil characteristics at Fei Ngo Shan (Maunsell Geotechnical Services Ltd., 2007).*

| Material | | Colluvium | CDT | MDT | SDT |
|---|---|---|---|---|---|
| Cohesion, c (kPa)[1,2] | | 4 | | 6.5 | 10 |
| Internal friction angle, ϕ' (deg.) [1,2] | | 35 | | 38.5 | 45 |
| Unit weight (kN/m$^3$) [1,2,3] | | 14 | 17 | 19 | 23 |
| Modulus of deformation (MPa) [4] | | 16 | 20 | 5000 | |
| Poisson's ratio (υ) [4] | | 0.35 | 0.3 | 0.2 | |
| $K_x$ (m/s)[1,5] | | 2.4 x 10$^{-6}$ | 1.5 x 10$^{-5}$ | 1.8 x 10$^{-5}$ | 2 x 10$^{-7}$ |
| Fredlund – Xing Parameters[5,6,7,8] | a (kPa) | 20.3 | 63.14 | 110.48 | Saturated |
| | n | 2.356 | 2.118 | 2.015 | |
| | m | 0.562 | 0.665 | 10.618 | |
| | $\theta_s$ ($\theta_r$) | 0.452 (15%) | 0.444 (15%) | 0.43 (15%) | |

[1](Maunsell Geotechnical Services Ltd., 2007); [2](Wong, 2020); [3](Liu et al., 2022); [4](GCO, 2000); [5](Davis, 1952; Ng & Shi, 1998); [6](Fredlund, 1998); [7](Fredlund et al., 1996); [8](Carsel & Parrish, 1988).

In addition to the datasets described above, the research team collected 33 soil samples from 11 geographically distributed sites. The samples were predominantly moist, brown, and composed of silty-clayey and gravelly sands. Laboratory testing yielded an average effective cohesion of 3.67 kPa (range: 0.4–9.6 kPa) and an average internal friction angle of 30.5° (range: 20.4°–35.8°) (ETS-Testconsult Ltd., 2023) (refer to Figure A3; Table 4.4). These experimental results show the shear stress at failure for each test under the three normal stress conditions.

Statistical analysis of the data reveals: a) At 20 kPa normal stress: Shear stresses range from 11.6 kPa to 19.6 kPa with a mean of 15.4 kPa and standard deviation of 2.8 kPa; b) At 30 kPa normal stress: Shear stresses range from 17.0 kPa to 27.4 kPa with a mean of 21.0 kPa and standard deviation of 3.0 kPa; and c) At 50 kPa normal stress: Shear stresses range from 24.0 kPa to 41.3 kPa with a mean of 32.7 kPa and standard deviation of 4.8 kPa. Furthermore, the sample's state was characterized by measuring initial and soaked moisture contents across all 33 samples. The initial moisture content ranged from 0.063 to 0.219, with the standard deviation of the response (soaked minus initial) being 0.038.

*Table 4.4. Summary of direct shear test statistics.*

| Normal Stress (kPa) | Mean Shear Stress (kPa) | Standard Deviation (kPa) | Range (kPa) |
|---|---|---|---|
| 20 | 15.4 | 2.8 | 11.6-19.6 |
| 30 | 21.0 | 3.0 | 17.0-27.4 |
| 50 | 32.7 | 4.8 | 24.0-41.3 |

## 4.4. AUGMENTED DATASET

### *4.4.1. Subsurface Landslide Response Augmentation.*

To overcome the temporal limitations of sensor-based systems and account for varying spatial extents, this section presents the augmented dataset of subsurface responses, focusing on VWC, designed to support and strengthen sensor-driven prediction models. After numerically validating the case study, data from the Hong Kong Observatory (available online at DATA.GOV.HK) is used in the numerical simulations for the shallow colluvium layer at depths of 0.50 and 1.50 meters. The virtual monitoring point used is located at T1 (refer to Figure 4.5). The simulated rainfall patterns span from 2010 to 2024. This virtually monitored dataset serves three main purposes: a) enriching the dataset by considering a different location; b) improving model training; and c) testing the proposed model.

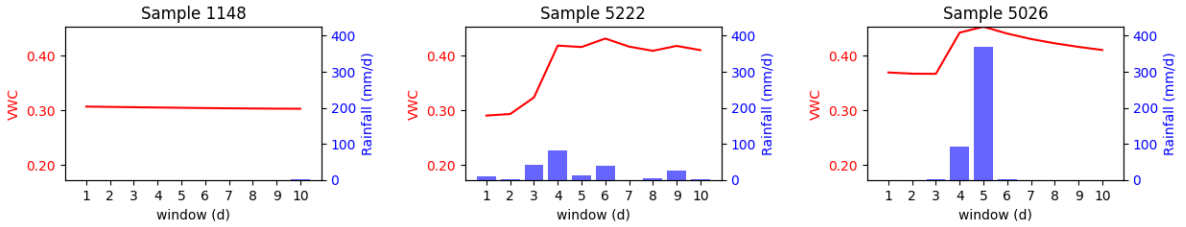
Using a 10-day window, matching the Observatory’s forecast, which covers the current day plus the following nine days, the time series data were segmented into subcases. Each subcase used the initial VWC as a boundary condition and included rainfall features such as

patterns, duration, intensity, and non-rainfall durations as input features. This process resulted in 5439 subcases for the responses at a depth of 0.50 meters and an equal number for the sensors at a depth of 1.50 meters at T1. Figure 4.6 presents a sample of these subsets, showcasing various randomly selected patterns versus the VWC response. These patterns are modeled as time-dependent, an approach that is rarely addressed in the literature. For example, the figure illustrates patterns such as dry, prolonged low-intensity rainfall and short high-intensity events; each pattern yields a distinct VWC response, underscoring the need for accurate modeling.

*4.4.2. Laboratory Tests Augmentation.*

To support automated parameterization and enable an automated stability assessment system, this section presents the augmented dataset used by the APHs tool. Figure 4.7 shows the frequency distribution of the HS parameters, with histograms depicting how frequently each parameter takes specific values across the dataset. Figure 4.8 presents a scatter pair plot, displaying the relationships between pairs of HS parameters, with each point representing a case and its corresponding parameter values. Most parameters show a relatively uniform frequency distribution. During the development of the numerical simulation, each parameter set corresponding to laboratory tests was ensured to have an apparent failure occurring within the strain range. Consequently, parameters such as the $R_f$ and $m$ exhibit higher frequencies for specific values (Figure 4.7) as they were manually adjusted, ensuring that failure is achieved. To illustrate, $m$ was manually adjusted for the cases that correspond to a lower value of the $E_{50}^{ref}$, and the $E_{oed}^{ref}$ by increasing the value of $m$. This can be visualized in the pair plot between these parameters in Figure 4.8. Similarly, the upper limits of the $R_f$ were modified

(reduced for the cases with lower $E_{50}^{ref}$ and $E_{oed}^{ref}$ as visualized in Figure 4.9). This explains why specific values of these parameters occur more frequently than others.




(a) dry (b) prolonged low-intensity (c) short high-intensity

*Figure 4.6. Sample subcases of the numerically simulated VWC responses.*

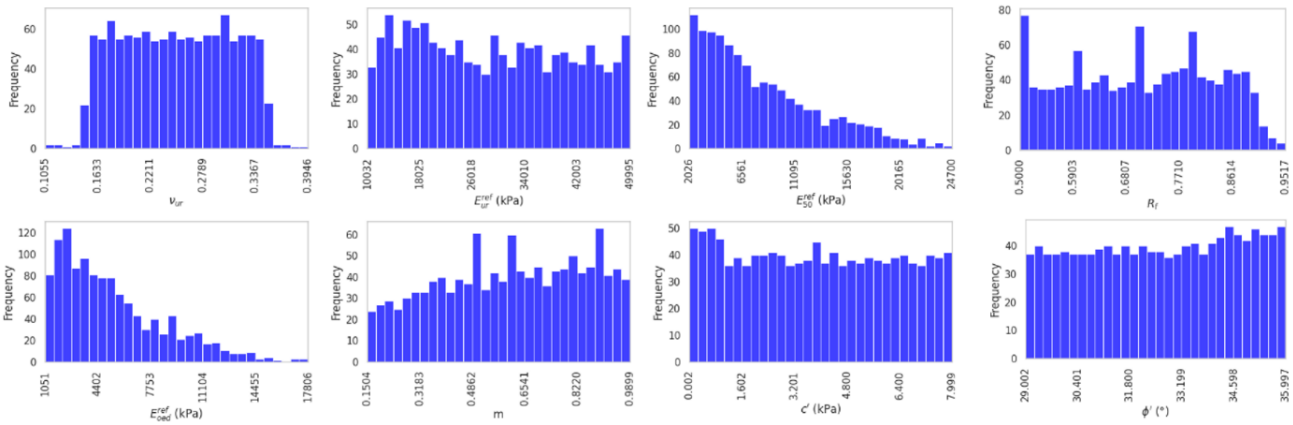



*Figure 4.7. HS parameters frequency distribution (augmented dataset)*

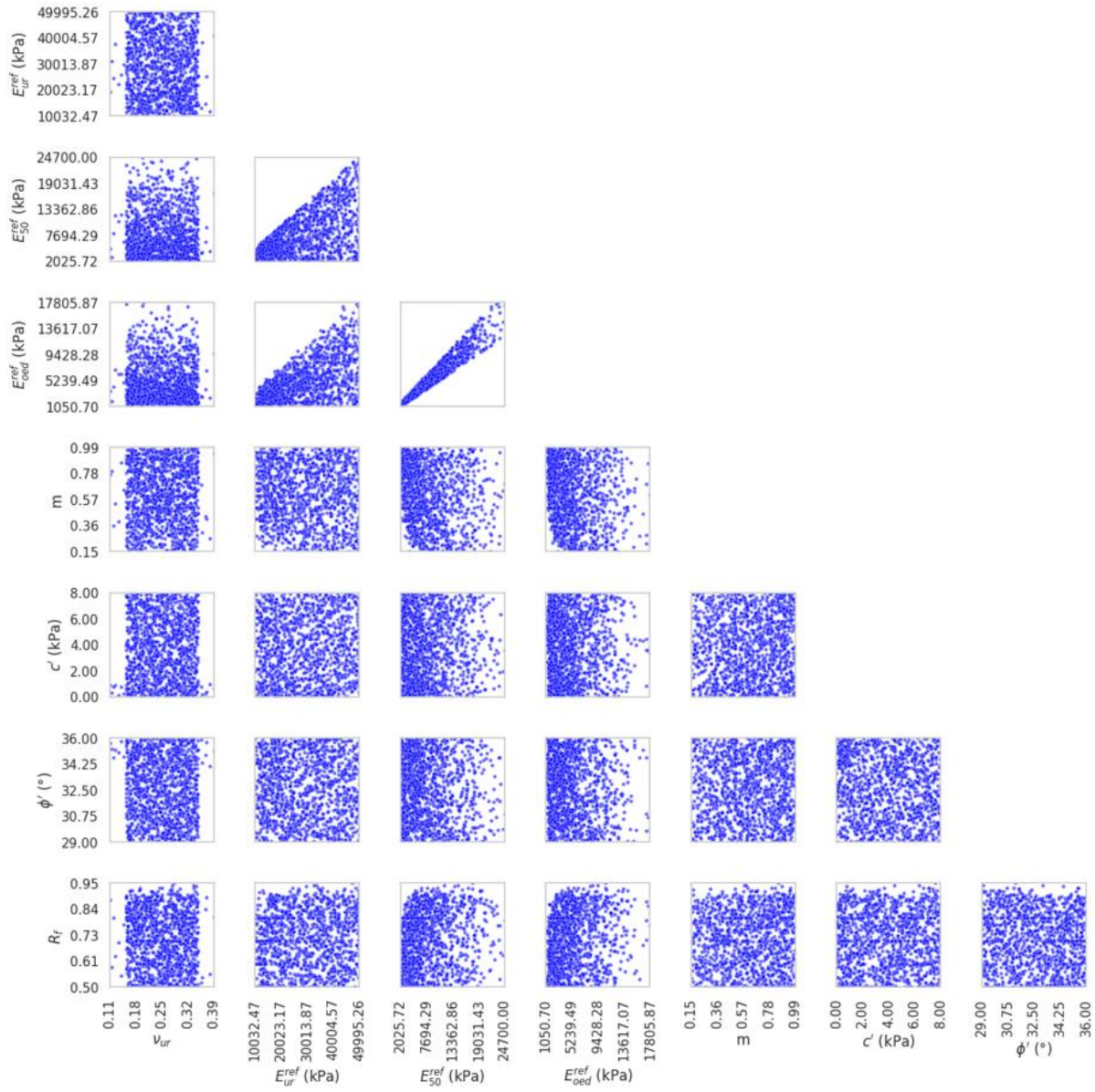



*Figure 4.8. HS parameters pair plot relationships (augmented dataset)*

The distribution of $E_{50}^{ref}$ is skewed, as it is constrained by the maximum value of $E_{ur}^{ref}$, where the maximum value of $E_{50}^{ref} = 0.5E_{ur}^{ref}$. A similar pattern is observed with $E_{oed}^{ref}$, which is assumed to be within the range of 0.5 to 0.8 * $E_{50}^{ref}$. Figure 4.8 confirms these relationships, demonstrating that they align with the theoretical assumption and conform to the expected range. Parameters such as $c'$, $\phi'$, and $\upsilon_{ur}$ are uniformly distributed. $\upsilon_{ur}$ is occasionally simulated outside the predefined bounds to enhance the model's learning capability, as this parameter has minimal influence on the simulation results. Figure 4.9 illustrates the

relationships between $\varepsilon_1$, $q$, and $\varepsilon_v$ as simulated using the SIGMA/W axisymmetric finite element model (GeoStudio, 2023). These simulations represent the augmented parameter space defined in Table A6 and detailed in Figures 4.7 and 4.8. The figure comprises six subplots, each corresponding to a specific $\sigma_3$, and visualizes 1200 curves representing the response of 1200 unique cases, each defined by a labeled parameter set. Additionally, these plots include a heatmap representation of $\upsilon_{ur}$. The relationships are integral to both model training and validation as input features.

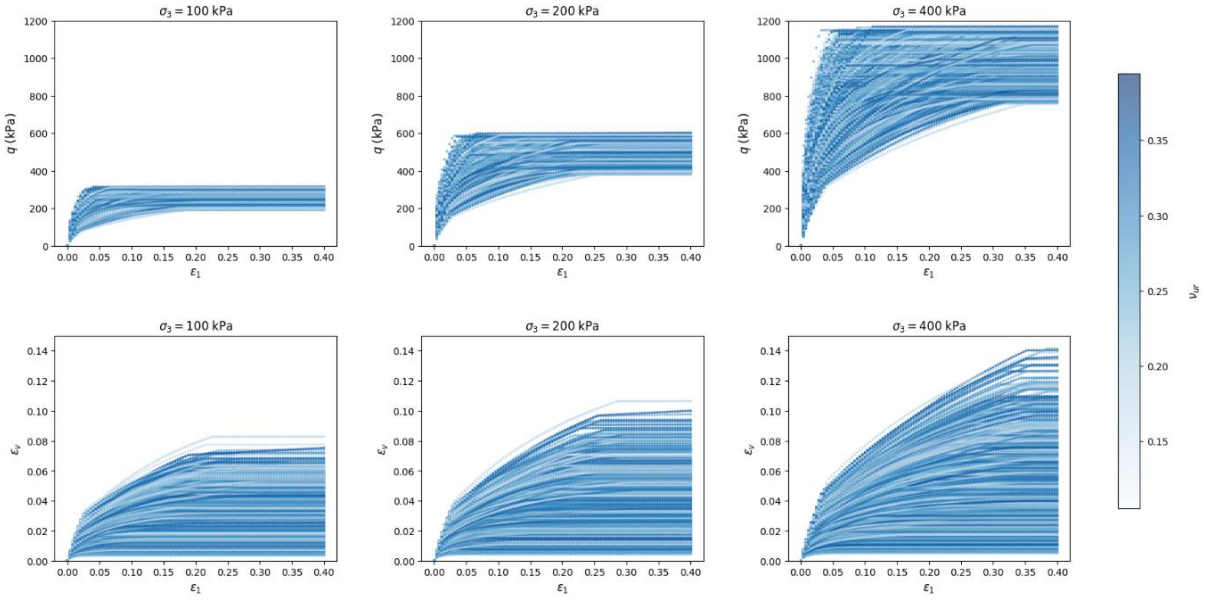



*Figure 4.9. The augmented stress-strain curves*

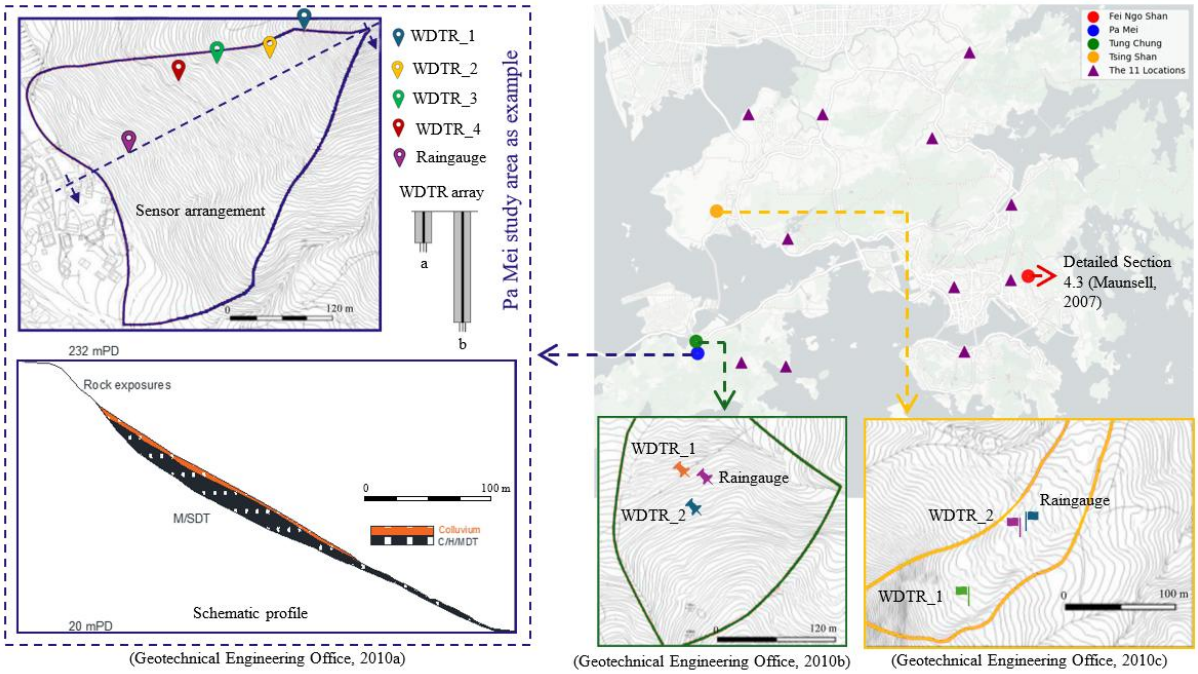
## 4.5. SENSOR-BASED DATASET

The data utilized in this section were collected from various study areas across Hong Kong, including Tsing Shan, Tung Chung, and Pa Mei. The monitoring data are documented in Geotechnical Engineering Office reports (2010a–c), which provide detailed descriptions of the geological setting, instrumentation setup, and the data collected from these monitoring efforts. The collection and use of this data were made possible through the gracious permissions granted by the Head of the Geotechnical Engineering Office and the Director of Civil Engineering and Development (CEDD). This data serves a crucial role in validating and

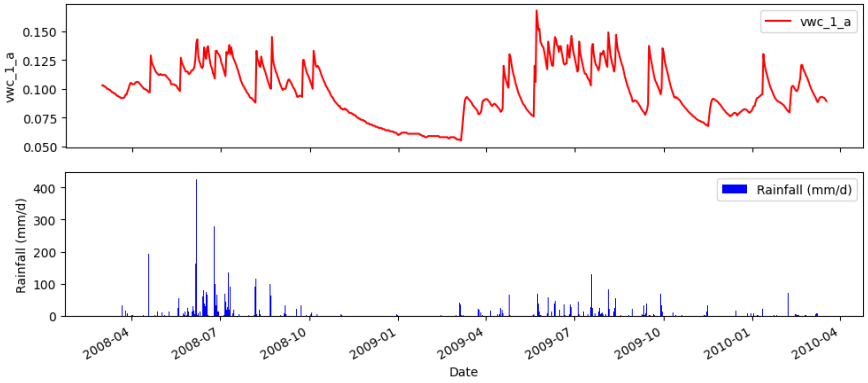
assessing the generalization of the methods illustrated in Section 3, thereby ensuring that the proposed models are effectively tested with real-world geotechnical data.

For the OWHK tool, Figure 4.10 shows a schematic of the study area locations and collected samples. The sensor arrangement is depicted at the three locations through a monitoring array installed at two depths: the upper sensor (labeled "a") is positioned at 0.50 m, and the lower sensor (labeled "b") at 1.50 m. A schematic of the soil profile at Pa Mei is provided as an example. The soil profile is as follows: the top layer is colluvium (displayed in orange), which overlies a layer of moderately to completely decomposed tuff (indicated by a hatched pattern). Beneath this is a zone of highly decomposed tuff, and the profile is capped by a layer of moderately to slightly decomposed tuff at the base. The thickness of each layer varies among the study locations. The sensors used to monitor the VWC were situated within the colluvium layer in the response zone.

In this study, data were collected from three distinct locations to capture VWC at shallow depths. At Pa Mei, four WTDR soil moisture probes equipped with dual-sensor arrays were installed at depths of 0.50 m and 1.50 m, while in Tung Chung, two probes with sensors at the same depths were employed. At Tsing Shan, measurements were recorded using two probes at 0.50 m and one probe at 1.50 m, with outlier and deeper sensor data not integrated into the current analysis. At all locations, a site-specific tipping bucket rain gauge continuously recorded rainfall, following installation procedures detailed in the Geotechnical Engineering Office reports (2010a–c). Figure 4.10.b illustrates representative sensor data, showing the relationship between rainfall and the corresponding VWC over an approximately two-year period starting in March 2008. Additionally, to support further geotechnical analysis, a numerical simulation was conducted of the Fei Ngo Shan landslide (Maunsell Geotechnical Services Ltd., 2007), which is underpinned by extensive laboratory and field test data as described in Section 4.4.1.




(a) Schematic map of approximate study area locations and collected samples (modified from CEDD) (© OpenStreetMap contributors and © CARTO).




(b) Sample data from one sensor utilized: rainfall vs. VWC at Pa Me.

*Figure 4.10. Sensor-based study area map and representative sensor record*

For deeper soils and to improve predictive robustness across heterogeneous sites, Figure 4.11 presents the sensor sharing network. At Pa Mei (Figure 4.10.a and Figure 4.11.b), Soil moisture probes (WTDR) with three-sensor arrays were installed at depths of 0.50 m, 1.50 m, and 2.50 m (for example, vwc_2_a denotes the WTDR_2 sensor at depth a). Each site is equipped with a tipping-bucket rain gauge and an evaporimeter to monitor rainfall and evaporation. Monitoring data indicate that baseline groundwater levels at the study sites are deep, below the rockhead. Deformation monitoring on brittle-landslide slopes proved challenging, whereas VWC measurements more reliably identify areas at elevated failure risk.

Tung Chung (Figure 4.11.a and 4.11.b) was used to evaluate the generalizability of the framework. The Tung Chung soil profile consists of a colluvium layer overlying a sequence of tuffs: a moderately to completely decomposed tuff, a highly decomposed tuff, and a moderately to slightly decomposed tuff. Rupture surfaces were typically encountered between 0.6 m and 1.5 m depth, with isolated occurrences down to 3.0 m. Groundwater monitoring shows that the base groundwater lies well below the identified rupture surfaces, suggesting that transient moisture changes and variations in soil suction are primary controls on the observed movements at the site.

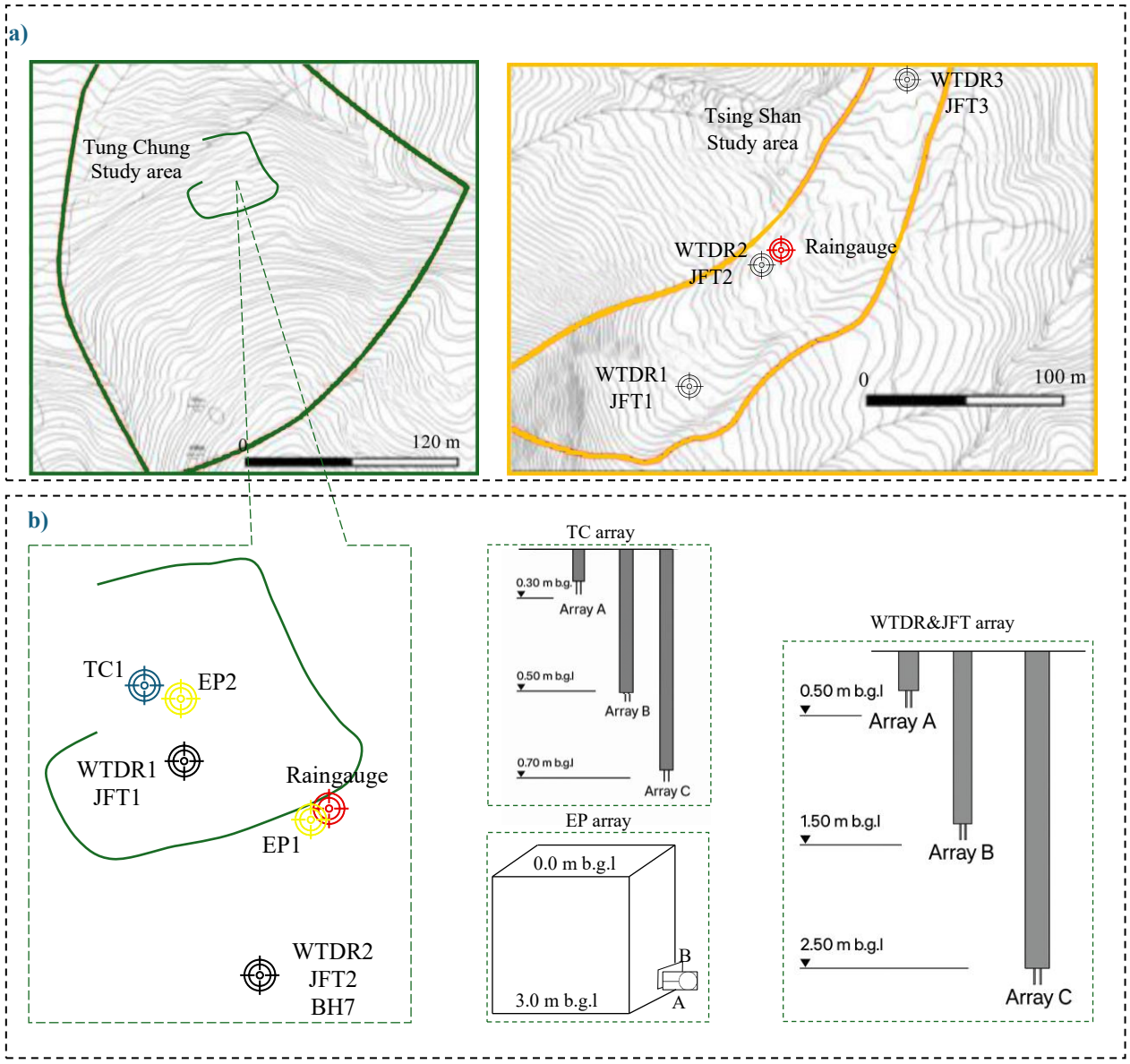



*Figure 4.11. The sharing network with a monitoring array among several spatial locations*

To enhance model-sharing capabilities, multiple hydrological and mechanical subsurface responses were collected from the Tung Chung case study, including earth pressure (EP) and suction (refer to Figure 4.11.a and 4.11.b). Two suction sensor types were deployed: Thermal Conductivity (TC) and Jet Fill Tensiometers (JFT). One TC sensor with three arrays was installed at depths of 0.30 m, 0.50 m, and 0.70 m, and two JFT sensors, each with three arrays, were installed at depths of 0.50 m, 1.50 m, and 2.50 m. Two VWC sensors (WTDR), each with three arrays, were installed at 0.50 m, 1.50 m, and 2.50 m. Two EP sensors were installed to capture the mechanical subsurface response in two dimensions. Groundwater records from the initial ground investigation (2005) for borehole BH7 (response zone 3.00 m bgl to 10.20 m bgl) were used for further analysis.

Data from Tsing Shan were included (refer to Figure 4.11.a and 4.11.b) to assess framework generalization: three JFTs (each with arrays at 0.50 m, 1.50 m, and 2.50 m), three WTDRs (each with arrays at 0.50 m, 1.50 m, and 2.50 m), and two EP sensors were installed. For sensors No. 1 and No. 2, the three arrays (A, B, and C) of the JFT and WTDR sensors correspond to depths of 0.50 m, 1.50 m, and 2.50 m, respectively; array A is located in the colluvium, while arrays B and C are located in the completely decomposed andesite. For sensor No. 3, all arrays are located within the colluvium. This study used data from March 2008 to September 2009 to ensure that both training and validation datasets include rainy and dry seasons for robust and generalizable model evaluation. Suction monitoring during this period is considered comparable between Tung Chung and the Tung Chung case study. For earth pressure, only EP1P and EP2P data from March 2009 to August 2010 were used, since EP1 and EP2 provided insufficient records for model training.

## 4.6. SUMMARY

This chapter describes the data-collection methods used to develop models for the study objectives. To support the second objective, operational landslide prediction modeling, a historical incident dataset was compiled, including rainfall sensor records and related spatial attributes. Case studies from CEDD records were numerically modeled for model validation. For the third objective, sensor-based prediction, field datasets obtained from CEDD were used to train and develop predictive models. To support the final objective, augmented data were generated by combining numerical-analysis outputs with the assembled datasets to expand training coverage and improve model robustness.

# CHAPTER 5: RESULTS AND DISCUSSION

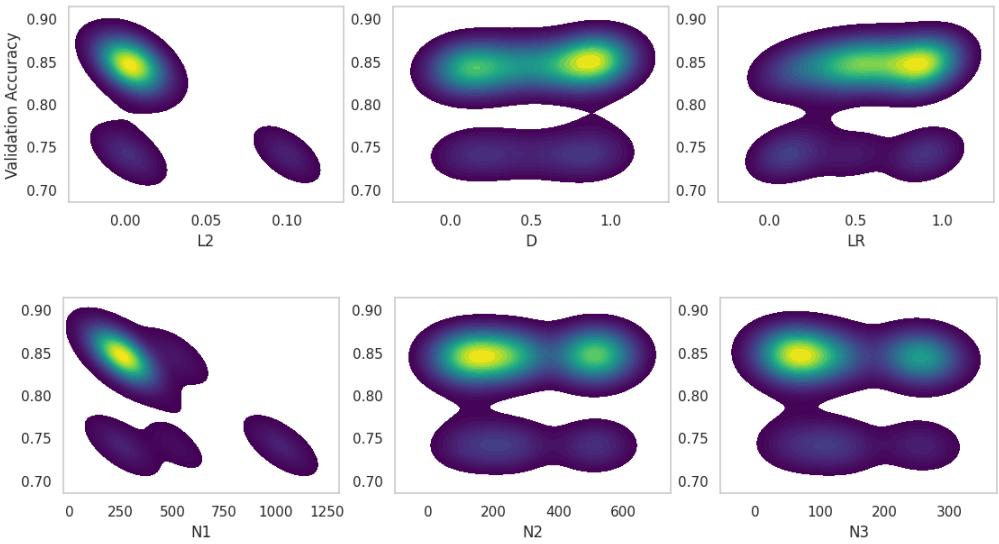
## 5.1. INTRODUCTION

This chapter provides the results of the research methodologies developed to fulfill the objectives of the current study. The results are divided into three parts: the first presents the landslide prediction model developed from historical incidents; the second describes the sensor-based system used to provide more reliable predictions and to refine ambiguous cases produced in the first stage; the third presents focused analyses that improve accuracy and support the development of an automated system. Detailed results are presented below.

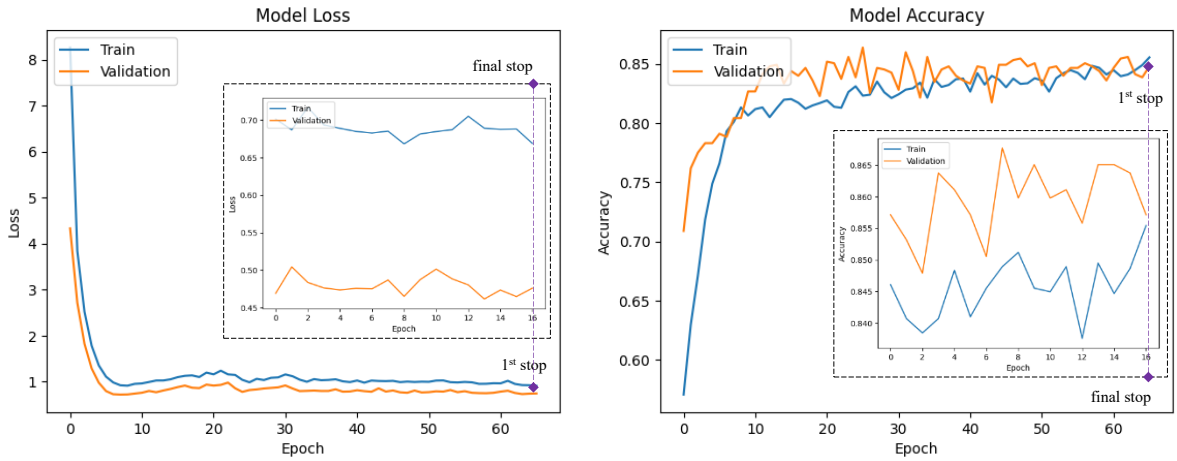
## 5.2. LANDSLIDE PREDICTION MODEL (TIER 1)

### *5.2.1. Model Tuning and Training*

DL requires precise parameter tuning to attain optimal performance. This study applied Bayesian optimization (Snoek et al., 2012) over a maximum of 100 trials, each with 250 epochs, to efficiently explore the parameter space. The best parameter set was refined using a coarse-to-fine-tuning strategy (see Figure A4). Early stopping was employed, and loss curves were visually monitored to iteratively adjust settings, thereby reducing the risk of local minima and balancing overfitting with underfitting. This approach enhanced computational efficiency and accuracy. To ensure model robustness and reliability, a hold-out validation strategy was implemented by randomly partitioning the dataset into 70% training, 15% validation, and 15% testing. Figure 5.1.a shows the coarse tuning approach. Figure 5.1.b shows the learning curves (i.e., the fine-tuning) for LSTM as an example, where a rapid loss decline during the initial epochs is followed by stabilization at a low level. The close correspondence between the training and validation loss curves suggests minimal overfitting despite minor fluctuations, likely due to the Adam optimizer and stochastic dropout.




a) Coarse tuning using Bayesian optimization (Notations are listed in Table 3.1).




b) Fine-tuning: training curves with loss and accuracy monitoring.

*Figure 5.1. Model coarse-to-fine tuning*

Table 5.1 displays the optimized parameters for the LSTM, GRU, and CNN models. Table 5.2 summarizes the optimized parameter settings for the LR models, which were developed to evaluate the individual contributions of engineered features to model performance. The hyperparameter tuning durations reveal distinct computational demands among the models. The LSTM required approximately 9 hours and the GRU about 11 hours, indicating that their architectures' sequential nature and recursive processing impose significant computational overhead. In contrast, the CNN, benefiting from its parallelizable structure for feature extraction, completed tuning in roughly 5 hours. The LR model, characterized by its

more straightforward, convex optimization process, required only 17 minutes. These variations underscore the trade-offs between model complexity and tuning efficiency.

*Table 5.1. The optimized model parameters.*

| Parameter | Value (target) | | | | |
|---|---|---|---|---|---|
| DL model | LSTM | LSTM | LSTM | GRU | CNN |
| Features | Set (1) | Set (2) | Set (3) | Set (3) | Set (3) |
| N1 | 512 | 512 | 512 | 1024 | 512 |
| N2 | 256 | 256 | 512 | 128 | 128 |
| N3 | 128 | 256 | 64 | 128 | 128 |
| D | 0.9 | 0.7 | 0.9 | 0.3 | 0.1 |
| LR | 0.9 | 0.7 | 0.9 | 0.7 | 0.1 |
| L2 | 0.0001 | 0.0001 | 0.01 | 0.001 | 0.001 |
| Kernel size | - | - | - | - | 3 |
| Initial tuning durations* | 09h 32m 10s | 08h 29m 32s | 09h 10m 54s | 10h 32m 52s | 04h 51m 51s |

* These durations are based on the high computational power provided by the HPC server.
Set (1) = Rainfall as time series feature; Set (2) = Rainfall as time series + curvature + slope + elevation; Set (3) = Rainfall as time series + all features listed in Table 4.1

*Table 5.2. The optimized spatiotemporal LR model parameters.*

| Parameter | Search space/value | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Features | CR1W | CR2W | CR3W | CR4W | Initial month | MRI | RD | All |
| Window | 1 week | 2 weeks | 3 weeks | 4 weeks | - | - | - | |
| C | 0.7798 | 0.0608 | 0.0334 | 0.7798 | 0.0124 | 0.2186 | 8.9564 | 0.0631 |
| Solver | liblinear | liblinear | liblinear | liblinear | saga | liblinear | saga | lbfgs |
| Tuning durations | 2 min | 2 min | 2 min | 2 min | 2 min | 2 min | 2 min | 17 min |

Penalty: L2; Scoring: F1; Weights: Balanced; Iterations: 5000
All: Model incorporating CR1W, CR2W, CR3W, CR4W, Initial Month, MRI, and RD

### *5.2.2. Fei Ngo Shan Landslide Validation*

Figure 5.2 illustrates the model boundary conditions. The initial water level was set to the lowest monitored level. Top boundaries (lines 3–4) were assigned a water flux equal to the net rainfall, while bottom boundaries (lines 1–6 and 2–3) maintained a zero-flux condition (q = 0). Lateral boundaries (lines 1–2 and 5–6) were defined by specified total head values, and drainage was enforced along lines 4–5 (water rate = 0 $m^3$/sec). Net rainfall equals total rainfall minus runoff. Premchitt et al. (1992) reported runoff ratios of 0–100% (mean 40%) on Hong Kong slopes; here, the ratio is back-calculated to fit observed subsurface responses (30%). Figure 5.3 compares monitored and simulated groundwater levels at BH1, BH2, BH3, and QP4 alongside rainfall records. Despite minor discrepancies, the agreement confirms that the

boundary conditions accurately represent the equipotential boundaries. Similarly, Figure 5.4 demonstrates strong concordance between measured and simulated suction values, further validating the back-calculated parameters and the assigned boundary conditions. The error at each time was calculated, and the normal distribution for the pore water pressure error (PWP), considering the automatic reading time and all values, is presented in Figure A5. a. The error in PWP was considered while calculating the FOS by adding additional or subtracting PWP cases to calculate the corresponding error in the FOS. The overall average error in PWP was 1.18 kPa, as shown in Figure A5. a. The corresponding average error in the FOS almost equals zero, while the minimum and maximum FOS errors were around 5% except for some minor time steps, as presented in Figure A5. b.

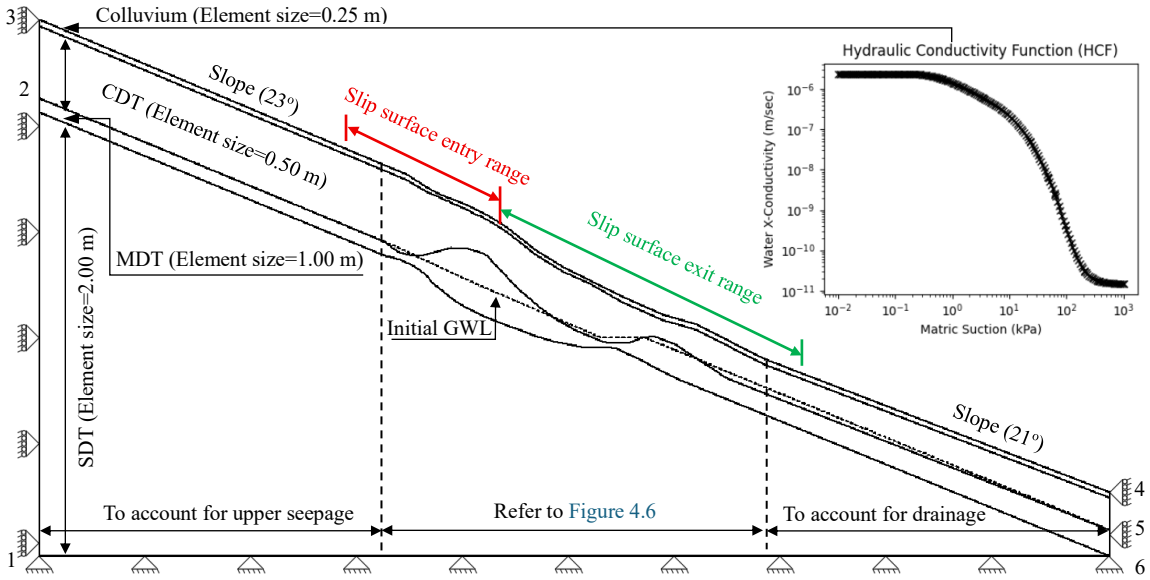



*Figure 5.2. Fei Ngo Shan landslide boundary conditions*

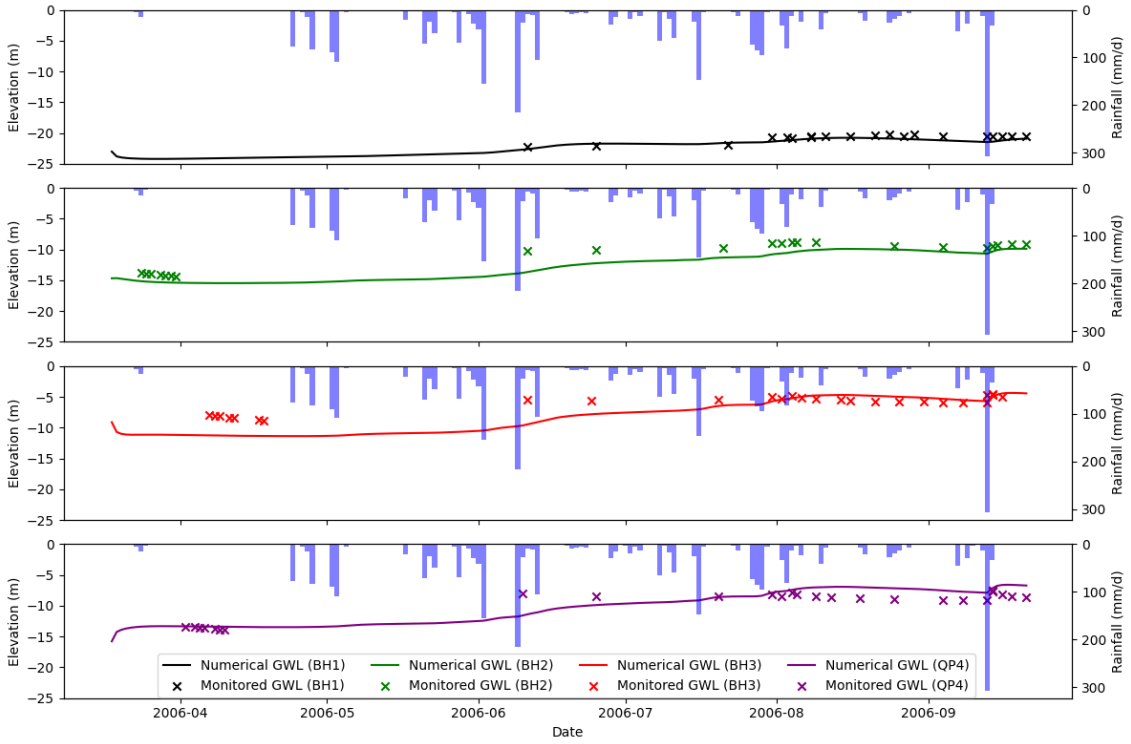



*Figure 5.3. Numerically simulated vs. monitored groundwater levels*

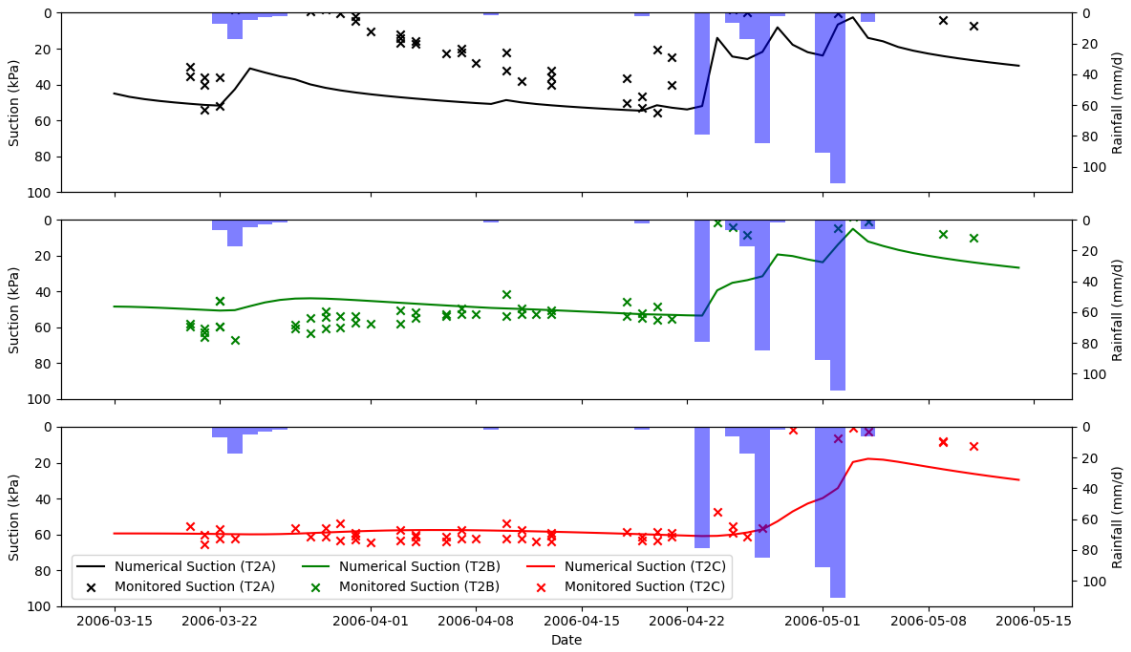



*Figure 5.4. Numerically simulated vs. monitored suction at T2*

### *5.2.3. Model Evaluation with Feature Analysis*

An LR model was used to evaluate several engineered features. Figure 5.5 illustrates the sigmoid curves with optimal thresholds, alongside confusion matrix evaluations for the

training, validation, and testing sets. All features outperformed random chance (50%), with one-week cumulative rainfall showing the strongest correlation with landslide occurrences. Figure 5.5 reveals critical rainfall thresholds for landslide triggering: 177 mm (7-day), 270 mm (14-day), 341 mm (21-day), 424 mm (28-day) cumulative rainfall, 146 mm/day/month maximum intensity, and 7 rainy days/month.

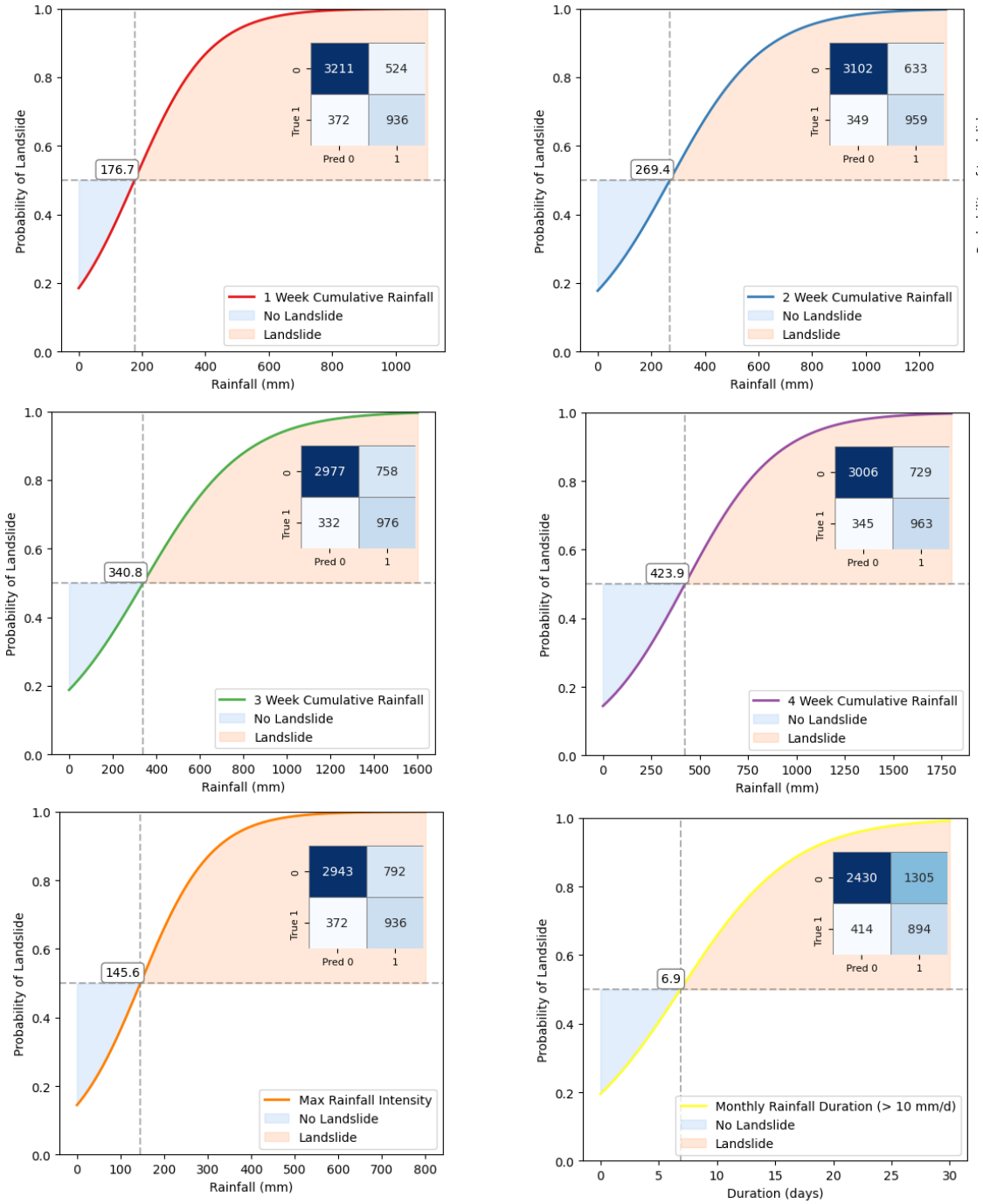



*Figure 5.5. Multi-rainfall thresholds for several features: CR1W; CR2W; CR3W; CR4W; MRI; RD (definitions in Table 4.1). Figure A6 details the confusion matrix analysis.*

Figure 5.6.a compares the LR and three LSTM models, each utilizing a distinct set of features, displaying their overall confusion matrices and performance metrics. Although the LR model outperformed a single-feature baseline, the LSTM models significantly surpassed the LR model. Moreover, incorporating spatiotemporal features further enhanced performance across all metrics. Notably, precision increased from 0.64 to 0.80 when the problem was modeled as a time series with spatiotemporal features. Figures 5.6. (b-c) present confusion matrices and ROC curves for four models: LSTM, GRU, CNN, and LR. The LSTM model demonstrates the highest performance with an AUC of 0.92, indicating superior discrimination capacity between classes. Both the GRU and CNN models follow closely, each registering an AUC of 0.90, while the LR model records a comparatively lower AUC of 0.87. The performance hierarchy reflects each model's intrinsic strengths. The LSTM excels in capturing temporal dependencies with its memory and gating mechanisms, while the GRU effectively models sequential patterns despite its simpler structure. The CNN, though primarily designed for spatial feature extraction, adapts well to temporal data.

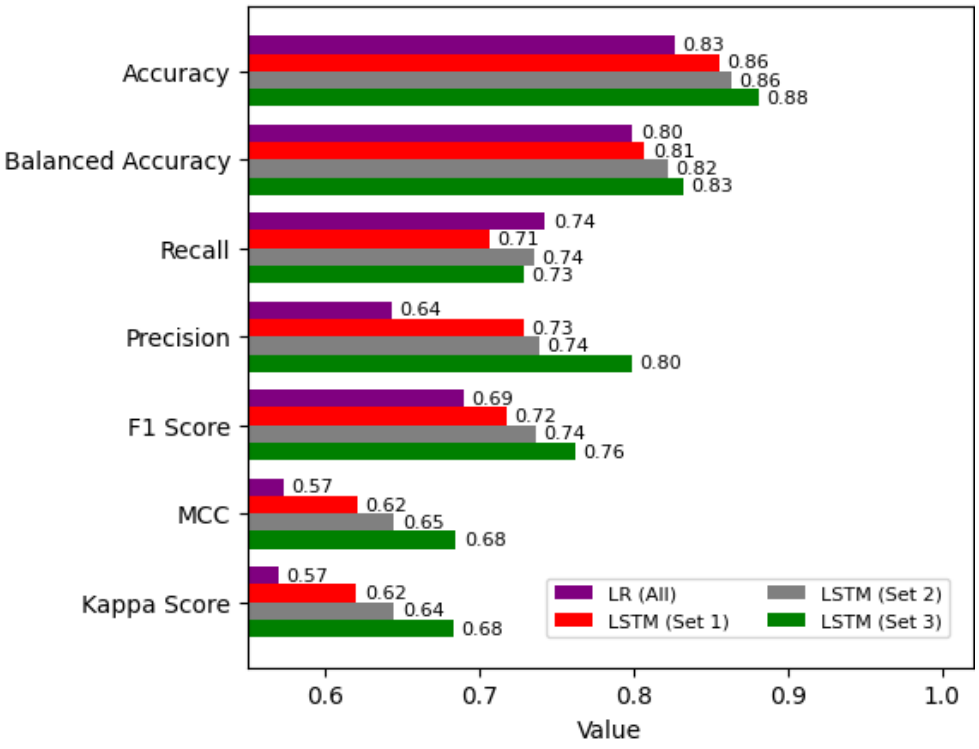
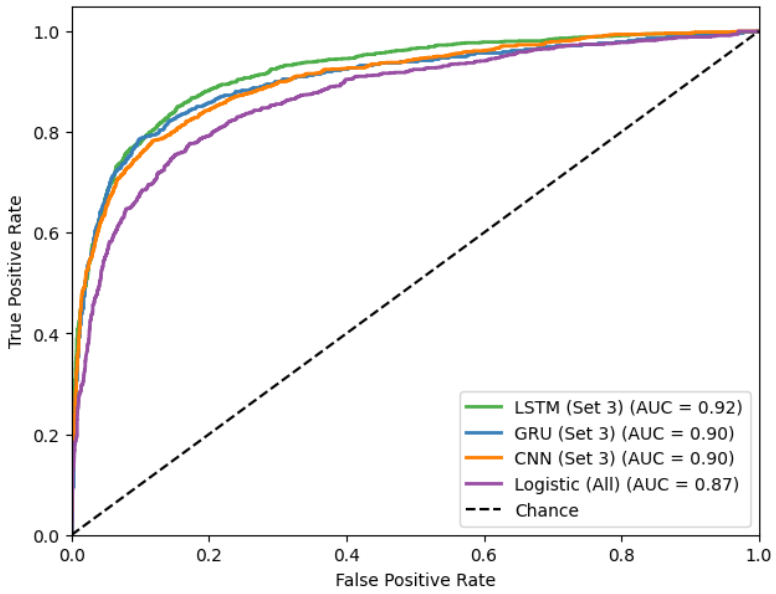
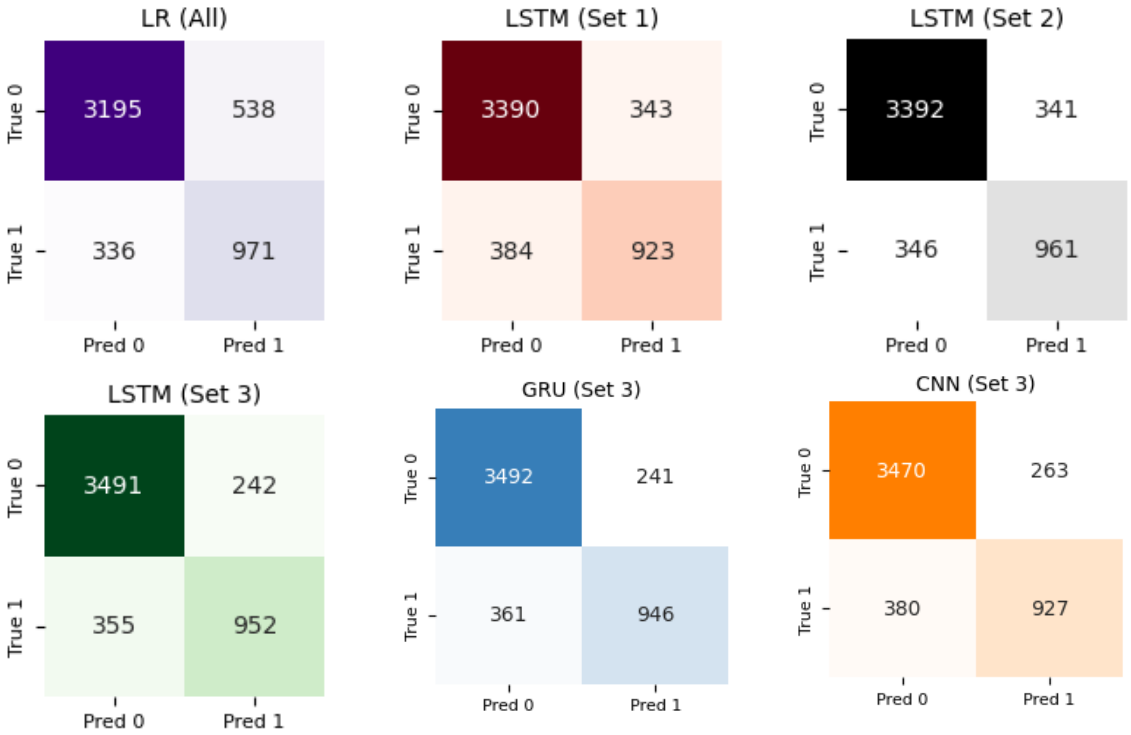



*Figure 5.6. Comparison of a non-temporal vs. temporal models: a) Evaluation metrics*




b) ROC curves




| LSTM (Set 3) | | | | | |
|---|---|---|---|---|---|
| Training (70%) | | Valid. (15%) | | Testing (15%) | |
| 2482 | 137 | 509 | 55 | 500 | 50 |
| 227 | 682 | 59 | 133 | 69 | 137 |

c) Confusion matrices

All: Model incorporating CR1W, CR2W, CR3W, CR4W, Initial Month, MRI, and RD.
Set (1) = Rainfall as time series feature;
Set (2) = Rainfall as time series feature + curvature + slope + elevation;
Set (3) = Rainfall as a time series feature + all features listed in Table 4.1.

*Figure 5.6. Cont. b) ROC curves; and c) Confusion matrices*

Figure 5.7 presents the most influential features identified through permutation importance analysis, quantifying their relative contributions to model accuracy. Rainfall-related features demonstrated the most substantial impact, collectively accounting for approximately 36% of the model's predictive accuracy. This significance substantially exceeds random chance performance (50% baseline for binary classification). The temporal distribution of rainfall influence revealed distinct patterns: antecedent rainfall measures (T1-T26) contributed over 10%, with the two-day antecedent rainfall feature alone accounting for approximately 15%. Real-time rainfall patterns explained an additional 10% of model accuracy. Although spatiotemporal features contributed a smaller proportion to overall accuracy (~2%), their incorporation yielded a 7% increase in model precision compared to rainfall-only models. This enhancement reflects critical geotechnical mechanisms: transitional months (March-May) exhibit elevated temporal susceptibility due to suction transition from dry to wet conditions, while converging terrains demonstrate heightened spatial vulnerability, consistent with established geomorphological principles (Giuseppe et al., 2016). The analysis identified spatial heterogeneity in the model's predictions. Specifically, zones N53, N35, and N02 exhibited relatively higher values compared to other zones. This suggests these zones may encompass unique geological formations or a concentration of historical landslide events. The analysis further identified that Paths, Roads, Residential Areas, and Hillside exhibited relatively higher importance than other nearby structures. Paths and Roads often necessitate slope cutting, destabilizing natural terrain. Residential Areas contribute through structural surcharge or altered drainage. The importance of the Hillside category confirms the model's ability to capture the inherent susceptibility of undeveloped slopes to rainfall-induced failures (Ho & Roberts, 2016). Overall, while rainfall is a direct trigger, incorporating spatiotemporal features improves the model's discrimination of damage severity.

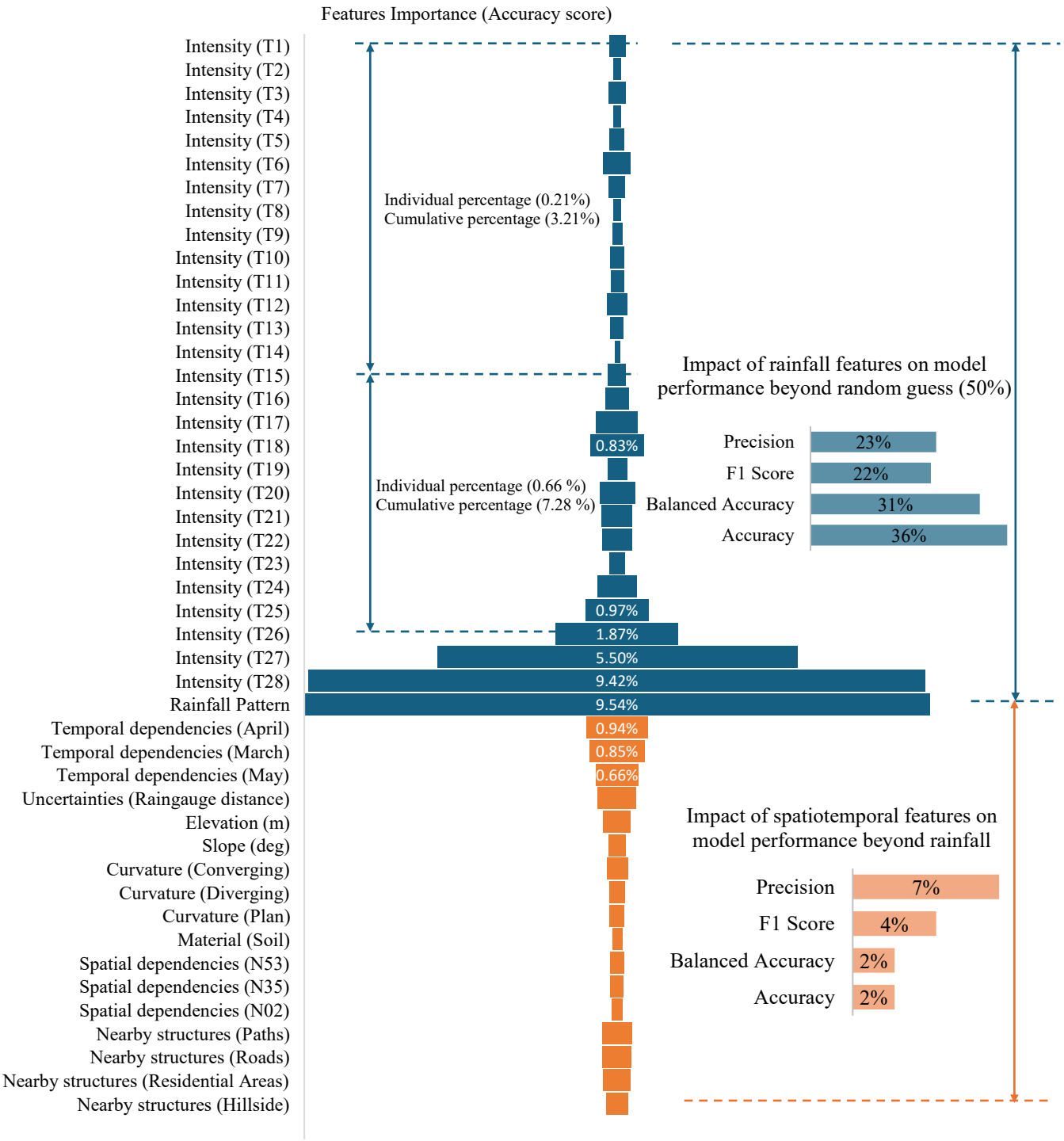



*Figure 5.7. Spatiotemporal LSTM model permutation-based feature importance analysis*

It is important to note that the feature importance visualized in Figure 5.7 is derived from permutation importance analysis, which quantifies the performance degradation when feature values are randomly shuffled relative to the baseline model. This approach is inherently influenced by several factors: (a) the model architecture employed; (b) dataset characteristics,

including volume, quality, and distribution; and (c) the methodological approach to feature perturbation. To substantiate this, Table 5.3 presents comparative importance scores for selected features using alternative methods: specifically, replacing feature values with their minimum and maximum values, respectively. The analysis reveals that importance scores are sensitive to the permutation methodology, as evidenced by the substantial variation observed for feature T23 (0.001 vs. 0.559). Furthermore, no consistent ranking pattern emerges across permutation methods; for instance, Slope exhibits higher importance than Hillside under maximum value permutation, while the inverse relationship holds under minimum value perturbation. This inconsistency underscores the complex interdependencies among features. The preceding feature importance analysis provides a critical, albeit limited, window into the "black box" of DL models. The observed importance scores are influenced by inherent feature interdependencies and the model's architecture, meaning they illustrate how the model behaves rather than providing a definitive, causal ranking of geotechnical factors.

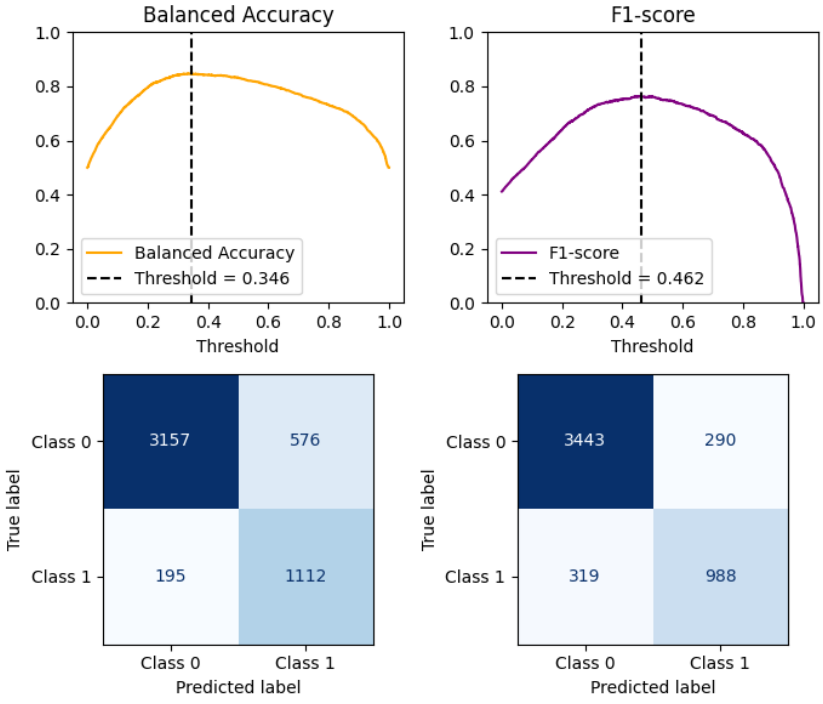
*Table 5.3. Alternative feature importance assessment: Feature's minimum vs maximum value.*

| Selected features | Accuracy score | |
|---|---|---|
| | Min value importance | Max value importance |
| T23* | 0.001190 | 0.55892 |
| T24* | 0.004365 | 0.60853 |
| T25* | 0.005556 | 0.61924 |
| T26* | 0.004563 | 0.62222 |
| T27* | 0.014683 | 0.62222 |
| T28* | 0.041270 | 0.62142 |
| Uncertainties (Rainguage distance) | 0.0026 | 0.0415 |
| Slope (deg) | 0.0024 | 0.0403 |
| Spatial dependencies (N53) | 0.0018 | 0.0145 |
| Nearby structure (Hillside) | 0.0026 | 0.0137 |
| Temporal dependencies (April) | 0.0083 | 0.0131 |
| Curvature (Converging) | 0.0018 | 0.0040 |
| Material (Soil) | 0.0046 | 0.0006 |
| *Time step Tn represents the rainfall intensity at the nth temporal interval in the sequence in days | | |

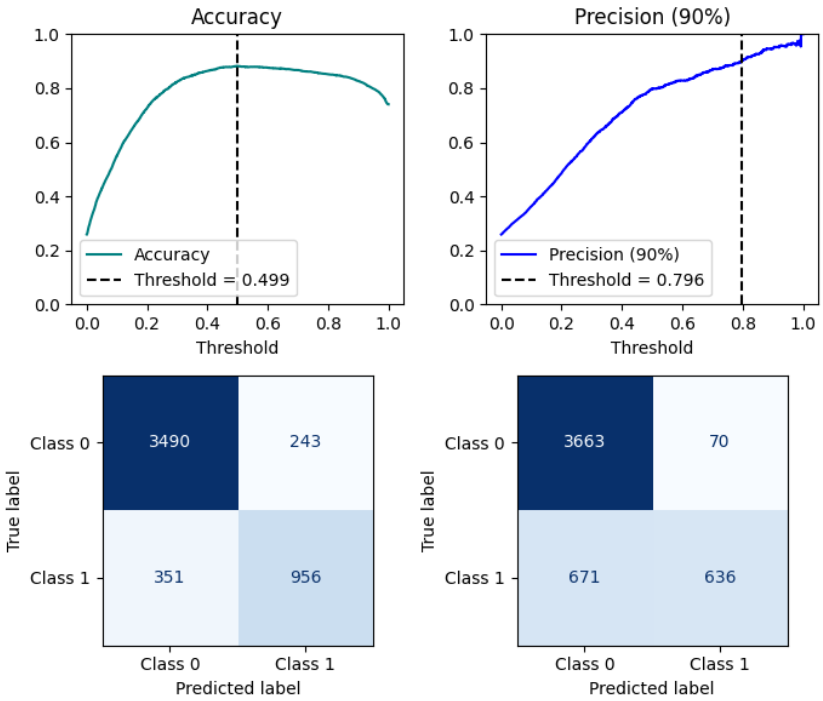
### *5.2.4. Model Warning Thresholds*

Relying on a fixed 0.5 probability threshold for classification can yield misleading results. To address this, this study introduces a dual-threshold operational framework that moves beyond

a single classification cutoff. Multiple performance metrics, including Balanced Accuracy, Accuracy, F1-score, and Precision, were evaluated across a range of probability thresholds (Figure 5.8). Among these metrics, the F1-score demonstrated better performance for establishing a lower classification boundary, effectively balancing false positives and false negatives to exclude non-landslide cases with greater reliability than either Balanced Accuracy or Accuracy. For the upper threshold, a 90% precision level was selected to isolate instances with high landslide probability. Unlike the F1, which exhibits a clear maximum, precision increases with threshold level. Therefore, the 90% precision threshold represents a conservative, operationally-driven choice rather than a mathematical optimum. Cases falling between these two thresholds are categorized within a "warning zone" that indicates the need for further investigation through geotechnical analysis or slope monitoring, enabling proactive risk management.




a) Balanced accuracy and F1




b) Accuracy and Precision

*Figure 5.8. Threshold curves and corresponding confusion matrices (LSTM as an example)*

### *5.2.5. Model Validation and Testing*

The developed models were rigorously evaluated on an independent, real-world case study to assess their operational robustness. Their performance was benchmarked against a deterministic model that computes the slope FOS by explicitly modeling the infiltration and seepage processes. The modeling process of the physical model began with seepage analysis using SEEP/W. The soil water characteristic curve and the hydraulic conductivity functions are back-calculated to match the field-monitored data and laboratory tests. Then, the pore water pressure is exported to the SLOPE/W using Bishop analysis with the entry-exit method boundary condition to calculate the deterministic FOS.

Figure 5.9 presents the mean predictions obtained via Monte Carlo dropout over the period from 17 May 2005 to 21 August 2005. The time series encompasses 97 events, each defined by 28 sequential time steps with a one-step overlap between consecutive events. Among these, 96 events correspond to non-landslide conditions, corroborated by field

investigations that reported no significant displacement or tension cracks before failure (Maunsell Geotechnical Services Ltd., 2007). By contrast, the event on 21 August 2005 was associated with a landslide involving an estimated volume of approximately 3500 $m^3$. Overall, a close alignment was observed between the predictions from the data-driven models and the numerical model with the observed reality of landslide behavior. Several key insights are detailed below:

A) The event between 20 and 24 May 2005 prompted a warning from the LR model, which struggles to capture the precise rainfall pattern. In contrast, the DL models accurately accounted for these patterns. This comparison demonstrates that incorporating the actual rainfall pattern alongside spatiotemporal features reduces false warnings.

B) The event from 21 June 2005 to 24 June 2005 triggered a warning across all models. This alert was due to an intense rainfall event that delivered 426.5 mm over four days, following an antecedent event of 161.5 mm from 14 to 16 June 2005. However, no outright failure was reported during this period. To further investigate the origin of this warning, a numerical analysis model was developed, which also produced a FOS below the 1.2 threshold, as recommended by GCO (1984), a value designed to account for uncertainties in monitoring data, flow rules, mesh discretization, boundary condition approximations, solution techniques, and soil property variability. These results suggest that the adopted warning thresholds effectively reduce false alarms while prompting decision-makers to pursue further investigations through more precise methods, such as numerical modeling or subsurface monitoring, to account for unmonitored features, including runoff, vegetation effects, and other factors (detailed in Section 5.3).

C) The numerical model, grounded in physical principles, required additional time for the drainage process to heal the FOS above the critical threshold of 1.2. The data-driven models, particularly the CNN model, partially capture this temporal delay. After the

event on 21 June, the CNN model maintained elevated probabilities, even as subsequent rainfall intensities declined, thereby partially mirroring the delayed recovery.

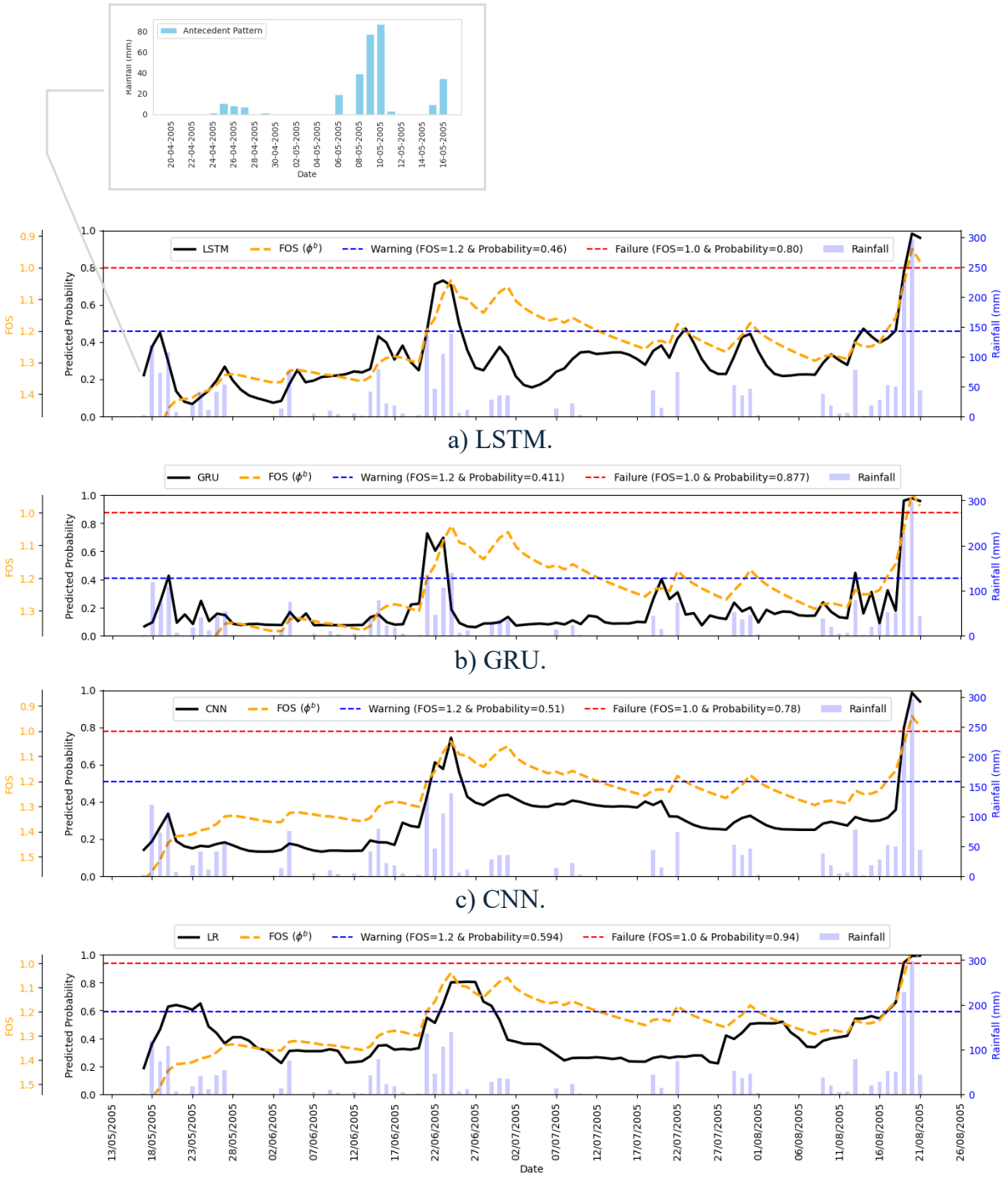



*Figure 5.9. Model Monte Carlo mean predictions vs. the numerical stability*

In Figure 5.9, the warning thresholds for the data-driven models are derived from the analysis in Section 5.2.4. For comparative analysis, the warning threshold for numerical

modeling is set at FOS = 1.2, following GEO (1984) recommendations to account for various uncertainties, while the failure threshold remains at FOS = 1.0. Figure 5.10 presents the predicted probability distributions obtained via Monte Carlo at the landslide event alongside the mean standard deviations across all events. This figure reveals that, although the LSTM model achieves higher accuracy overall, its predictions exhibit moderate uncertainty. In contrast, the CNN model consistently shows lower uncertainty, suggesting that it provides more stable and reliable confidence estimates.

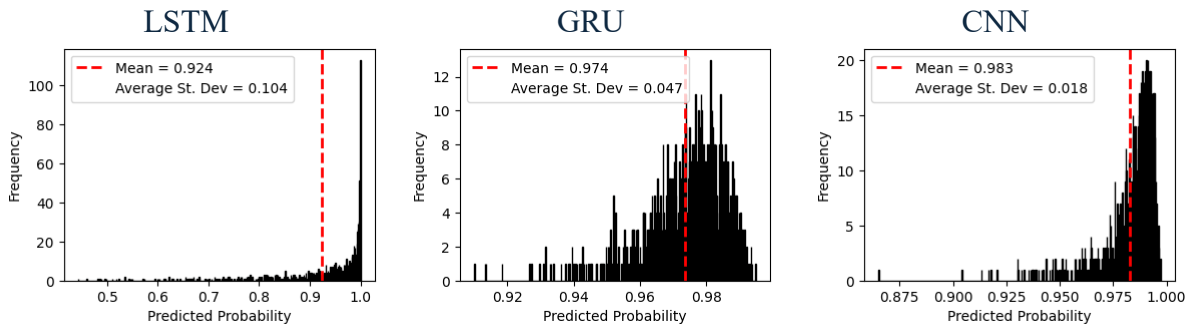



*Figure 5.10. Monte Carlo predictions for the landslide event alongside the mean standard deviations across all events*

### *5.2.6. Optimal Model Selection, Benchmarking, and Uncertainty*

Table 5.4 summarizes a comparative analysis of the models based on: (1) Performance, (2) computational efficiency, (3) uncertainty, and (4) real-world operational reliability to recommend the optimal model. Although the evaluation may involve some subjectivity, it indicates that the CNN strikes an optimal balance among high precision, moderate computations, low uncertainty, and superior operational reliability (partially considering the drainage process with the narrowest warning zone). Nonetheless, all models are integrated into the decision-support interface, allowing users to choose the appropriate one for each scenario.

The model's performance was benchmarked against relevant studies, acknowledging the dataset differences. Compared to Mondini et al. (2023), who used short-term cumulative rainfall events, the time-dependent approach slightly outperformed theirs with an AUC of 0.92

and a Balanced Accuracy of 0.83. However, considering the accuracy as a comparison is misleading for an imbalanced dataset. In this context, Xiao et al. (2023) emphasized that addressing data imbalance is critical. The data preparation methodology explicitly accommodates both temporal and spatial imbalances. As a result, precision improved, surpassing that reported by Xiao et al. (2023), reaching nearly 80%. These findings underscore that detailed spatiotemporal modeling and careful handling of data imbalance can substantially enhance predictive performance.

*Table 5.4. Comparative assessment of the proposed data-driven models.*

| Model | Performance (Precision) | Computational power (Tuning) | Uncertainties (St. deviation) | Reliability | |
|---|---|---|---|---|---|
| | | | | Drainage response (refer to Figure 5.9) | Warning zone width |
| LSTM | High (80%) | High (9 hr) | High (10.4%) | Partial response | Moderate (0.46-0.8) |
| GRU | High (80%) | High (11 hr) | Moderate (4.7%) | Immediate response | Wide (0.41-0.88) |
| CNN | High (78%) | Moderate (5 hr) | Low (1.8%) | Partial response | Narrow (0.51- 0.78) |
| LR | Good (64%) | Low (0.5 hr) | None | Immediate response | Moderate (0.59-0.94) |

Despite the model's outstanding performance, minor false predictions persist due to inherent uncertainties, both aleatory and epistemic, as discussed by Mondini et al. (2023). In this study, uncertainties arise from model-related and data-related challenges, such as imprecise event timing and spatial variability in rainfall, slope mechanics, hydrology, and topography. These issues were mitigated through a comprehensive framework, which includes: a) Monte Carlo dropout was employed to assess uncertainty across different model architectures (LSTM, GRU, CNN, and LR), demonstrating that architectural choices yield distinct outcomes; b) a coarse-to-fine tuning strategy, combined with sensitivity analyses on various feature sets, was adopted to refine model performance; c) given that the dataset (1984–2022) includes events with uncertain failure timings, longer antecedent durations were used to capture consistent pre-landslide patterns; and d) by including the distance between the nearest rain gauge and the landslide location, along with additional spatial features, the model's ability to account for uncertainties affecting the infiltration process was improved.

## 5.3. SUBSURFACE SENSOR-BASED PREDICTION MODEL (TIER 2).

### *5.3.1 VWC Prediction Operational Model (OWHK)*

Following the unique temporal process illustrated in Section 3.3.1, a total of 1,000 randomly selected numerical subcases were merged with field data to form the final dataset. Figure 5.11 presents the density distribution of the VWC measured at both 0.50 m and 1.50 m depths, with the VWC values ranging between 0.055 and 0.493. Figure 5.11.b shows the distribution of cumulative rainfall per pattern, where rainfall intensities recorded at each time step vary from 0 to 426 mm/day. Overall, 13,161 rainfall patterns were analyzed, with cumulative rainfall values ranging from 0 to 898 mm; among these, 2,493 cases correspond to dry patterns, while 10,668 cases correspond to wet patterns. The dataset comprises 6,871 cases for subsurface responses at 0.50 m and 6,306 cases at 1.50 m. Missing data from sensors (e.g., Tsing Shan 1A and Tsing Shan 1B) were removed, and outliers were eliminated during the process to reduce potential bias in subsequent model development.

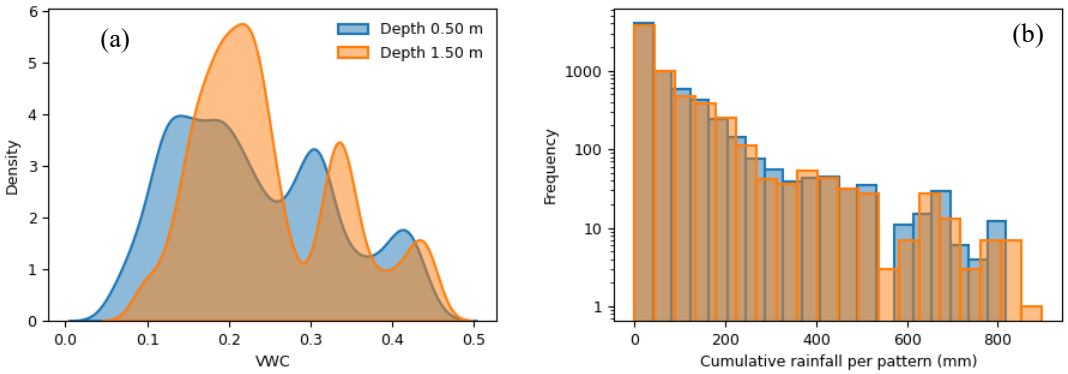


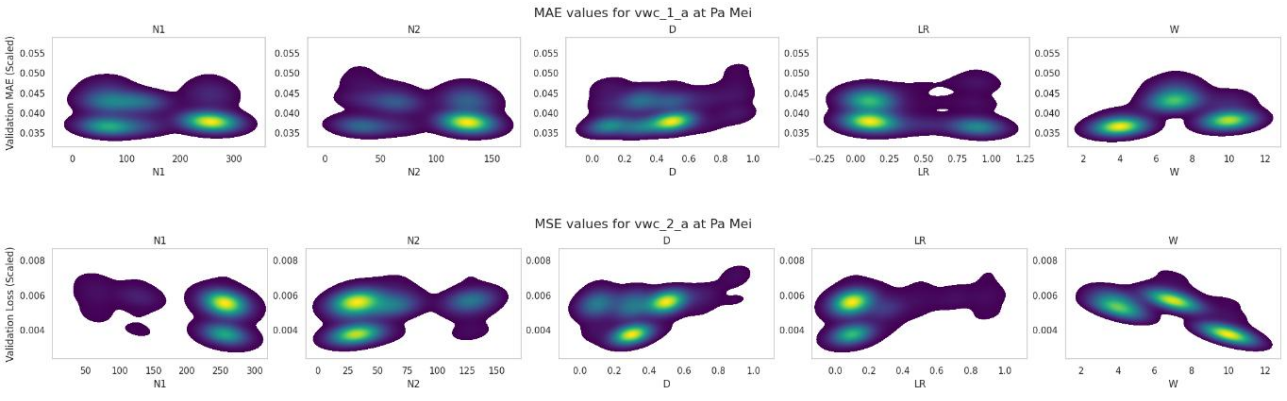

*Figure 5.11. (a) Kernel density estimate of VWC measured at 0.50 and 1.50 m; (b) Frequency distribution of rainfall patterns for the corresponding subcases*

Analysis of the figures reveals a substantial overlap between VWC responses at 0.50 m and 1.50 m for the same rainfall patterns. Moreover, even at a given depth, spatial variability may lead to different VWC responses under similar rainfall triggers. This variability necessitates clustering the data to distinguish among the multiple response patterns associated with similar inputs. In this study, clustering the data based on both target depth and initial VWC

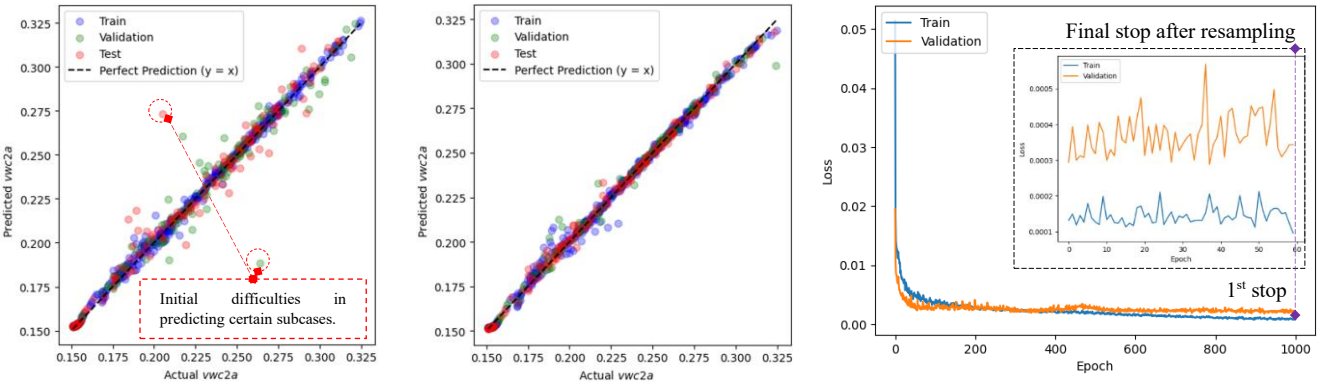
is proposed. The grouping strategy is based on several considerations: a) each group must include both dry and wet rainfall patterns, covering a full range of intensities at each time step; b) the initial VWC boundaries defining the clusters are non-overlapping; and; c) the grouping should balance model performance with data volume per cluster, thereby minimizing the risk of underfitting or overfitting. This balance is achieved through iterative training and fine-tuning. For each cluster, a unique LSTM model was trained, forming the basis of the multi-LSTM approach, which is illustrated earlier in Figure 3.8.

a) Model training

In this study, the Bayesian optimization technique (Snoek et al., 2012) was employed to optimize the parameters for each DL model. A maximum of 300 trials were conducted, with 250 training epochs per trial, to initially explore the parameter space with a validation ratio of 30%. Figure 5.12 presents the density map of each variable against the validation MSE (loss) and MAE for two different sensors as an initial coarse-tuning approach. Some model parameters exhibit a consistent pattern with various sensors (i.e., D and W), while others do not show a clear pattern (i.e., N1, N2, and LR). It is observed that multiple combinations of hyperparameters can result in similar low MSE or MAE values. Coarse tuning discards the model training behavior and alone is insufficient for the final model parameters; hence, a coarse-to-fine-tuning approach is adopted (refer to Figure A4). For each site-specific model, the subsets were randomized and split into 60% training, 20% validation, and 20% testing subcases. While the operational model was randomized and split into 80:10:10%. Early stopping was used during training to prevent overfitting, and the optimized weights were saved. This process was repeated with different randomized splits, using the previously saved weights as initialization for further optimization. Given the uniqueness of each subcase in the dataset, this strategy allowed the model to generalize effectively by exposing it to various combinations of training, validation, and testing subcases.




a) Model initial coarse tuning (density map).




b) Before resampling c) After resampling d) Monitoring of learning curves

b-d) Model fine-tuning and training process

*Figure 5.12. VWC operation model coarse-to-fine model tuning*

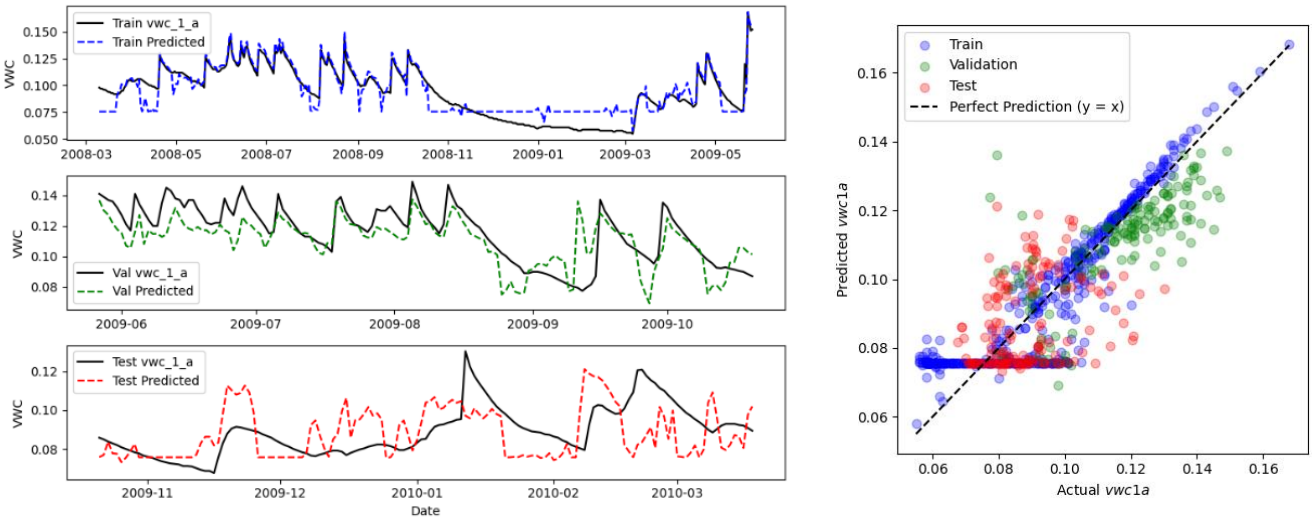
For example, Figure 5.12.b illustrates the training process for a site-specific model at Pa Mei (vwc_2_a), where initial difficulties in predicting certain subcases were resolved through resampling in Figure 5.12.c. Figure 5.12.d shows the learning curves for the optimized model using benchmark parameters (N1 = 128, N2 = 64, D = 0.3, LR = 0.5, W = 10) that are applied across models for direct comparison (see Figure 3.6 for annotations). The curves demonstrate the model's robust performance, with loss values decreasing significantly during the initial epochs and stabilizing at a consistently low level. The close alignment between the training and validation loss curves suggests a minimal risk of overfitting. However, some minor noise and oscillations are present during the training process, likely due to the Adam optimization algorithm and random dropout techniques. The model was trained with early stopping, saving the weights at the first stop (Figure 5.12.b), then resampled and retrained until

reaching the saturation level of training (Figure 5.12.c), marked as the "final stop" (Figure 5.12.d).

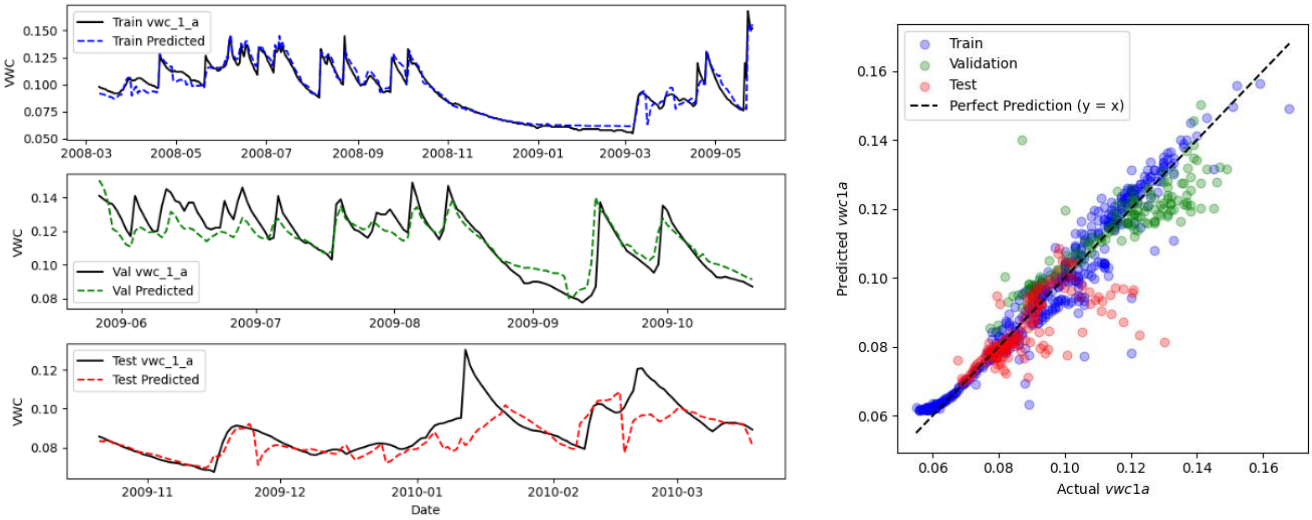
### b) Model results and discussions

This study proposes a unique data preparation technique designed to achieve high forecasting accuracy while maintaining model simplicity. This objective arises from the widely available rainfall data in Hong Kong, despite the absence of controlling features for subsurface moisture dynamics. Given this context, is it possible to achieve higher accuracies using only rainfall features? In response to this question, Figure 5.13 shows the prediction of a site-specific case (vwc_1_a) in the Pa Mei study area. This Figure compares three scenarios: a) the traditional method; b) the traditional method with improved features; c) the proposed improved method.

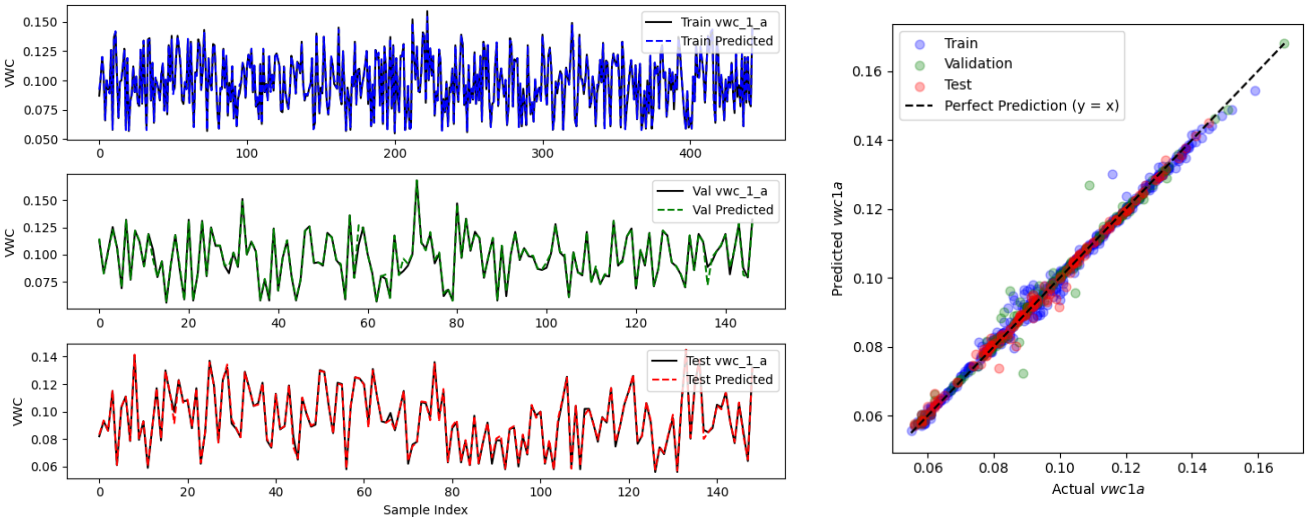
For each scenario presented in Figure 5.13, the left-hand panel features time series subplots depicting the training (blue), validation (green), and testing (red) datasets to illustrate the temporal model performance. The right-hand panel provides corresponding scatter plots for these predictions, enabling a comprehensive evaluation of the model's accuracy as follows: In the traditional method (Figure 5.13.a), the dataset was split based on its temporal ordering, with the first 60% of the data used for training and the remaining data sequentially used for validation and testing. This approach considered only rainfall features as input. The predictions showed relatively poor performance, as the model struggled to forecast unseen patterns in the testing sets, resulting in significant deviations from the actual values with a test MAE of 0.0123. Additionally, during the dry season, the model failed to accurately capture those mechanisms, as visualized in the training set. The challenge lies in the fact that identical rainfall patterns can result in varying responses, which the model struggles to learn.




(a) Traditional method (Rainfall features)




(b) Traditional method (Rainfall features + Boundary condition)




(c) The proposed independence-oriented time series windowing method

*Figure 5.13. Comparison of predictions among different scenarios: a) the traditional method; b) the traditional method with improved features; c) the proposed improved method*

As a result, this study considers an additional feature as a boundary condition. This feature indirectly represents mechanical, hydrological, topographical, and antecedent environmental characteristics. The rainfall features account for the dynamic changes in infiltration, storage, and drainage processes. Figure 5.13.b illustrates the predictions with these additional features while maintaining temporal ordering. At this stage, the model performed better compared to using only rainfall features, with a test MAE of 0.0057. However, due to the temporal ordering, some patterns were not captured. In response, the proposed method incorporates boundary conditions to help the model differentiate between similar rainfall patterns that yield different responses. It employs an independence-oriented time series windowing technique by shuffling the subcases, thereby removing the inter-temporal ordering between subcases while preserving the internal temporal structure within each subcase. The challenge then becomes selecting the optimal combination of training, validation, and testing sets. This is achieved through an iterative process, as described in the Model tuning section, until the best performance is obtained. Figure 5.13.c illustrates the predictions using the proposed method, which outperforms the previously mentioned methods with a test MAE of 0.0012. This figure demonstrates that the training, validation, and testing sets were selected to have similar response ranges, effectively addressing the overfitting or underfitting issues evident in the traditional method (Figures 5.13.a and 5.13.b).

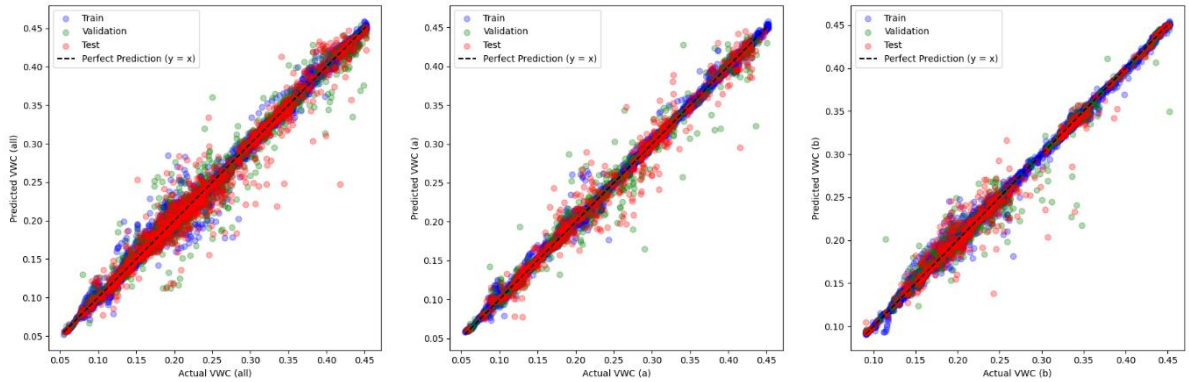
To ensure the proposed model's generability, it was applied to the 15 sensors. The prediction performance across the training, validation, and testing sets is summarized in Table 5.5. The model demonstrates good performance, with an $R^2$ greater than 0.9478, MAE lower than 0.0043, and NMAE below 0.0227. Referring to literature, such as that by Dassanayake et al. (2023), Filipović et al. (2022), and Granata et al. (2023) on irrigation and landfill capping systems, reported MAE values range from 0.0067 to 0.018, while $R^2$ values span from 0.643 to 0.946. These studies provide valuable benchmarks for evaluating predictive performance.

*Table 5.5. The evaluation metrics for the VWC site-specific model: NMAE, R2, and MAE.*

| Metrics | | NMAE | | | $R^2$ | | | MAE | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Case | Location | Train | Validate | Test | Train | Validate | Test | Train | Validate | Test |
| Pa Mei | vwc_1_a | 0.0099 | 0.0133 | 0.0147 | 0.9933 | 0.9919 | 0.9869 | 0.0011 | 0.0012 | 0.0012 |
| | vwc_2_a | 0.0105 | 0.0120 | 0.0085 | 0.9951 | 0.9946 | 0.9983 | 0.0018 | 0.0019 | 0.0014 |
| | vwc_3_a | 0.0075 | 0.0080 | 0.0083 | 0.9953 | 0.9949 | 0.9895 | 0.0006 | 0.0006 | 0.0007 |
| | vwc_4_a | 0.0092 | 0.0052 | 0.0114 | 0.9853 | 0.9949 | 0.9922 | 0.0012 | 0.0011 | 0.0011 |
| | vwc_1_b | 0.0114 | 0.0155 | 0.0148 | 0.9948 | 0.9895 | 0.9868 | 0.0021 | 0.0027 | 0.0027 |
| | vwc_2_b | 0.0056 | 0.0107 | 0.0127 | 0.9985 | 0.9919 | 0.9902 | 0.0008 | 0.0014 | 0.0015 |
| | vwc_3_b | 0.0047 | 0.0116 | 0.1362 | 0.9990 | 0.9885 | 0.9830 | 0.0003 | 0.0008 | 0.0009 |
| | vwc_4_b | 0.0047 | 0.0089 | 0.0361 | 0.9948 | 0.9777 | 0.9478 | 0.0010 | 0.0016 | 0.0018 |
| Tung Chung | vwc_1_a | 0.0101 | 0.0098 | 0.0146 | 0.9909 | 0.9926 | 0.9838 | 0.0009 | 0.0009 | 0.0012 |
| | vwc_2_a | 0.0069 | 0.0114 | 0.0100 | 0.9913 | 0.9907 | 0.9923 | 0.0013 | 0.0016 | 0.0015 |
| | vwc_1_b | 0.0102 | 0.0174 | 0.0152 | 0.9961 | 0.9788 | 0.9900 | 0.0006 | 0.0010 | 0.0009 |
| | vwc_2_b | 0.0077 | 0.0246 | 0.0206 | 0.9923 | 0.9809 | 0.9797 | 0.0026 | 0.0037 | 0.0039 |
| Tsing Shan | vwc_1_a | 0.0163 | 0.0176 | 0.0183 | 0.9798 | 0.9793 | 0.9785 | 0.0041 | 0.0046 | 0.0041 |
| | vwc_2_a | 0.0226 | 0.0165 | 0.0227 | 0.9840 | 0.9940 | 0.9853 | 0.0034 | 0.0025 | 0.0034 |
| | vwc_1_b | 0.1307 | 0.0127 | 0.0139 | 0.9856 | 0.9913 | 0.9902 | 0.0032 | 0.0030 | 0.0033 |

Operational modeling suggests that a single model cannot capture soils' heterogeneous responses, as identical rainfall patterns can trigger distinct responses depending on soil characteristics. Figure 5.14 illustrates this by comparing a single LSTM model trained on all 15 sensors (Figure 5.14.a) with a dual-LSTM approach, one LSTM for soils at depth (a) (Figure 5.14.b) and another for soils at depth (b) (Figure 5.14.c). No resampling was employed in this comparison to isolate the performance differences between the two configurations and to quantify the soil variability. The overall MAE across training, validation, and testing is 0.0051 for the single model, compared to 0.0043 for the dual-LSTM setup. This improvement is expected because soils at different depths respond differently, and a single LSTM, with one shared set of weights, cannot adequately differentiate these responses. Within each soil depth, differences among sensor locations necessitate deploying additional models to mimic the SWRC accurately. Although further segmenting the data can enhance the model's ability to mimic the SWRC, it also results in smaller training groups that are more susceptible to underfitting or overfitting. Consequently, determining the optimal segmentation is an empirical

process governed by the available dataset volume and the performance of each subgroup relative to an established benchmark. In this study, site-specific models achieving an MAE below 0.0043 were adopted as the benchmark.




(a) MAE = 0.0051 (b) MAE = 0.0042 (c) MAE = 0.0044

*Figure 5.14. Scatter plot: predictions vs. actual data (a) single LSTM (15 sensors); (b) LSTM for 0.50 m depth; (c) LSTM for 1.50 m depth*

Each model within the multi-LSTM framework was rigorously fine-tuned and trained. Figure 5.15 presents the scatter plots for each group, while Table 5.6 summarizes the corresponding evaluation metrics. Each subplot represents a group with an annotation representing the initial VWC range and the number of dry and wet patterns. Key considerations include ensuring that each group encompasses both dry and wet rainfall patterns, maintaining non-overlapping initial VWC boundaries, and balancing model performance with sufficient data volume per cluster to minimize underfitting or overfitting. The models achieve robust performance, with $R^2$ values exceeding 0.891, MAE values below 0.0031, and NMAE values under 0.0306. Moreover, the close agreement between training, validation, and testing errors indicates minimal overfitting or underfitting. These findings demonstrate that the proposed nine-group configuration yields performance metrics comparable with those depicted in Table 5.5. Although alternative grouping strategies may produce different outcomes, the benchmarks in this study were defined based on the MAE of single model performance (Figure 5.14) and site-specific performance (Table 5.5).

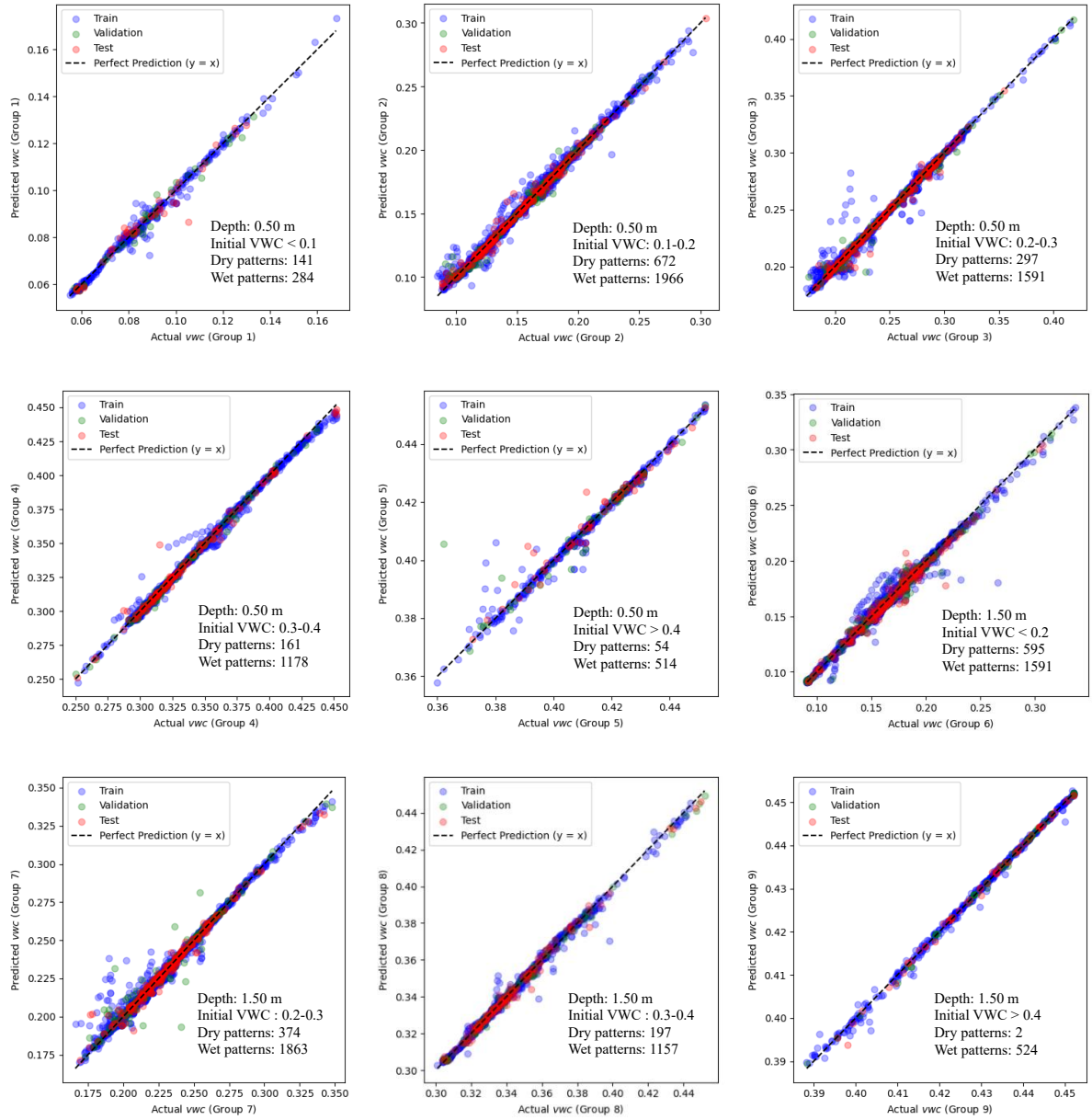



*Figure 5.15. Predicted vs. actual VWC across the multi-operational models*

To assess how this approach replicates the SWRC, feature importance was quantified using permutation (Fisher et al., 2019) (Figure 5.16). For each feature being evaluated, the values of that feature were randomly shuffled across different cases in the dataset. This means that the feature values were reassigned to different samples, thereby breaking the statistical relationship between that feature and the target variable while preserving the overall distribution of the feature. For example, if the feature was rainfall intensity at a specific time step, the intensity values were randomly permuted among all samples, so that each sample

received a rainfall intensity value from another sample's history. This disruption isolates the feature's contribution by removing its genuine correlation with the target.

*Table 5.6. The evaluation metrics for the VWC operational model: NMAE, R2, and MAE.*

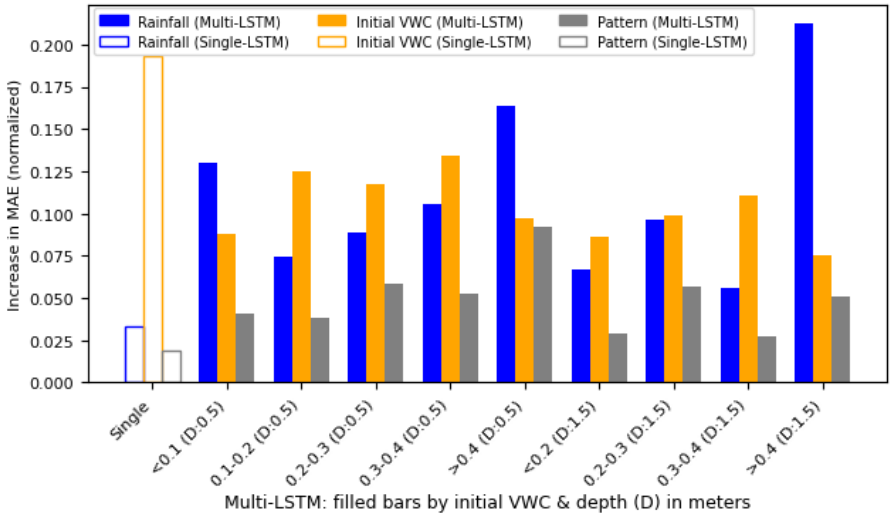
| Metrics | | NMAE | | | $R^2$ | | | MAE | |
|---|---|---|---|---|---|---|---|---|---|
| Group | Train | Validate | Test | Train | Validate | Test | Train | Validate | Test |
| 1 | 0.0120 | 0.0224 | 0.0306 | 0.9893 | 0.9859 | 0.9618 | 0.0014 | 0.0017 | 0.0022 |
| 2 | 0.0083 | 0.0110 | 0.0083 | 0.9918 | 0.9902 | 0.9928 | 0.0017 | 0.0018 | 0.0018 |
| 3 | 0.0082 | 0.0076 | 0.0111 | 0.9840 | 0.9910 | 0.9898 | 0.0020 | 0.0018 | 0.0019 |
| 4 | 0.0080 | 0.0078 | 0.0087 | 0.9950 | 0.9962 | 0.9899 | 0.0016 | 0.0015 | 0.0018 |
| 5 | 0.0124 | 0.0271 | 0.0223 | 0.9785 | 0.8910 | 0.9656 | 0.0011 | 0.0024 | 0.0018 |
| 6 | 0.0107 | 0.0128 | 0.0145 | 0.9817 | 0.9858 | 0.9757 | 0.0026 | 0.0029 | 0.0031 |
| 7 | 0.0093 | 0.0153 | 0.0119 | 0.9814 | 0.9585 | 0.9828 | 0.0017 | 0.0027 | 0.0020 |
| 8 | 0.00757 | 0.00729 | 0.0092 | 0.9929 | 0.9972 | 0.9931 | 0.0011 | 0.0011 | 0.0013 |
| 9 | 0.0077 | 0.0057 | 0.0077 | 0.9974 | 0.9987 | 0.9973 | 0.0005 | 0.0004 | 0.0004 |

To specifically assess the importance of rainfall temporal patterns, a within-case shuffle of rainfall intensities was performed. For each individual case (i.e., each rainfall event sequence), the sequence of rainfall intensity values was randomly reordered while preserving the cumulative total rainfall. This means that the same set of intensity values was retained, but their chronological order was randomized. This manipulation disrupts the temporal structure, such as whether rainfall was concentrated early or late in the event, while keeping the total rainfall volume unchanged. The increase in prediction error after this shuffle quantifies how much the model relies on the specific sequence of rainfall, rather than just the total amount. For both randomization types, performance was measured as the increase in Mean Absolute Error (MAE) relative to the baseline model, averaged over 10 independent randomization iterations to ensure statistical robustness.

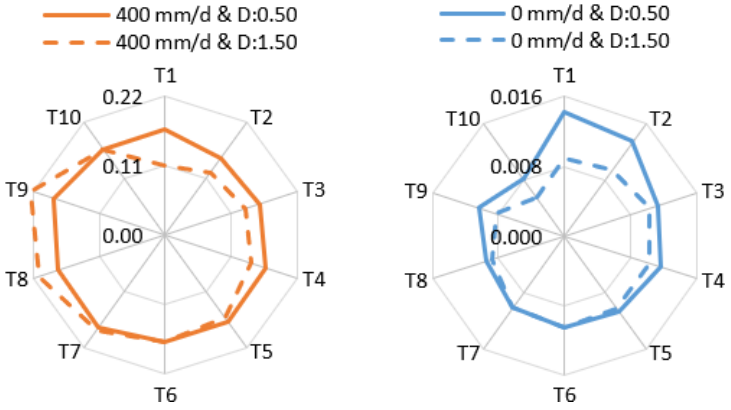
In contrast to the traditional single-model approach, which assigns uniform weights, the fine-tuned multi-LSTM model dynamically adjusts weights based on different response segments (refer to Figure 5.16.a). Specifically, the results indicate that rainfall receives higher importance in cases with low (< 0.1) and high (> 0.4) VWC, while a variant of moderate weighting is applied in the intermediate range (0.1–0.4). This differential weighting enables

the multi-LSTM model to mimic the changes in the slope of the SWRC. This approach recognizes that slope responses depend on initial conditions. Moreover, even when the cumulative rainfall is identical, variations in the rainfall pattern within each boundary lead to distinct responses.

Furthermore, each rainfall intensity within a window is weighted to reflect its role in the infiltration and seepage process. In this analysis, rainfall at each time step (T1 to T10) was set to either 0 mm/d or 400 mm/d, and the resulting average increase in normalized MAE for the multi-LSTM model is shown in Figure 5.16.b. When set to 0 mm/d, early time steps (T1–T4) exhibit slightly higher weights, especially at shallower depths (0.50 m), indicating their importance in initiating infiltration. In contrast, under high-intensity conditions (400 mm/d), later time steps (T7–T10), particularly T9 at deeper depths (1.50 m), have a greater impact, suggesting that intense rainfall later in the window is closely linked to delayed percolation and saturation processes. Figure 5.17 compares the overall performance of the proposed operational model with the site-specific model for some sensors. Table 5.7 lists the evaluation metrics. The overall performance of the proposed operational multi-LSTM model is found to be similar to the site-specific model. The MAE ranges between 0.0008 and 0.0085 for the operational model, compared to 0.0005 to 0.0042 for the site-specific model, emphasizing the effectiveness of the proposed methodology. Despite some anticipated deviations arising from the inherent assumptions outlined in Section 3.3.2, a detailed discussion is provided in the following lines.




(a) Rainfall, initial VWC, and pattern permutation-based importance




(b) Timestep ($T_{1\text{-}10}$) permutation-based importance with different intensities (0 & 400 mm/d)

*Figure 5.16. Operational VWC permutation-based evaluation of feature importance*

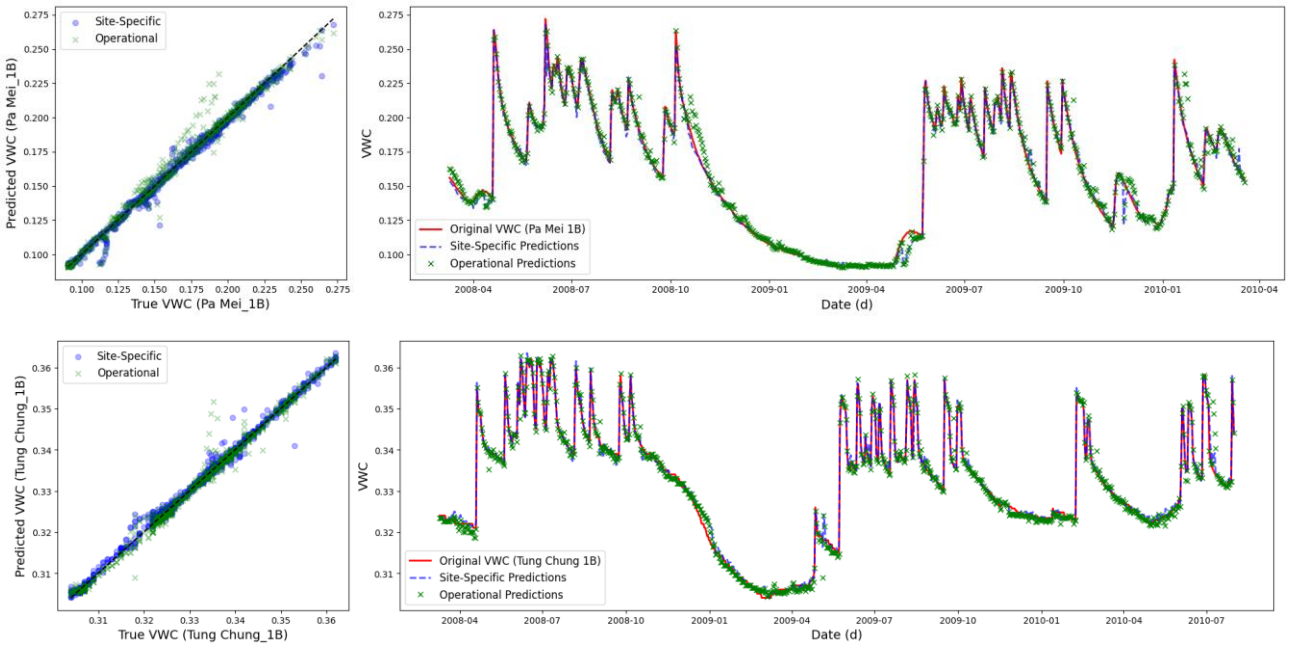
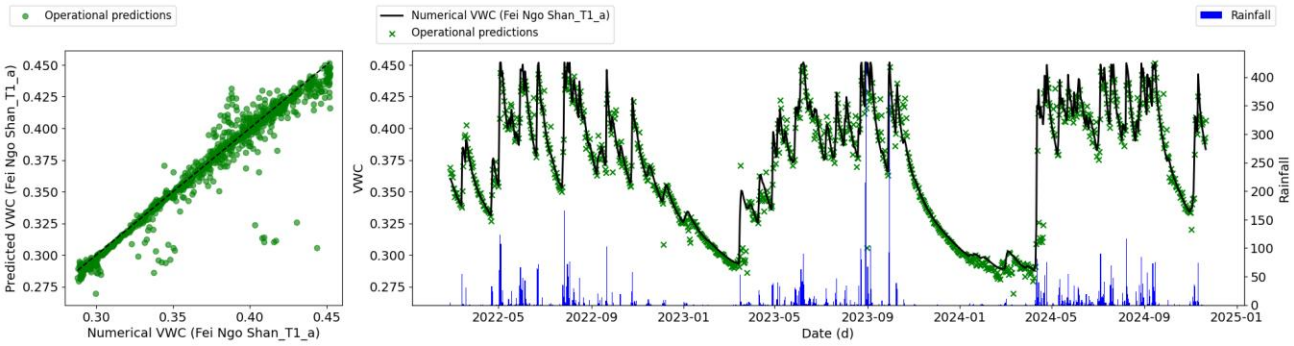



*Figure 5.17. Site-specific vs. operational predictions for some sensors as an example*

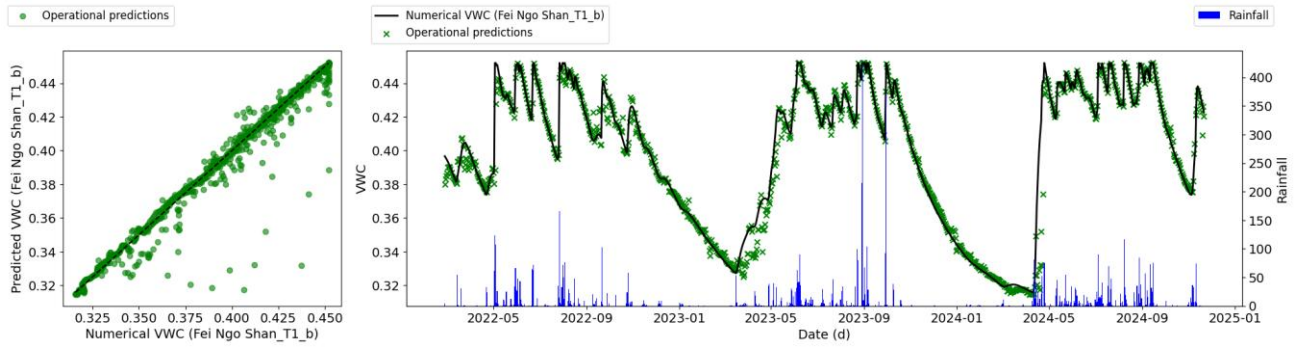
*Table 5.7. The overall evaluation metrics for the operational vs. site-specific model.*

| Metrics | | $R^2$ | | MAE | |
|---|---|---|---|---|---|
| Case | Location | Site-specific | Operational | Site-specific | Operational |
| Pa Mei | vwc_1_a | 0.9919 | 0.9736 | 0.0012 | 0.0019 |
| | vwc_2_a | 0.9956 | 0.9799 | 0.0018 | 0.0022 |
| | vwc_3_a | 0.9943 | 0.9684 | 0.0006 | 0.0015 |
| | vwc_4_a | 0.9890 | 0.9697 | 0.0012 | 0.0015 |
| | vwc_1_b | 0.9930 | 0.9857 | 0.0022 | 0.0028 |
| | vwc_2_b | 0.9962 | 0.8998 | 0.0010 | 0.0029 |
| | vwc_3_b | 0.9954 | 0.9681 | 0.0005 | 0.0015 |
| | vwc_4_b | 0.9890 | 0.6330 | 0.0012 | 0.0032 |
| Tung Chung | vwc_1_a | 0.9901 | 0.9745 | 0.0009 | 0.0016 |
| | vwc_2_a | 0.9914 | 0.9875 | 0.0014 | 0.0017 |
| | vwc_1_b | 0.9935 | 0.9900 | 0.0007 | 0.0008 |
| | vwc_2_b | 0.9890 | 0.8645 | 0.0029 | 0.0037 |
| Tsing Shan | vwc_1_a | 0.9795 | 0.9878 | 0.0042 | 0.0025 |
| | vwc_2_a | 0.9864 | 0.9938 | 0.0032 | 0.0020 |
| | vwc_1_b | 0.9872 | 0.8083 | 0.0032 | 0.0085 |

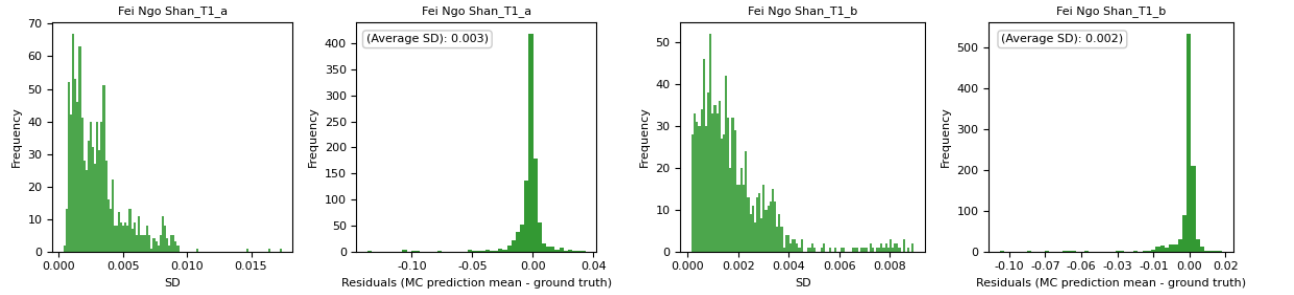
To assess the model's deterministic performance, numerically simulated patterns over 1000 days (February 2022 to November 2024) at T1 were utilized (refer to Figure 4.6). Figure 5.18. (a-b) presents the operational predictions at depths of 0.50 m and 1.50 m, where the model exhibits robust predictive capability with minor deviations. It is noteworthy that training used randomly shuffled subcases, whereas testing employed temporally ordered cases, maintaining a roughly equivalent training-testing ratio with minor overlap. This approach demonstrates how the adopted data handling and fine-tuning methodology enhances the prediction performance. The operational model exhibited superior performance at the 1.5 m depth (Figure 5.18.b) compared to the shallower 0.50 m depth (Figure 5.18.a), despite both using the same operational model. Specifically, the model achieved an MAE of 0.0060 at 0.50 m and 0.0031 at 1.50 m, along with $R^2$ of 0.921 and 0.955, respectively. The discrepancy highlights that deeper soils exhibit more stable moisture profiles. In contrast, shallow VWC responses show higher variability, driven by near-surface sensitivity to rainfall, suggesting incorporating additional features (e.g., evapotranspiration) may further enhance model performance.



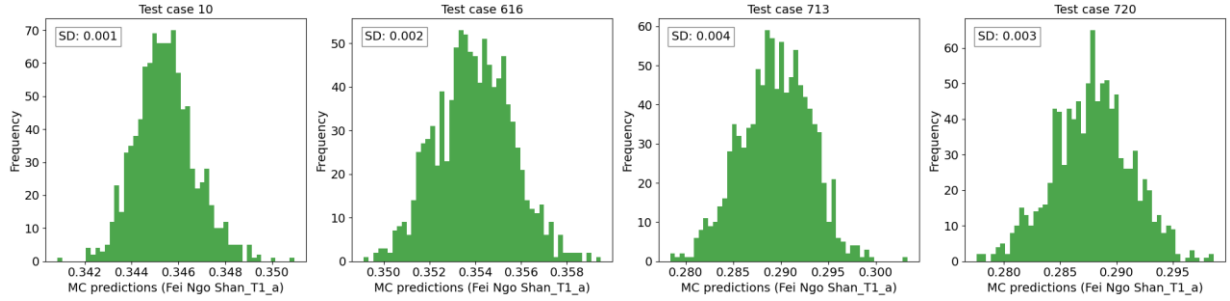

(a) Operational predictions vs. numerically simulated VWC at depth 0.50 m.



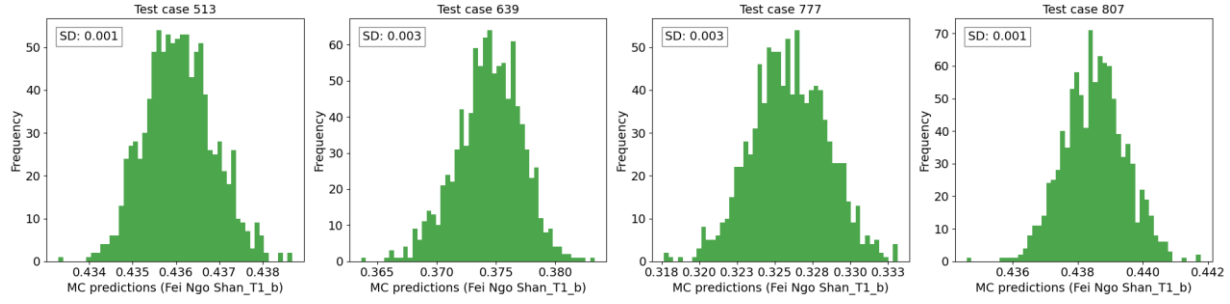

(b) Operational predictions vs. numerically simulated VWC at depth 1.50 m.




(c) MC uncertainties and residuals of the mean MC prediction distributions across test cases at 0.50 m and 1.50 m.




(d) MC predictions frequency from T stochastic simulations for four randomly selected test cases at 0.50 m.




(e) MC predictions frequency from T stochastic simulations for four randomly selected test cases at 1.50 m.

*Figure 5.18. Deterministic and stochastic evaluation of the operational model performance*

This study employs MC dropout to quantify predictive uncertainty in scenarios that may involve heterogeneous soil properties. For each subcase, T = 1000 simulations were performed, resulting in 2,000,000 simulations (2000 subcases × 1000 simulations). Figure 5.18.c presents histograms of the standard deviations (SD) for each subcase and the histograms of residuals at depths of 0.50 m and 1.50 m. The residuals exhibit predominantly bell-shaped distributions with an average SD of approximately 0.003. The low residual errors across test cases underscore the model's predictive performance. Most of the cases exhibit uncertainty values clustering around the reported mean of approximately less than 0.004, indicating consistent predictive performance. The results of the MC align well with the expected variability range. Specifically, applying a single model across all 15 sensors yields an MAE of approximately 0.005 (see Figure 5.14.a), which reflects the inherent soil variability in the study area. This performance is anticipated where the region exhibits similar soil responses, further validating the approach toward developing an operational tool.

Although most test cases exhibit uncertainties near the mean value of approximately 0.003, a small subset displays notably higher uncertainty with a slight skew in the distributions, as shown in Figure 5.18.c. This may be attributed to transitional hydrological conditions. Figures 5.18.a-b further confirm that the largest deviations occur during the soil's infrequent transition, only one or two events per year, from fully dry to partial saturation. Because these transitional states are sparsely represented in the primarily sensor-based training set, the model exhibits higher errors and uncertainty. Targeted data collection during these rare hydrological events is therefore essential to enrich the training database and improve generalization. Figure 5.18.e-f illustrates the MC stochastic prediction histograms for randomly selected samples. These distributions are generally bell-shaped, indicating an approximate Gaussian behavior and a consistent central tendency across the predictions. However, the variances differ among cases due to diverse triggering conditions, boundary constraints, and soil heterogeneity, with

the multi-LSTM approach minimizing the related uncertainties. Notably, the validity of this tool is inherently constrained by its reliance on predefined data boundaries, a limitation intrinsic to the data-driven methodology. Future data collection efforts to capture a broader spectrum of variability could enhance the model's generalizability, accuracy, and robustness.

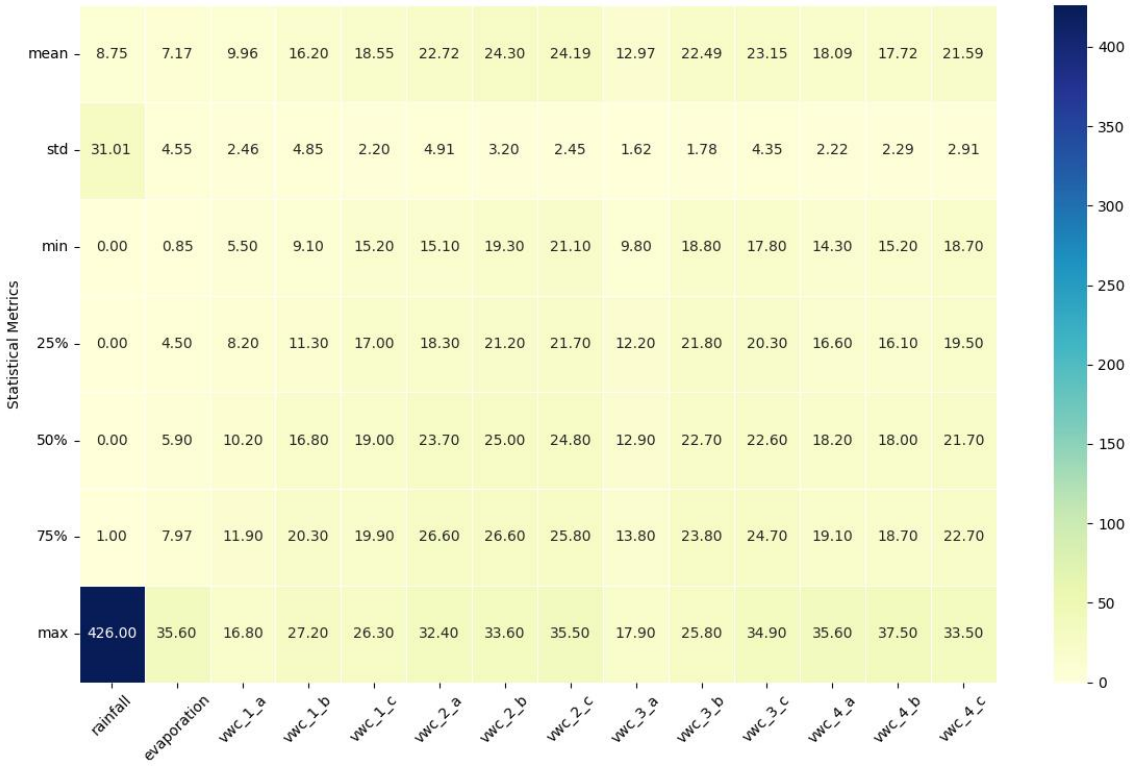
### *5.3.2 Hypertuning Strategy for Enhanced Model Performance.*

To extend predictive capability to deeper soil layers (e.g., 2.50 m) and enhance robustness across heterogeneous sites, this section introduces a unique hypertuning framework that synergizes empirical DL optimization with physical infiltration processes. The framework explicitly targets unmonitored rainfall-related features to provide both single and interval predictions, addressing monitoring-system limitations through integrated analysis of multiple subsurface hydrological and mechanical properties. Table 3.3 outlines comparative contributions, highlighting how this work advances and complements existing research.

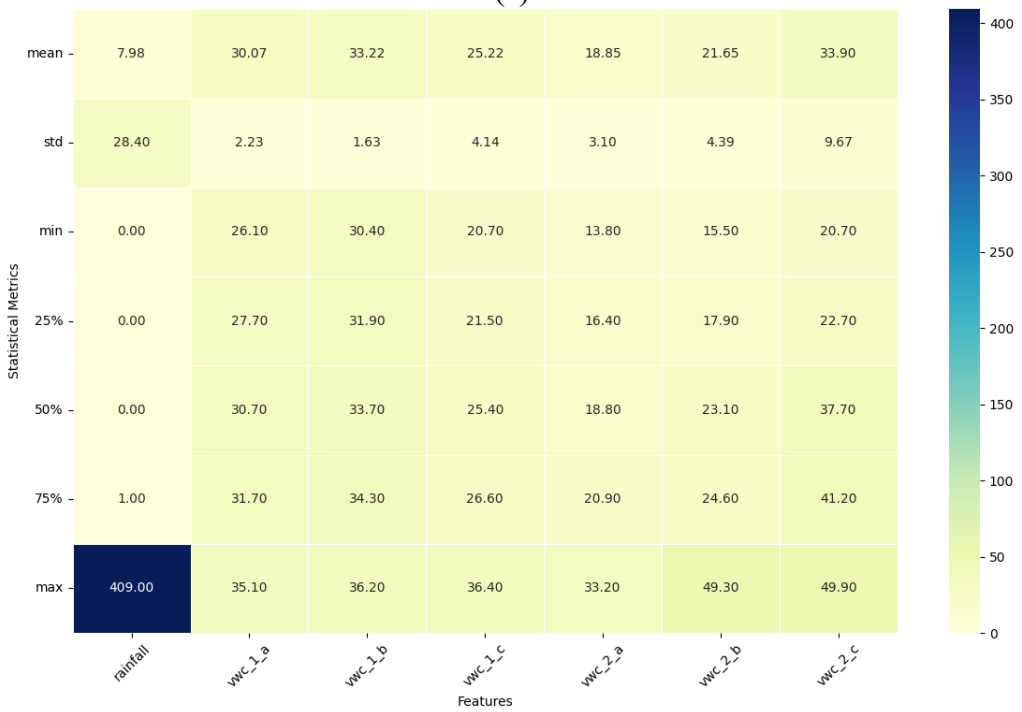
#### a) Model implementation

The key statistics of each input feature are shown in Figure 5.19, which shows that the rainfall distribution varies from the other features and that the data should be normalized. The evaporation data collected at Pa Mei have some missing data. The augmented Dickey-Fuller (ADF) test was used, where the value of p is smaller than 0.1. Since stationary data serve as the foundation for linear filling, these data are stationary, with missing data computed by linear filling. As a result, the missing during the first half of 2008 is assumed to be comparable to 2009. A 549-day record (from March 2008) split into a 400-day training set (~73%) and a 149-day validation set (~27%) was used, aligning with the conventional 70:30 ratio while preserving time-series continuity. This window captures multiple wet–dry cycles for robust pattern learning and provides an unbiased holdout. Although additional data are available, restricting analysis to this validation period ensures consistent inter-case comparisons and mitigates bias from predominantly dry intervals that can distort evaluation metrics. Figure

5.20's box plots (red: training; blue: validation) show largely comparable feature distributions, with only a few extreme values.




(a)




(b)

*Figure 5.19. Statistical characteristics of the input Features: a) Pa Mei; b) Tung Chung*

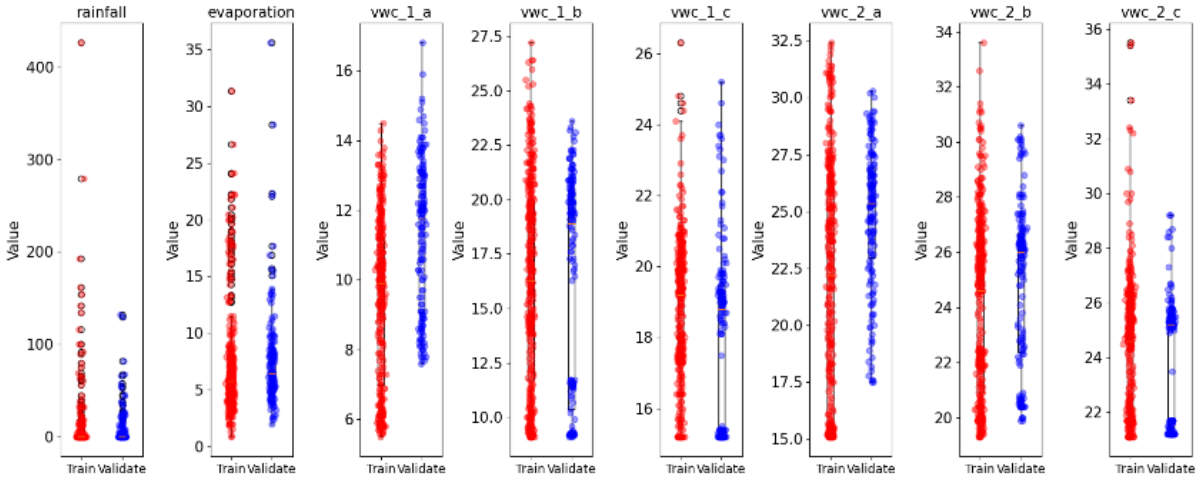



*Figure 5.20. Data distributions for training (red) and validation (blue) sets from some sensors in the Pa Mei as an example. Y-axis: rainfall and evaporation (mm/day); VWC (%)*

This study employs Kendell, Pearson, and MIC methods to investigate the link between the two variables (Table 5.8). Rainfall shows a modest positive connection at all depths, ranging from 0.23 to 0.42 (Kendall's coefficient). This suggests that heavier rainfall increases soil moisture content at various depths. While evaporation does not exhibit a strong association (i.e., from 0.10 to 0.19), Ahmed et al. (2021) revealed that evapotranspiration greatly influences seasonal soil hydrologic behavior. These variables (VWC_1, VWC_3, and VWC_4) have substantial positive correlations with VWC_2, with coefficients ranging from 0.68 to 0.94 (Kendall's coefficient). These findings highlight the need to address meteorological parameters (e.g., rainfall) and soil features (e.g., spatial moisture content at various depths) when studying VWC dynamics.

*Table 5.8. Statistical correlation among the input features and the output target at Pa Mei.*

| Feature | Correlation with the values of WTDR_2 as a target | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Sensor A: Depth = 0.50 m | | | Sensor B: Depth = 1.50 m | | | Sensor C: Depth = 2.50 m | | |
| | Kendell | Pearson | MIC | Kendell | Pearson | MIC | Kendell | Pearson | MIC |
| Rainfall | 0.42 | 0.34 | 0.34 | 0.32 | 0.37 | 0.25 | 0.23 | 0.41 | 0.22 |
| Evaporation | 0.19 | 0.26 | 0.22 | 0.16 | 0.26 | 0.20 | 0.10 | 0.21 | 0.18 |
| VWC_1 | 0.72 | 0.87 | 0.68 | 0.81 | 0.92 | 0.92 | 0.82 | 0.90 | 0.94 |
| VWC_3 | 0.69 | 0.85 | 0.62 | 0.74 | 0.88 | 0.73 | 0.83 | 0.87 | 0.96 |
| VWC_4 | 0.78 | 0.86 | 0.69 | 0.72 | 0.76 | 0.84 | 0.89 | 0.90 | 0.98 |

The antecedent condition has been demonstrated to influence the VWC response to rainfall. As a result, new features were added to enhance the model's performance, as shown in Table 5.9. These findings imply that rainfall history, particularly over extended periods, can improve VWC prediction, which is understood by the window size (antecedent effect). The hydrological conductivity of the landslide primarily determines the antecedent effect. This effect has been thoroughly discussed in the literature, where the proper levels of antecedent rainfall varied among cases. The literature lacks, however, a sensitivity analysis that looks at how DL can successfully extract features. The model tuning process, visualized in Figure 5.21, reveals exceptional performance characteristics: loss curves exhibit a steep decline during initial epochs, stabilize into a plateau around epoch 100, and maintain consistent low loss values thereafter. The close convergence of training and validation curves provides strong evidence of minimal overfitting, while the marginally lower validation loss compared to training loss suggests appropriate model complexity relative to dataset characteristics.

*Table 5.9. Pearson correlations for different rainfall antecedent periods at the Pa Mei.*

| Rainfall antecedent period (days) | Pearson correlation (WTDR_2 as target) | | |
|---|---|---|---|
| | Sensor A: Depth = 0.50 m | Sensor B: Depth = 1.50 m | Sensor C: Depth = 2.50 m |
| 1 | 0.34 | 0.37 | 0.41 |
| 5 | 0.58 | 0.57 | 0.59 |
| 10 | 0.67 | 0.65 | 0.65 |
| 15 | 0.69 | 0.69 | 0.69 |
| 20 | 0.69 | 0.71 | 0.72 |
| 25 | 0.68 | 0.70 | 0.72 |
| 30 | 0.67 | 0.70 | 0.72 |

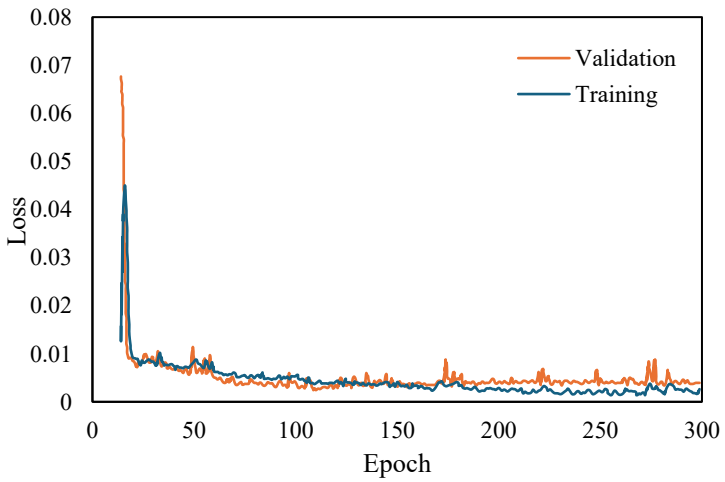



*Figure 5.21. Training and validation learning curves*

### b) Model results

The objective of the algorithm comparison is to evaluate which models can effectively handle the VWC time series relative to the naïve model and to establish a model benchmark. Figure 5.22 compares several models where the red dashed line reflects the naïve model as a benchmark for comparison. Consistent with the No Free Lunch theorem, model results demonstrate that DL does not substantially outperform shallow ML with limited training data, highlighting that tuning and data quality are more critical than algorithm choice (Tehrani et al., 2022; Merghadi et al., 2020). The results exemplify that simpler models, such as LR, and specific neural network architectures, such as CNN-LSTM, perform well in handling the dynamics of the VWC. "LSTM_Norm_Reg" surpasses all other models on a variety of measures. It has the lowest MSE, RMSE, and MAE and the highest $R^2$ and PRC, making it the ideal choice for this prediction task, consistent with Nava et al. (2023).

Several features were evaluated to determine how they could influence model performance and prediction outcomes. Table 5.10 depicts the comparison between integrating (rainfall), (rainfall and evaporation), (rainfall, evaporation, and VWC_1), and (rainfall, evaporation, VWC_1, VWC_3, VWC_4). Two benchmarks were used for comparison: the naïve model and the LSTM model, in which the output feature was VWC_2. Notably, incorporating rainfall, evaporation, and neighboring VWC data significantly improves model performance, with lower MSE, RMSE, MAE, and SDR values and higher $R^2$ and PRC. Considering the neighboring and spatially related monitored data, it is seen to be effective as it outperforms the Naïve model and LSTM model benchmarks. The sensitivity analysis was performed with two cases: case (1), in which the VWC_2 was considered as a feature to examine the maximum performance (feature benchmark) of the features, and case (2), in which a blind study was performed to select the features that outperform the naïve model and are close to the maximum performance in case (1). Rainfall and evaporation give reasonable results; however, they are still poorer than the naïve model. This can be illustrated as follows: adjacent sensors take into account a variety of unmonitored factors, including surface runoff, plant transpiration, spatial variability, hydraulic response, and other complicated mechanisms.

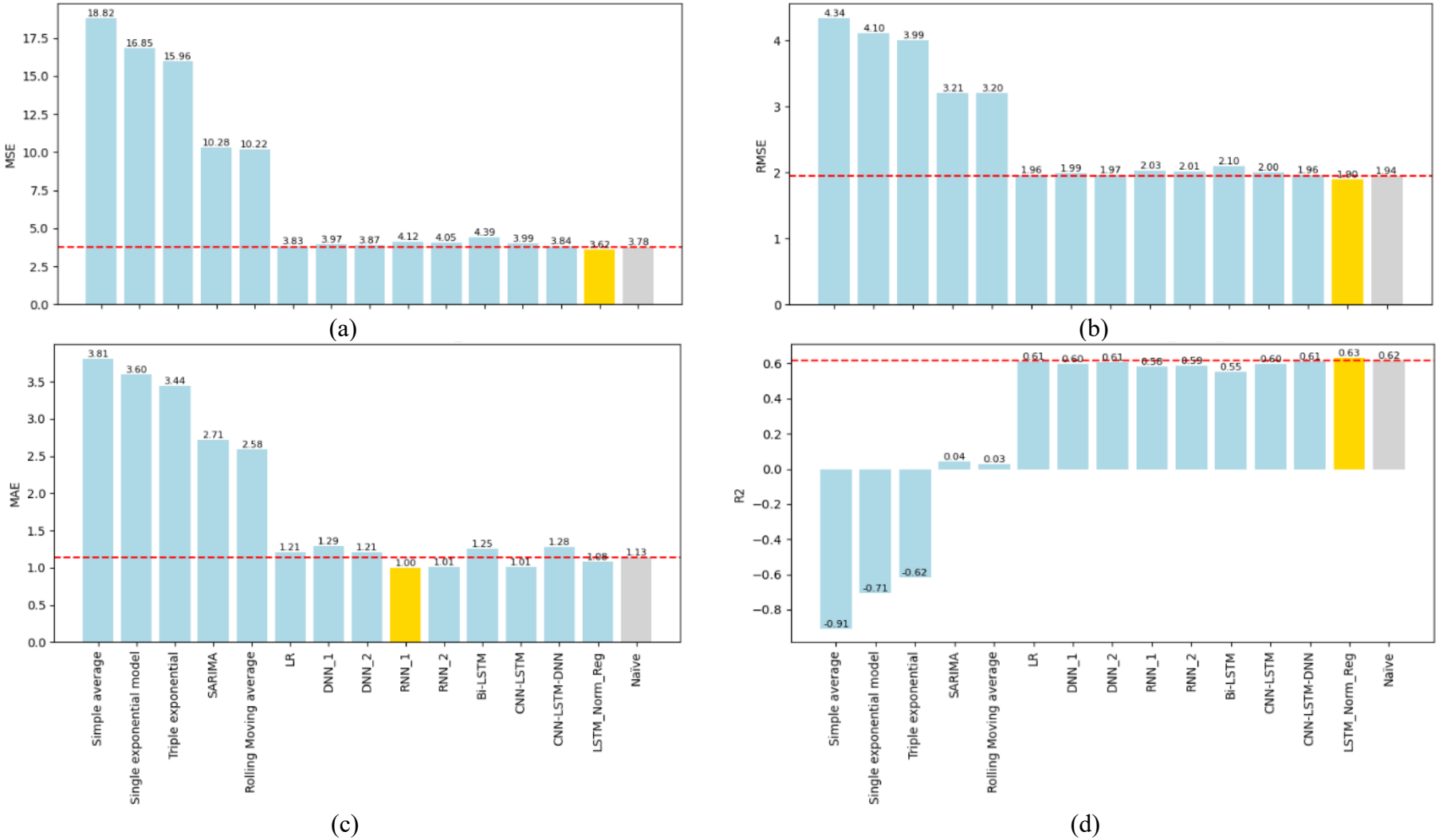



*Figure 5.22. Model selection comparison: a) MSE; b) RMSE; c) MAE; d) R2*

*Table 5.10. Feature selection comparison.*

| Case | Features | Metrics (Target VWC_2) | | | | | |
|---|---|---|---|---|---|---|---|
| | | MSE | RMSE | MAE | R2 | SDR | PRC |
| - | Naïve model benchmark | 3.78 | 1.94 | 1.13 | 0.62 | 1.94 | 0.81 |
| | LSTM model benchmark | 3.62 | 1.90 | 1.08 | 0.63 | 1.88 | 0.80 |
| 1 | Rainfall + VWC_2 | 2.97 | 1.72 | 0.97 | 0.70 | 1.72 | 0.84 |
| | Rainfall and evaporation + VWC_2 | 3.14 | 1.77 | 1.07 | 0.68 | 1.77 | 0.83 |
| | Rainfall, evaporation, VWC_1 + VWC_2 | 2.93 | 1.71 | 1.05 | 0.70 | 1.71 | 0.84 |
| | Rainfall, evaporation, VWC_1, VWC_3, and VWC_4 + VWC_2 | 2.90 | 1.70 | 1.11 | 0.71 | 1.70 | 0.84 |
| 2 | Rainfall | 4.17 | 2.05 | 1.40 | 0.59 | 2.00 | 0.81 |
| | Rainfall and evaporation | 4.01 | 2.00 | 1.49 | 0.59 | 2.00 | 0.77 |
| | Rainfall, evaporation, VWC_1 | 3.48 | 1.87 | 1.40 | 0.65 | 1.85 | 0.81 |
| | Rainfall, evaporation, VWC_1, VWC_3, and VWC_4 | 3.56 | 1.89 | 1.30 | 0.64 | 1.85 | 0.81 |

A sensitivity analysis was performed to examine the effect of the antecedent period. This study is unique as, besides the statistical methods above, the LSTM model was employed to extract such delayed effects through several trials, as shown in Figure 5.23.a-c. Most importantly, the results show that the best antecedent period for the shallow depths (A and B) is ten days, while the optimum value for the deeper depths (C) is 20. To illustrate, the shallow depths are more responsible for rainfall than the deeper locations, as the infiltration process takes time to reach deeper soil and also takes time to discharge away from the soil.

Model tuning was performed independently for each factor to investigate if there was a trend in selecting the suitable search dimensions. Figure 5.24 shows the loss against each parameter. Figure 5.24.a highlights the results for the tuning with different initial weights. This figure elucidates that selecting the optimum parameters is challenging, as the initial weights significantly affect the loss value. The dropout, alpha, batch size, and hidden units show no trends; however, increasing the return sequence of hidden layers causes a loss increase. Therefore, a random search method was employed to search for the optimum parameters (dropout, alpha, and hidden units) (Figure 5.25). The same behavior is found in the dependent search, where the best model has no clear trend, and several combinations can give approximate results. This is illustrated by the initial assigned weights by the Adam optimizer and the early

stopping technique. For this reason, reasonable parameters are selected from these results and visualized in Figure 3.12.

So far, the ideal and shortest allowable length of the training set in landslide prediction models has yet to be studied. Several training and validation sets were studied to investigate how different ratios affect the model's performance (Figure 5.26). It can be seen that the training set between 50% and 70% has better performance than other sets, where a ratio of 50% hits the best performance. This finding is illustrated by the 2008 wetting season, where there was a significant amount of rainfall between March and October (i.e., around 240 days), then from November to February (i.e., around 120 days), there was no significant rainfall, which means that the model is well-trained during these periods. Reducing or increasing the training set during non-rainy times will not affect the training process; however, it will increase the validation points, lower the average, and improve the performance metrics. The study evaluated various loss functions and optimizers (Figure 5.27), identifying the Adam optimizer with MSE loss as the most effective combination. Using these final parameters. Figure 5.28 illustrates the actual and predicted VWC alongside the 95% confidence interval and the scatter plots. Most VWC values fall within the prediction boundaries, with minor deviations. The results, evaluated using PICP metrics, indicate better performance for deeper VWC predictions compared to shallow depths. This is attributed to the shallow layers' complex interactions with rainfall, runoff, erosion, and spatial variability, whereas deeper layers exhibit a more consistent response to infiltrated rainfall.

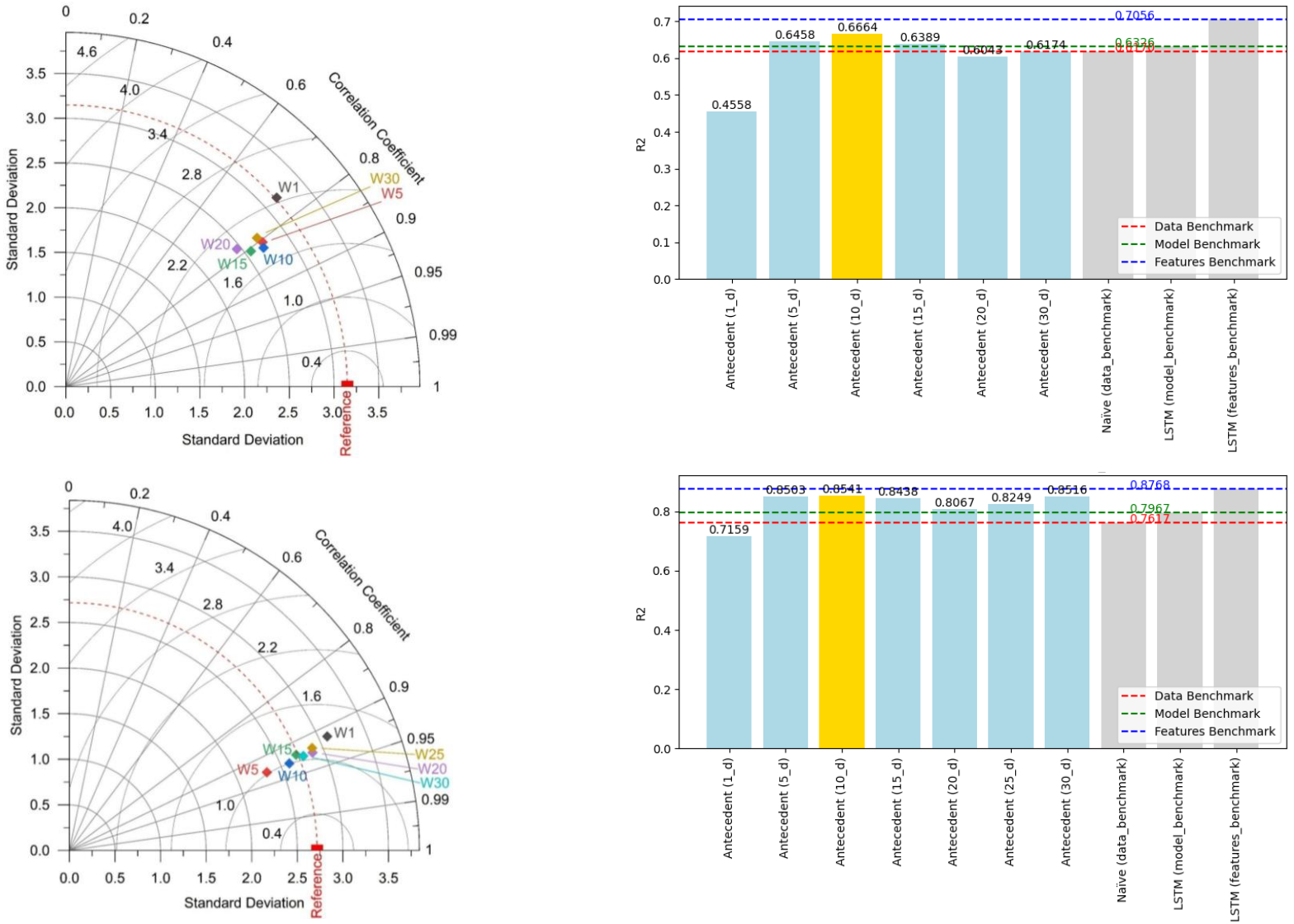



*Figure 5.23. Antecedent window size (W) selection: a) depth A; and b) depth B: (right panel) Taylor diagram; (left panel) R² vs. benchmarks*

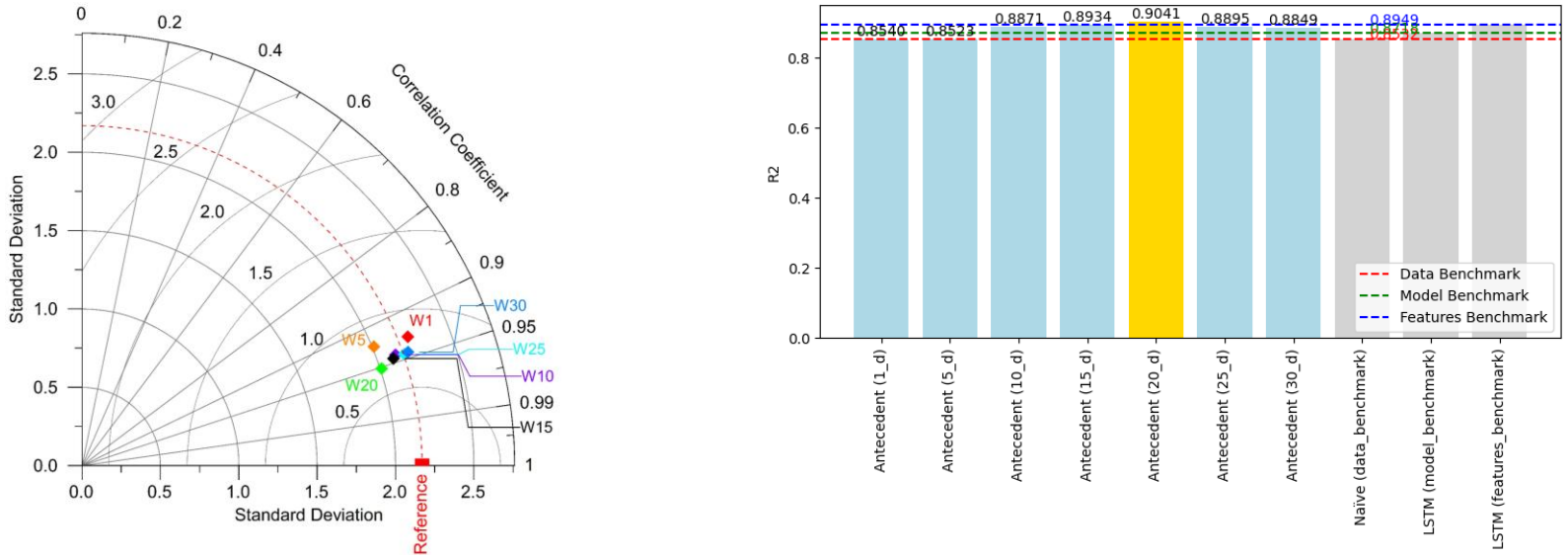
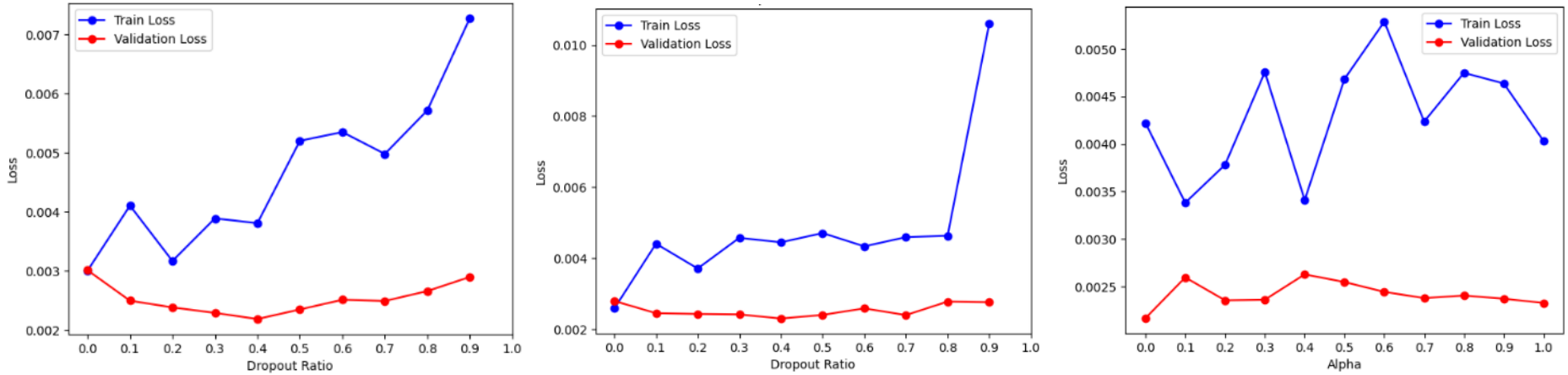
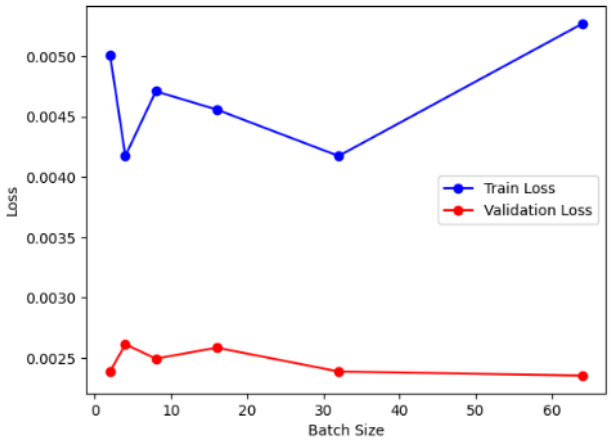



*Figure 5.23. Cont. c) depth C: (right panel) Taylor diagram; (left panel) R² vs. benchmarks*

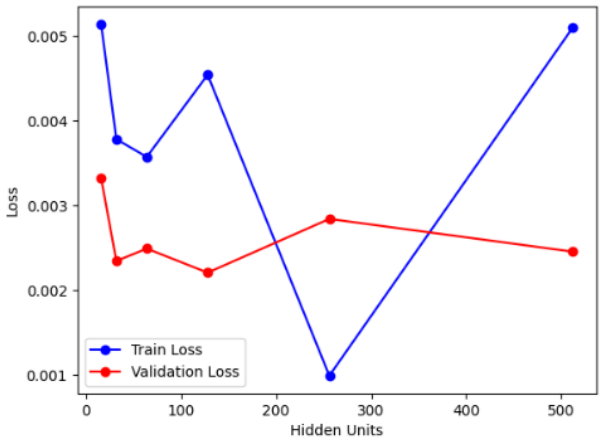



a) Dropout: First run b) Dropout: Second run c) Leaky ReLU (Alpha)

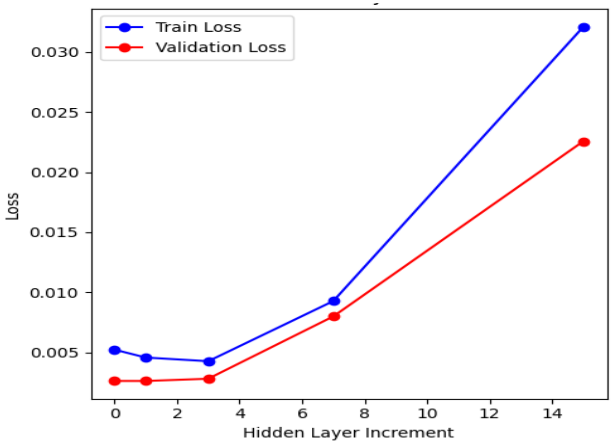
*Figure 5.24. Independent tuning search: a) dropout (1st run); b) dropout (2nd run); c) Leaky ReLU (VWC_2_C).*




d) Batch size




e) Hidden units (neurons)




f) Return sequence layers

*Figure 5.24. Cont.: d) batch size; e) hidden units; and f) return sequence layers (VWC_2_C).*

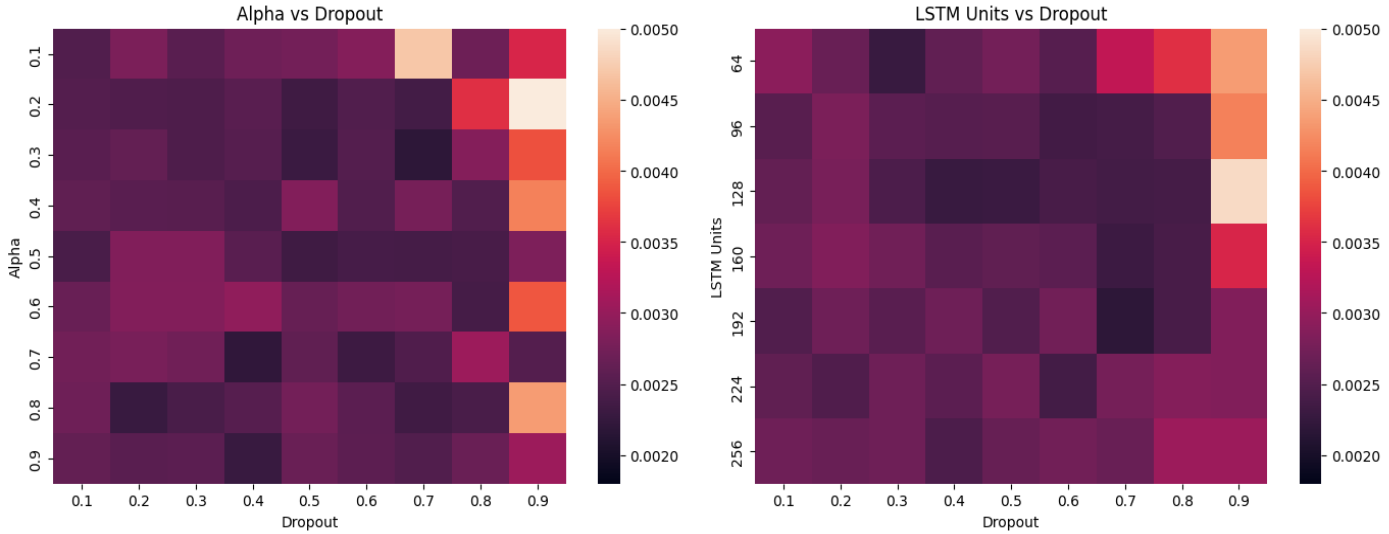
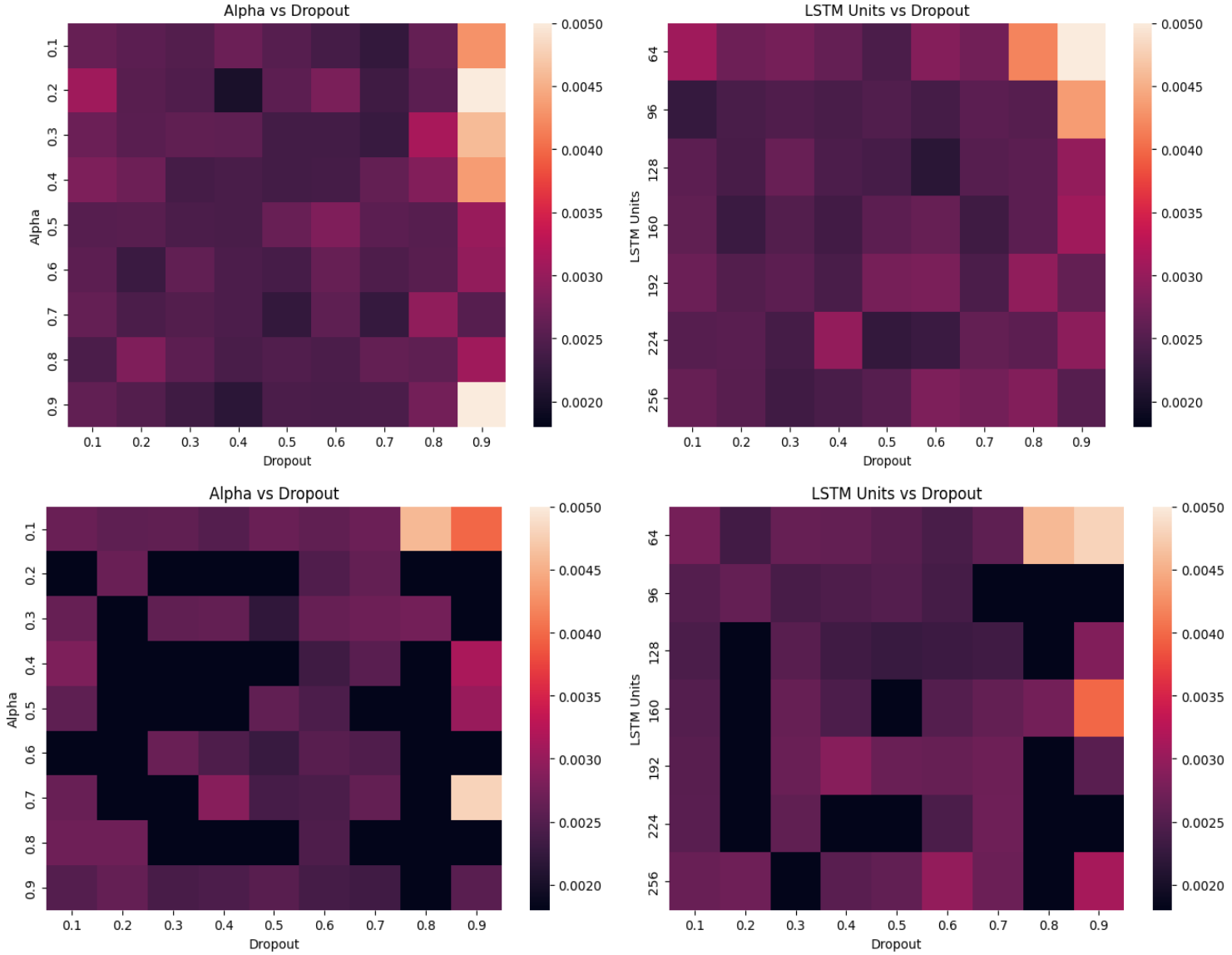


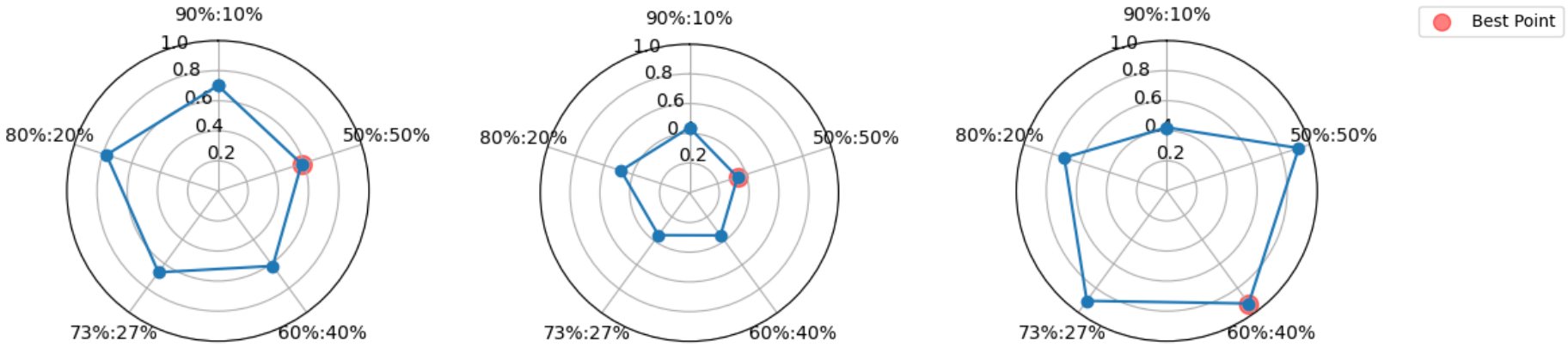

*Figure 5.25. Dependent tuning search heat map using random search (VWC_2_C): a) 1st trial random search*




(b)

(c)

*Figure 5.25. Cont.: b) 2nd trial random search; and c) Bayesian model search.*




a) RMSE b) MAE c) $R^2$

*Figure 5.26. Comparison between several training and validation sets (VWC_2_C).*

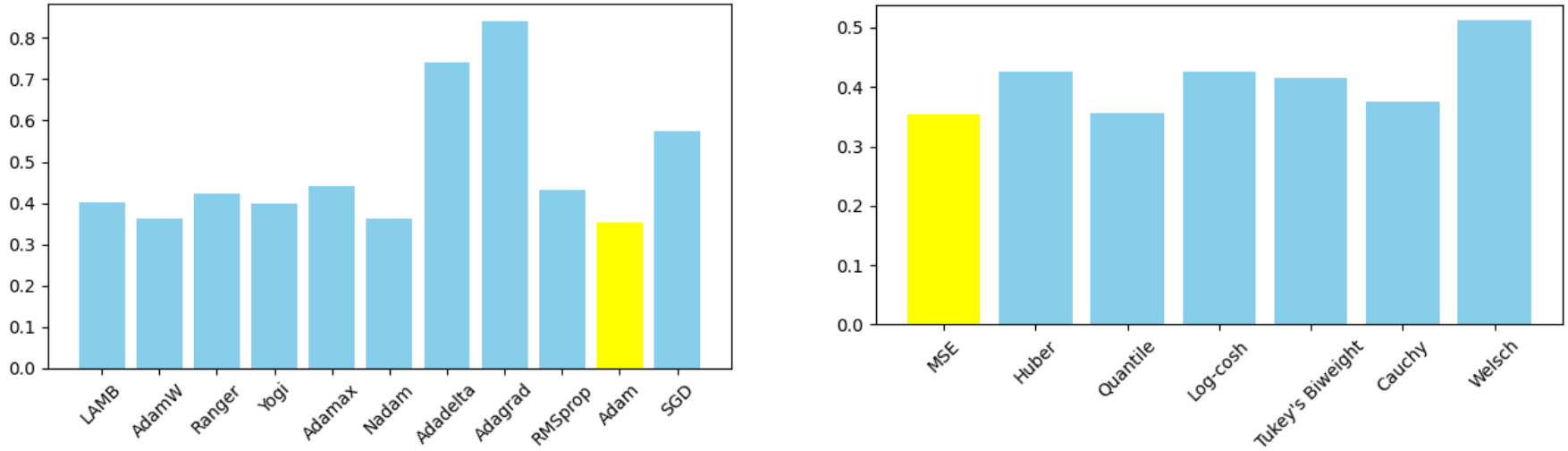
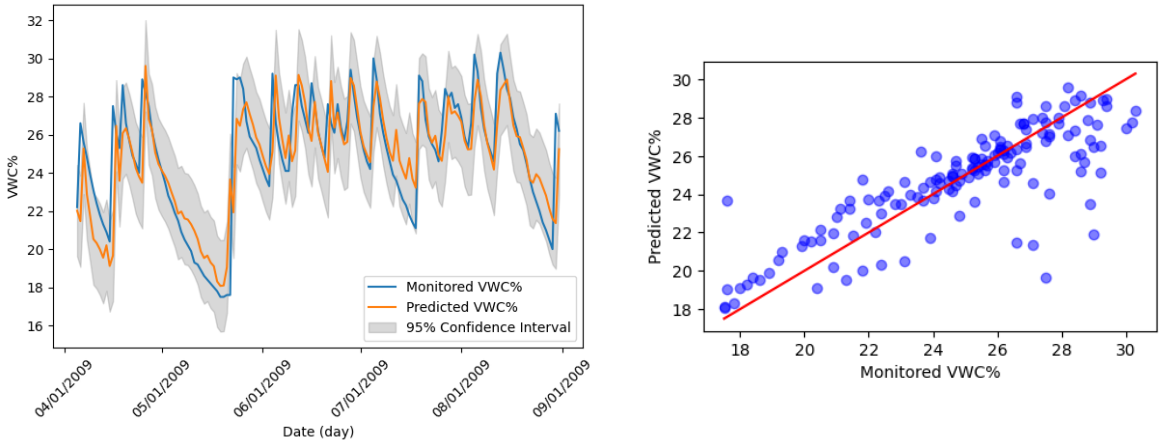
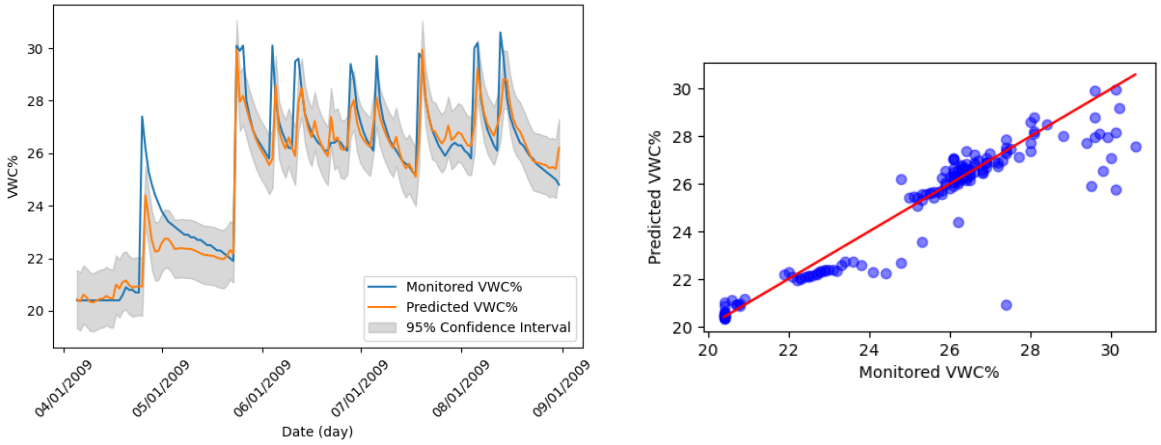


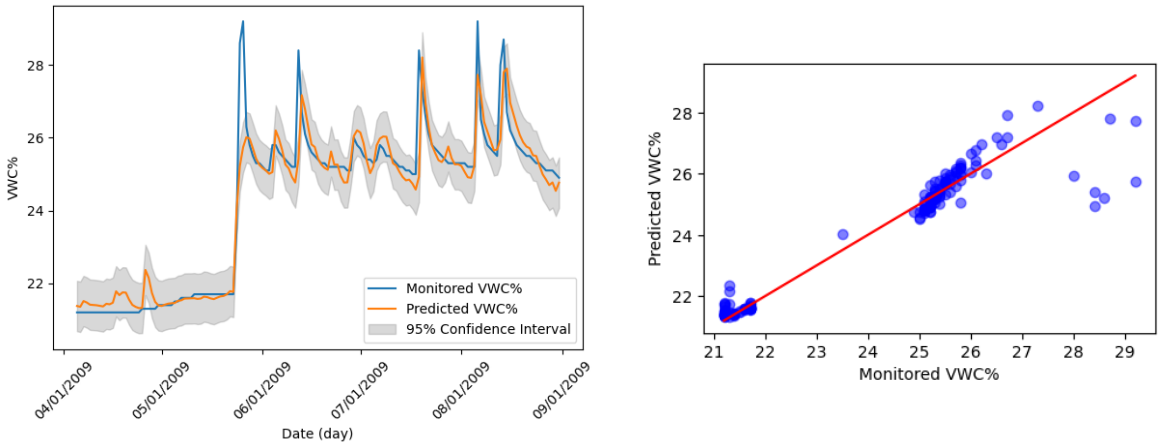

*Figure 5.27. MAE comparison between several optimization techniques (right) and several loss functions (left) (VWC_2_C).*




a) VWC_2_A




b) VWC_2_B




c) VWC_2_C

*Figure 5.28. Interval prediction and scatter plots for the VWC at three depths: (a) 0.50 m, (b) 1.50 m, (c) 2.50 m.*

Table 5.11 presents the final validation metrics, highlighting a comparison between this study and the related literature. It is found that the proposed model performs well compared

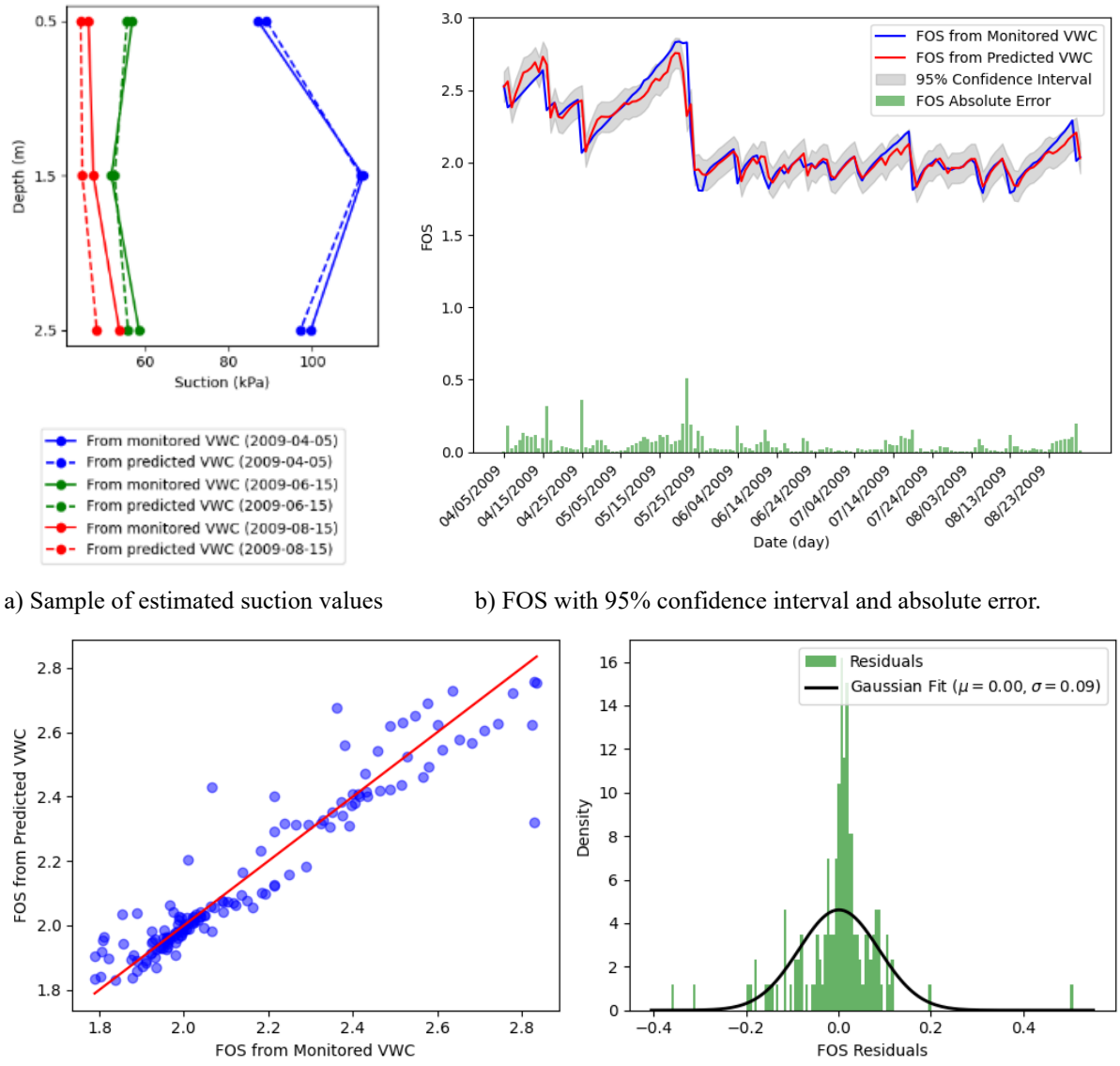
with the one-dimensional analytical transient infiltration model (Ahmed et al., 2021) and other related research about irrigation and landfill capping systems (e.g., Dassanayake et al., 2023; Filipović et al., 2022; Granata et al., 2023). The model of VWC_2_C shows better results compared to the literature available, as it improves the MAE and RMSE by around 47.53% and 23.64% compared to the best results of the related literature (Granata et al., 2023). Compared to the deterministic models (one-dimensional analytical infiltration model), the results of the VWC_2_C in this study show a good improvement in both R2 and RMSE values by a ratio of 2.87% and 33.03%, respectively. This proves that DL algorithms and the proposed model can capture the physical and internal mechanisms of such complex mechanisms. It should be emphasized that the model's results are contingent on the specific training dataset. The sharing networks allow the DL model to extract the transient infiltration mechanism and the missing unmonitored features. The complex hydrogeological settings, such as scouring, water concentration, runoff, vegetation, and so on, make it difficult to directly associate daily rainfall with daily VWC for the shallow layers (i.e., VWC_2_A), as demonstrated by Suk et al. (2022). However, neighboring monitoring can be valuable for improving model performance and making predictions for the next 24 hours, particularly for deep soils (i.e., VWC_2_C).

*Table 5.11. Comparative evaluation metrics and benchmark against related literature.*

| Metric | VWC_2_A | VWC_2_B | VWC_2_C | Ahmed et al. (2021) | Dassanayake et al. (2023) | Granta et al. (2023) | Filipović et al. (2022) |
|---|---|---|---|---|---|---|---|
| MSE | 3.2841 | 1.0698 | 0.4484 | - | - | - | - |
| RMSE | 1.8122 | 1.0343 | 0.6697 | 1 | 1.6 | 0.877 | - |
| R2 | 0.6664 | 0.8541 | 0.9041 | 0.8789 | 0.643 | 0.946 | - |
| MAE | 1.2206 | 0.5663 | 0.3531 | - | - | 0.673 | 1.8 |
| SDR | 1.8074 | 0.9984 | 0.6693 | - | - | - | - |
| PRC | 0.8184 | 0.8299 | 0.9521 | - | - | - | - |
| PICP | 0.859 | 0.879 | 0.913 | - | - | - | - |

To evaluate the overall performance of the model, the FOS for the shallow layer was calculated. Suction values, derived from the corresponding VWC measurements, were assumed

to remain steady above the measurement point. This assumption treats suction values as primarily time-dependent rather than spatially variable to address deviations arising from the predicted VWC at different depths. Figure 5.29.a shows samples of suction values derived from both monitored and predicted VWC under rainy and non-rainy conditions. The results align well with previous studies in Hong Kong, where Ng et al. (1998) reported highly variable suction values in colluvium soils, ranging from 0 to 80 kPa. Similarly, suction values monitored in related studies (Maunsell Geotechnical Services Ltd., 2007; Geotechnical Engineering Office, 2010b) are consistent with the values presented in Figure 5.29.a. The presented results and trends of suction validate the assumptions while acknowledging their simplified nature. Figure 5.29.b depicts the FOS along with the corresponding interval prediction and absolute error, where the PICP value is 0.8523. Figure 5.29.c depicts a scatter plot of the estimated FOS with the predicted VWC vs. the FOS estimated with the monitored VWC, with $R^2$, MSE, RMSE, MAE, PRC, SDR, and MBE values of 0.8916, 0.0075, 0.0865, 0.0536, 0.9443, 0.0865, -0.0001, respectively. Overall, the metrics indicate that the error in predicting VWC is not significant, demonstrating that the model performs well. Furthermore, Figure 5.29.d illustrates the residuals' distribution, which can reasonably be assumed to follow a normal distribution with minor deviations, validating the assumption of confidence intervals employed in this study. Utilizing the framework for the Tung Chung landslide case study provides a crucial opportunity to evaluate its transferability capabilities, applying the same methodology to a different scenario. The results obtained from the Tung Chung case study closely resemble those from the Pai Mei case study. Both datasets exhibit $R^2$ and MAE values ranging from 0.72 to 0.90 and 0.35% to 1.25%, respectively.




a) Sample of estimated suction values

b) FOS with 95% confidence interval and absolute error.

c) Scatter plot

d) Distribution of the FOS residuals

*Figure 5.29. FOS based on monitored and predicted VWC: a) sample of estimated suction values; b) FOS with 95% confidence interval and absolute error; c) FOS scatter plot; and d) residual distribution of FOS.*

### c) Model achievements and limitations

ML models are inherently empirical processes, heavily influenced by the selection of the model, auxiliary layers, features, evaluation metrics, and other hyperparameters. Landslide prediction, in particular, involves intricate temporal dependencies and complex dynamic mechanisms, requiring a careful and sensitive determination of the proper window size to account for these dependencies. In response, this study developed a hyperparameter-tuning

framework that balances the appropriate selection of models and optimization processes while addressing the complexities of landslide dynamics. The initial step in the framework involved testing several models using the original target variable and comparing their performance with a naïve baseline model. This step ensured that the model's performance would not be influenced by additional factors, such as selected features or window size, etc., providing a fair assessment of each model's capability. The results revealed that while multiple models showed comparable performance, the LSTM consistently delivered slightly better predictions. This justified its selection for further optimization steps, including fine-tuning of features, temporal window size, and other parameters.

The computational efficiency of the models varies with complexity. Simpler models like SARIMA and moving averages are the fastest, requiring 1–2 ms per step, making them ideal for rapid computations in straightforward scenarios. More complex models, such as DNN (2–11 ms) and RNN (4–11 ms), provide better accuracy at a moderate computational cost. Hybrid models like Bi-LSTM and CNN-LSTM take 20–45 ms per step, balancing accuracy and efficiency. The selected LSTM model, while the most computationally expensive at 39–95 ms per step, performs well in capturing complex dependencies, since all models were trained using Google Colab with a Python 3 Google Compute Engine backend, indicating that none of the models require extensive computational resources.

Overfitting is a concern in ML, particularly when working with relatively small datasets. In response, the selection of features was guided by both statistical methods and domain knowledge, ensuring the inclusion of critical variables like rainfall and VWC while accounting for unmonitored factors (e.g., surface runoff and transpiration). Neighboring sensors were also integrated to capture spatial variability and improve the model's robustness. A holdout validation strategy, combined with sliding window techniques, preserved temporal dependencies while testing various training-to-testing ratios to optimize data usage.

Hyperparameter tuning further refined the model, with dropout techniques and early stopping criteria applied to prevent overfitting. The training process was closely monitored, and the final model demonstrated acceptable performance, with loss curves stabilizing at a low plateau and validation loss aligning closely with training loss.

This study focuses on predicting VWC by addressing temporal dependencies rather than identifying controlling features, which are well-established in the literature. While cumulative rainfall values captured temporal effects (i.e., using statistical methods), this approach is unsuitable for other features (i.e., the spatial VWC) due to their spatial and temporal variability. An LSTM-based framework was employed to extract relevant features without prior assumptions, effectively handling these dependencies. Regarding the antecedent effect, the shallow colluvium layer's high permeability ($10^{-5}$ to $10^{-6}$ $m \cdot s^{-1}$) supports short antecedent periods (Au, 1998). Optimal antecedent windows of 10–20 days were identified, differences arising from site-specific subsurface conditions and boundary constraints. Natalia and Yang (2024) and Rahimi et al. (2011) show that rainfall randomness, intensity, duration, and pattern modulate these effects: delayed patterns enhance infiltration and speed FOS recovery, whereas advanced patterns delay failure. Soil conductivity further controls infiltration rate and healing time under identical hydraulic conditions, making antecedent responses highly sensitive to initial conditions and rainfall patterns.

To enhance practical relevance, it is acknowledged that, due to the challenges associated with deploying subsurface monitoring systems, decision-makers often prioritize remote sensing–based prediction and mitigation methods (Dou et al., 2025). These challenges include issues related to energy consumption, memory size, data collection, transmission, and harsh environmental conditions (Mucchi et al., 2018; Kumar et al., 2020b). Subsurface monitoring systems outperform remote sensing in both cost and accuracy. Yang et al. (2017) and Chu et al. (2021) estimated sensor system costs at roughly 1,500 USD and 1,000 USD,

highlighting subsurface monitoring's cost-effectiveness. Moreover, subsurface systems remain the only reliable method for capturing landslide internal mechanisms.

With that in mind, this study presents a DL-based framework for predicting subsurface responses, overcoming data-related challenges, enhancing subsurface monitoring applicability, and urging decision-makers to deploy these systems. Despite promising results, this study has several limitations that warrant future investigation. Only the drying branch of the SWRC was modeled; integrating the wetting path could better capture hysteresis (Johari & Hooshmand, 2018). Uncertainty quantification could be improved by adopting alternative probabilistic or ensemble methods to enhance robustness. Similarly, testing additional normalization techniques and activation functions could boost DL performance. Reliance on one field dataset underscores the scarcity of subsurface monitoring data; testing more data and developing numerical models validated by this dataset to generate synthetic observations will enhance generalizability. Future work will incorporate dynamic groundwater-table fluctuations into FOS calculations to better capture transient pore-water pressures in shallow layers.

### *5.3.3 Distributed Multi-Sensor (Subsurface) Prediction Model*

Existing approaches to these monitoring challenges fall into two main categories: pre-issue and post-issue techniques, as summarized in Table 2.6. Pre-issue methods save energy by collecting data less often, using fewer sensors in high-risk areas, or improving battery systems. These approaches extend monitoring life but may collect less data and miss early warning signs. Post-issue methods estimate missing data using statistics or machine learning, assuming similar conditions across sites. However, both approaches rely on simplified relationships that cannot capture how rainfall, water, and soil interact over time to trigger landslides. With that in mind and building on the techniques described earlier, this section applies a multi-node sharing network to overcome monitoring limitations such as data-collection challenges, randomness, triggering uncertainty, and to improve model performance. This objective aims to predict the

subsurface mechanical and hydrological responses using DL models to achieve better predictions, with explanations coming in second. Then, in the case of data-related issues, it can help solve the following issues: a) enhanced the level of uncertainty due to the unavailability of rainfall controlling features by learning from best correlated hydrological and mechanical characteristics; b) missing hydrological or mechanical responses; c) subsurface hydrological and mechanical response prediction of future scenarios, supporting decision making and early warning; and d) virtual sensing and understanding the spatiotemporal correlations, aiming to reduce the number of active sensors.

a) Model implementation

This section focuses on the implementation of the developed DL model in two real-world case studies in Hong Kong and details the step-by-step process of implementing the model within these case studies as the implementation process begins with data preparation: The main statistical data are listed in Table 5.12 before normalizing and processing, emphasizing that the feature distribution varies (Bolded) with some erroneous monitoring (Reded) and that the data should be normalized and processed. For the Tsing Shan case study, the Augmented Dickey-Fuller (ADF) test was applied to assess the stationarity of the time series from sensor arrays WTDR1A, WTDR1B, and WTDR1C. The resulting p-values for WTDR1A (0.50) and WTDR1B (0.17) were both greater than the 0.1 significance threshold, indicating that these datasets are non-stationary. In contrast, the p-value for WTDR1C (0.02) confirmed its stationarity. To address the non-stationarity in WTDR1A and WTDR1B, the subsequent analysis step involved using a linearly fitted dataset as input to the LSTM model, which helped update and refine the filled data for these series. Figure 5.30 shows a sample dataset after utilizing the LSTM. The distorted data is handled by dropping out these values, and then the vailed suction values, along with the corresponding VWC, are used to fit the best parameters ( $\theta_s$ , $a_f$ , $n_f$ , $m_f$ ) of the Fredlund-Xing equation (Equation 3.20). It should be emphasized that

the DL model will not be suitable for the issue, as the distorted data occurred during the dry season, when no data were available for the DL model to learn from. Figure 5.31 shows the corrected time series using the proposed model for the JFT1B, for instance.

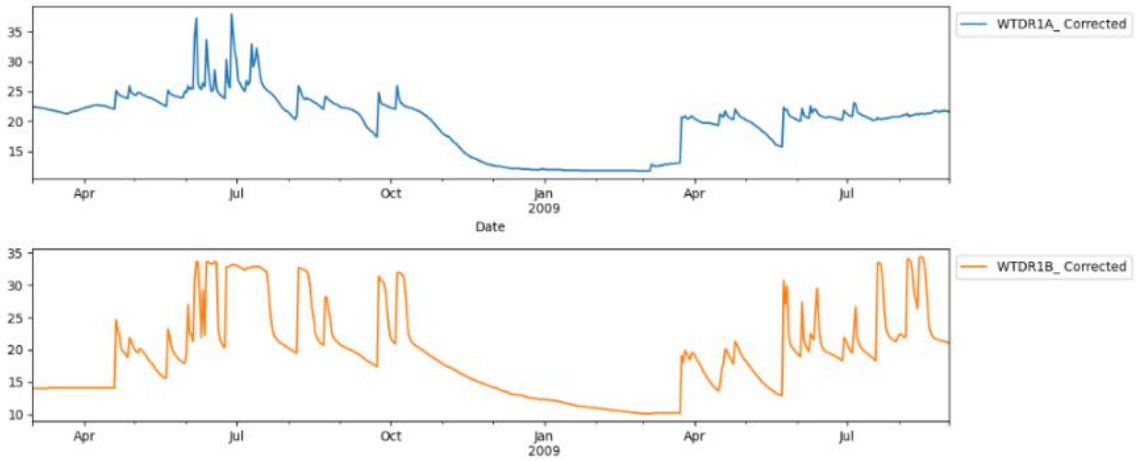



*Figure 5.30. The corrected VWC (WTDR1A, B) - Tsing Shan. The Y-axis unit WTDR is %.*

*Table 5.12. Main statistical characteristics of the dataset (Tung Chung).*

| Metrics | Features | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Rainfall | EP1A | EP1B | EP2A | EP2B | JFT1A | JFT1B | JFT1C | JFT2A | JFT2B | JFT2C |
| Mean | 7.98 | -0.06 | -0.08 | 5.71 | -0.21 | -11.53 | -27.97 | -42.19 | -11.97 | -19.70 | -31.11 |
| Std | 28.40 | 2.53 | 2.13 | 6.33 | 2.86 | 15.07 | 18.92 | 21.60 | 17.84 | 18.92 | 22.81 |
| Min | 0.00 | -2.80 | -2.90 | -2.00 | -3.50 | -89.80 | -82.10 | -86.10 | -83.90 | -89.10 | -88.90 |
| 25% | 0.00 | -1.60 | -1.20 | 0.40 | -1.50 | -10.10 | -31.80 | -67.10 | -10.90 | -18.90 | -32.00 |
| 50% | 0.00 | -0.90 | -0.40 | 4.40 | -1.00 | -8.00 | -19.50 | -31.90 | -7.90 | -17.00 | -23.30 |
| 75% | 1.00 | 0.40 | 0.30 | 9.90 | -0.30 | -6.50 | -16.70 | -26.70 | -4.60 | -14.00 | -20.40 |
| Max | **409.00** | 14.40 | 13.80 | 25.70 | 14.50 | 2.60 | **-0.40** | -4.20 | 3.30 | 2.40 | 2.80 |
| Metrics | Features | | | | | | | | | | |
| | TC1A | TC1B | TC1C | WTDR 1A | WTDR 1B | WTDR 1C | WTDR 2A | WTDR 2B | WTDR 2C | BH7 | WTDR 1A |
| Mean | -61.95 | -50.00 | -49.99 | 30.07 | 33.22 | 25.22 | 18.85 | 21.65 | 33.90 | 88.17 | 30.07 |
| Std | 65.16 | 47.69 | 49.35 | 2.23 | 1.63 | 4.14 | 3.10 | 4.39 | 9.67 | 1.15 | 2.23 |
| Min | -191.3 | -154.1 | -164.2 | 26.10 | 30.40 | 20.70 | 13.80 | 15.50 | 20.70 | 87.76 | 26.10 |
| 25% | -91.70 | -62.90 | -48.50 | 27.70 | 31.90 | 21.50 | 16.40 | 17.90 | 22.70 | 87.77 | 27.70 |
| 50% | -21.30 | -21.20 | -22.90 | 30.70 | 33.70 | 25.40 | 18.80 | 23.10 | 37.70 | 87.77 | 30.70 |
| 75% | -18.20 | -19.70 | -20.90 | 31.70 | 34.30 | 26.60 | 20.90 | 24.60 | 41.20 | 87.77 | 31.70 |
| Max | -17.10 | -18.90 | -19.80 | 35.10 | 36.20 | 36.40 | 33.20 | 49.30 | 49.90 | 95.39 | 35.10 |

Regarding data splitting and temporal ordering, this research used a training-to-testing ratio of (73%:27%). Figure 5.32 illustrates the training (Red circles) and validation (Blue circles) datasets with a box plot. The data distributions in the training and testing sets are comparable, except for a few extreme values. Figure 5.33 demonstrates the exceptional performance of the model since the loss curves are declining, particularly in the early epochs,

hitting a plateau around epoch 150, and staying stable at a low loss. Additionally, the proximity of these curves to one another suggests that overfitting is not present. However, it should be noted that the training process may have some noise and oscillation due to small batch sizes, Adam optimization, and random dropout techniques. These noises were addressed through fine-tuning the model by monitoring training and validation loss curves and using early stopping with model checkpointing to select the best model.

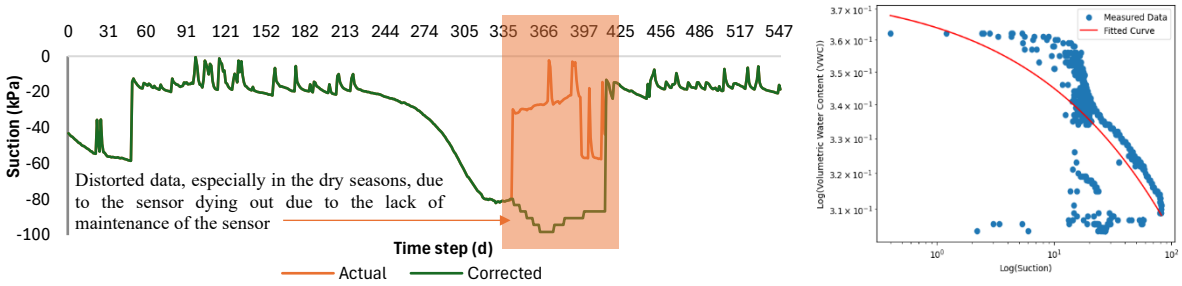



*Figure 5.31. a) Actual versus corrected JFT1B; b) SWCC fitted using Fredlund-Xing eq.*

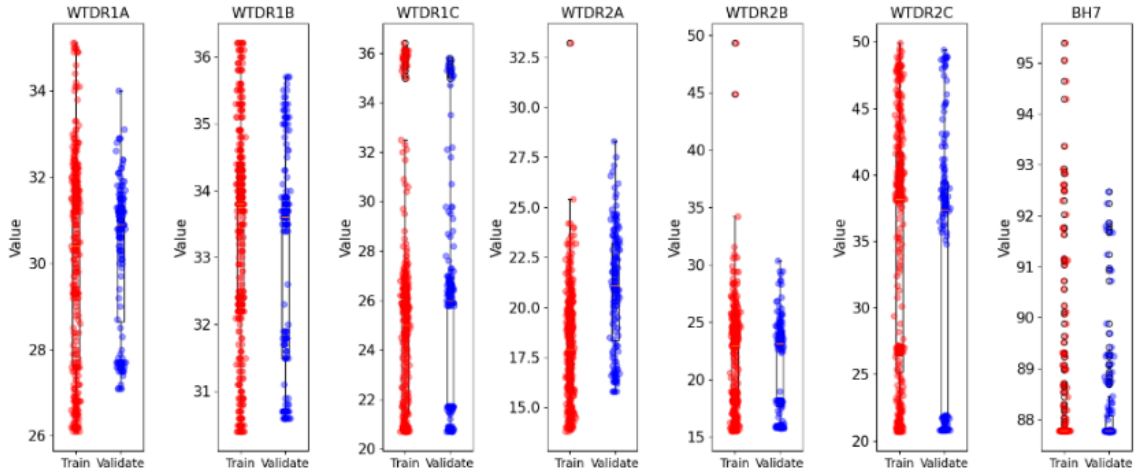
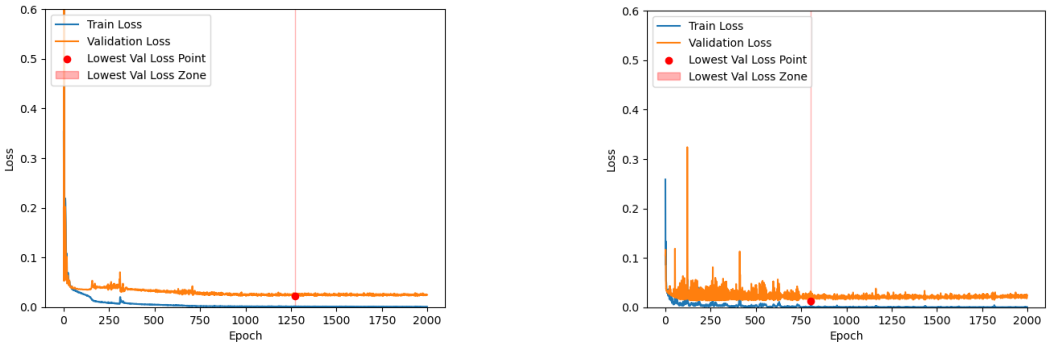



*Figure 5.32. Data distribution in the training and validation sets with a ratio of 73%:27%, for WTDR sensors as an illustration – Tung Chung.*

As for feature selection, Figure 5.34 shows the Pearson correlation among the original dataset features. Knowing that the Pearson correlation cannot account for the temporal correlations among these features, Table 5.13 sensitively examines and lists the Pearson correlation between amended rainfall as a feature and the other responses as a target. It is shown that the cumulative sum of the antecedent rainfall sequence (window size) significantly correlated with other features. However, this temporal correlation is not necessarily a

cumulative sum. As a result, a sensitivity analysis is performed for each model using LSTM to examine such complex temporal correlations.




a) Batch size = 500 b) Batch size = 16

*Figure 5.33. The training learning curve with different batch sizes: a) 500 and b) 16.*

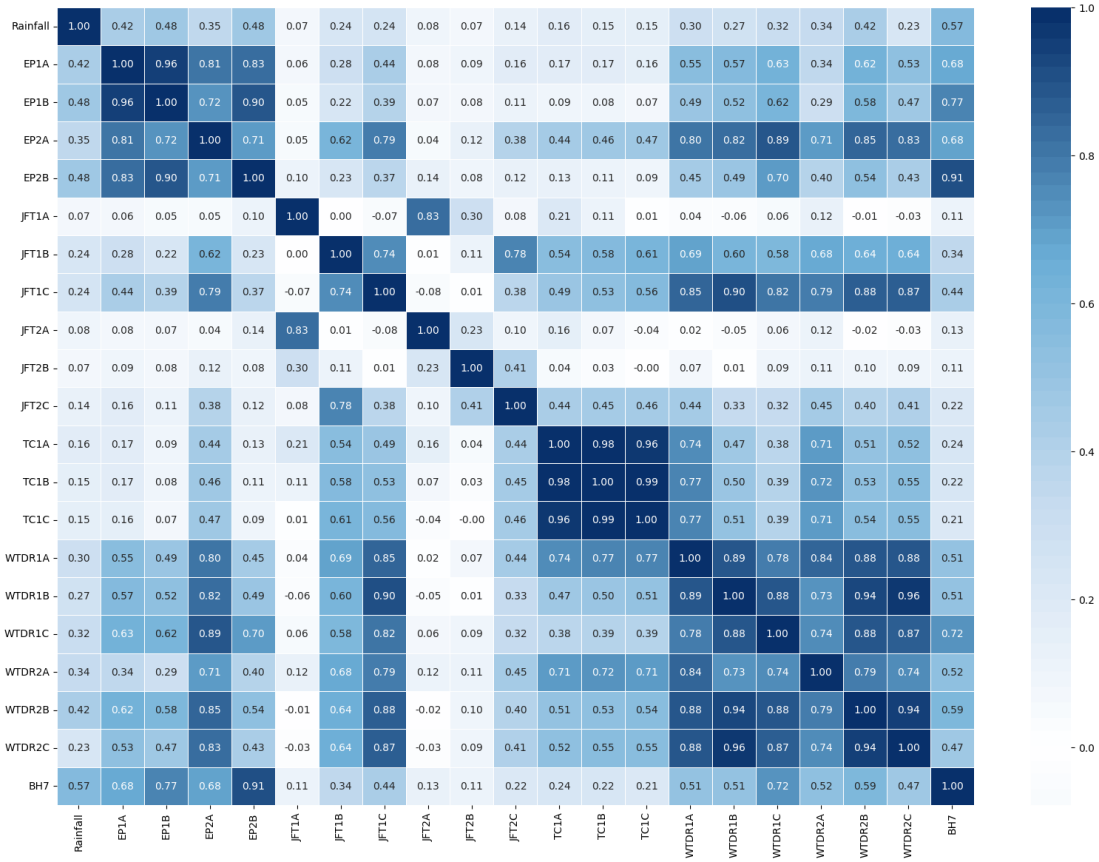



*Figure 5.34. Pearson correlation coefficient among the input features and the output target.*

It should be emphasized that the insights from Figure 5.34 assisted in selecting the best features and the knowledge-based method to be integrated into the LSTM (Table 3.5). Moreover, the dimensions of the antecedent period illustrated in Table 5.13 helped to select the window size trials using the LSTM (Table 3.6). To illustrate, rainfall shows an extremely low

to low positive connection at EP, JFT, TC, VWC, and BH7, ranging from 0.48 to 0.07 (Figure 5.34). On the contrary, as shown in Table 5.13, rainfall with antecedent effect shows a moderate to high correlation at EP, VWC, and BH7, ranging from 0.80 to 0.57. These data show that rainfall history, particularly over long periods, may significantly enhance EP, JFT, TC, and VWC forecasts. The Pearson correlation is extremely low to moderate, ranging from 0.13 to 0.62 at the JFT and TC sensors, necessitating the consideration of new features with LSTM, such as those listed in Table 3.5.

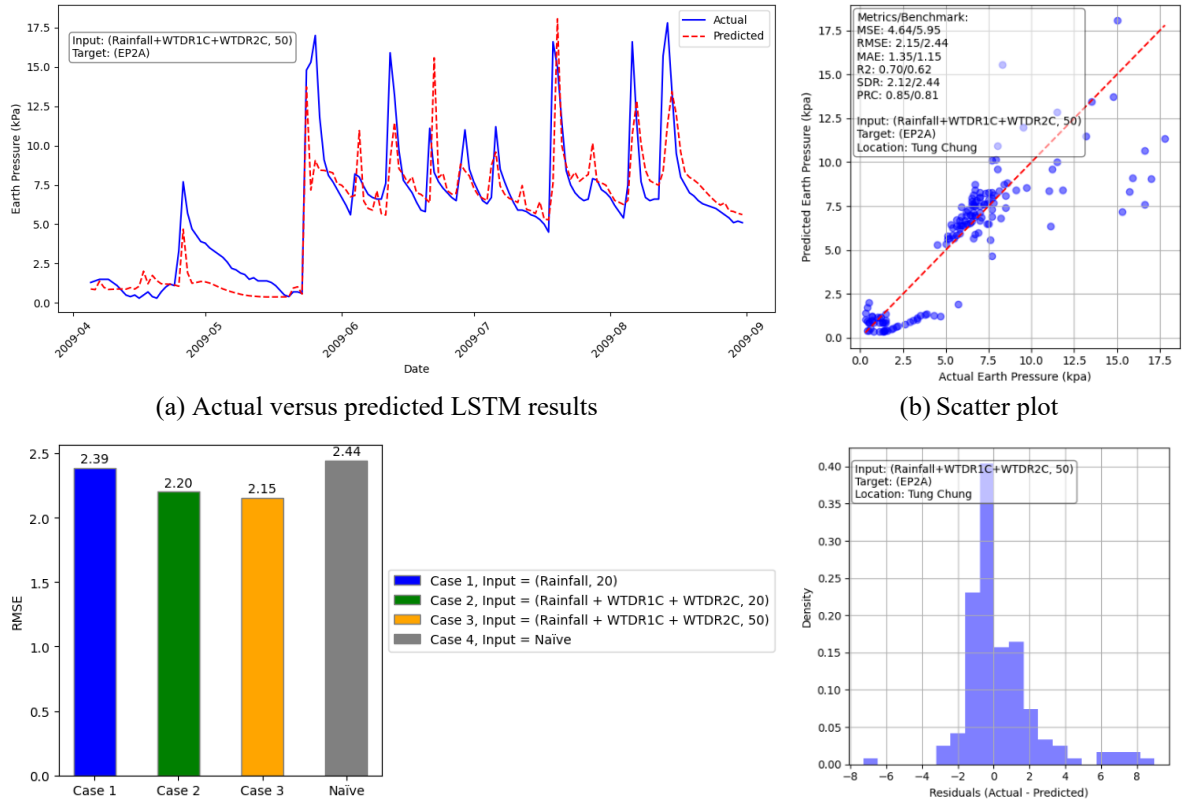
*Table 5.13. Pearson coefficient among the input amended features and the output target.*

| Target | Amended feature (Rolling window size of rainfall - days) | | | | | |
|---|---|---|---|---|---|---|
| | 0 | 10 | 20 | 30 | 40 | 50 |
| EP1A | 0.42 | 0.67 | **0.72** | 0.70 | 0.69 | 0.67 |
| EP1B | 0.48 | 0.61 | **0.64** | 0.61 | 0.59 | 0.55 |
| EP2A | 0.35 | 0.70 | 0.78 | 0.79 | 0.80 | **0.80** |
| EP2B | 0.48 | 0.64 | **0.65** | 0.58 | 0.54 | 0.47 |
| JFT1A | 0.07 | 0.13 | **0.14** | 0.12 | 0.10 | 0.07 |
| JFT1B | 0.24 | 0.43 | 0.48 | 0.49 | 0.50 | **0.51** |
| JFT1C | 0.24 | 0.51 | 0.58 | 0.60 | 0.61 | **0.62** |
| JFT2A | 0.08 | **0.15** | 0.14 | 0.11 | 0.07 | 0.02 |
| JFT2B | 0.07 | 0.12 | **0.13** | 0.12 | 0.11 | 0.10 |
| JFT2C | 0.14 | 0.30 | **0.34** | 0.33 | 0.33 | 0.33 |
| TC1A | 0.16 | 0.39 | **0.44** | 0.44 | 0.43 | 0.42 |
| TC1B | 0.15 | 0.38 | 0.43 | **0.44** | 0.44 | 0.44 |
| TC1C | 0.15 | 0.36 | 0.42 | 0.44 | 0.44 | **0.45** |
| WTDR1A | 0.30 | 0.59 | 0.67 | 0.68 | **0.69** | 0.69 |
| WTDR1B | 0.27 | 0.58 | 0.64 | 0.65 | **0.67** | 0.67 |
| WTDR1C | 0.32 | 0.65 | **0.70** | 0.69 | 0.68 | 0.66 |
| WTDR2A | 0.34 | 0.55 | **0.60** | 0.58 | 0.57 | 0.56 |
| WTDR2B | 0.42 | 0.61 | 0.68 | 0.68 | **0.69** | 0.69 |
| WTDR2C | 0.23 | 0.55 | 0.62 | 0.64 | 0.66 | **0.67** |
| BH7 | **0.57** | 0.57 | 0.57 | 0.50 | 0.47 | 0.41 |

### b) Earth pressure (EP) response

Table 5.13 illustrates that the best temporal sequence is 20 days for the earth pressure, except for EP2A, which slightly improves with a 50-day sequence. The DL LSTM was modeled considering these temporal dependencies and incorporating the spatially varied responses. Figure 5.35.a shows the time series plot for the best LSTM where the target is EP2A and the input features are: rainfall, WTDR1C, and WTDR2C. Figure 5.35.b presents the scatter plot and six evaluation metrics compared to the naïve model. Figure 5.35.c compares different input

scenarios to emphasize how spatially nearby VWC can improve the model performance and account for the rainfall uncertainties. Figure 5.35.d presents the error distribution plot for the best-fitted LSTM model, highlighting that the predicted results have a minor error distribution. To illustrate, the results show that integrating spatially varied VWC improves the prediction, as visualized in the comparative plot in Figure 5.35.c. This is because the subsurface water content considers the spatial uncertainties of the rainfall, surface runoff interaction, and vegetation effect, among other unmonitored complex features, illustrated earlier. In this comparison, two factors are studied: a) considering the VWC as a spatially nearby feature that best accounts for the rainfall uncertainties (i.e., (rainfall only) and (rainfall + WTDR2C+WTDR1C)), and b) the temporal dependences between these features (i.e., the window size of 20 days and 50 days). Considering RMSE as a comparison metric for the EP2A sensor, the VWC improves the model performance by around 8% compared to rainfall only. As for the temporal dependencies, longer sequences, as found by the Pearson correlation analysis, can further improve the performance to a ratio of around 12% compared to the naïve one.




(a) Actual versus predicted LSTM results

(b) Scatter plot

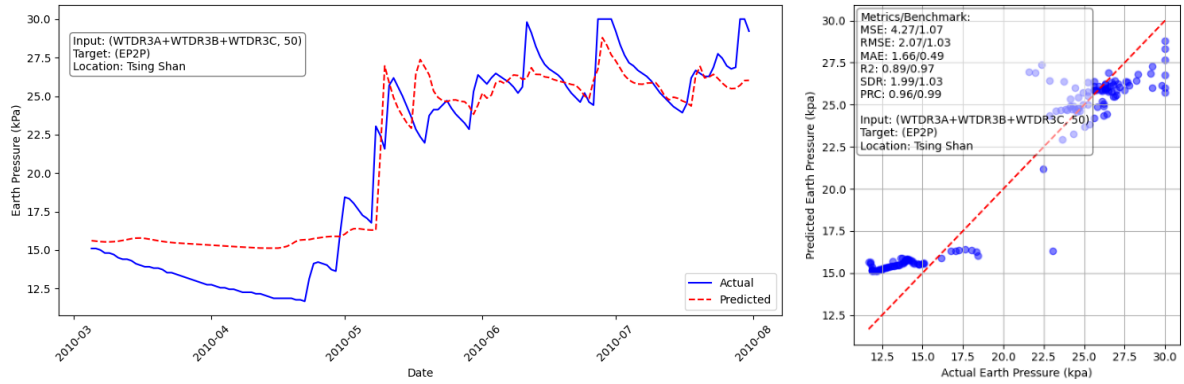
(c) EP2A comparative analysis of several input scenarios

(d) Error distribution plot

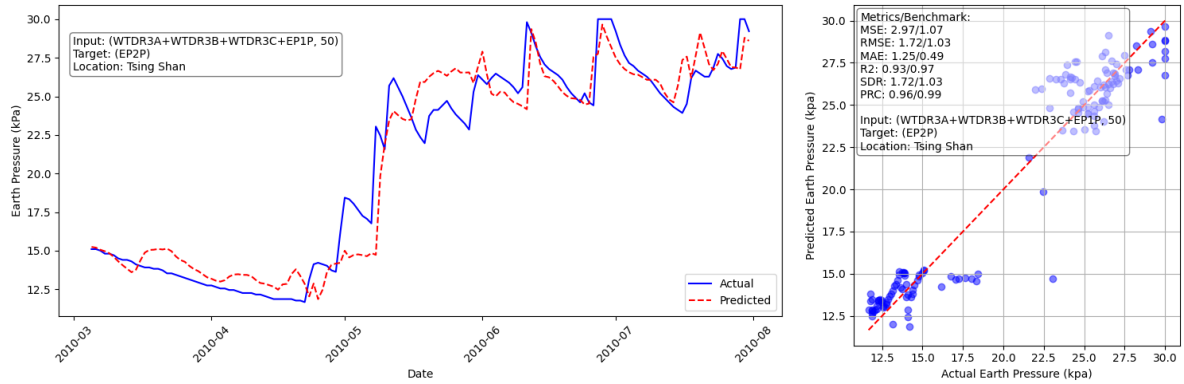
*Figure 5.35. Prediction analysis of EP2A at Tung Chung.*

Another case study (Tsing Shan) is examined to check the model's generality. Figure 5.36.a shows the predicted and the actual results, while Figure 5.36.b shows the scatter plot and the best metrics obtained for EP2P. The rainfall data was missing while monitoring EP1P and EP2P. As a result, in this case, the spatially varied VWC were integrated as input features without considering rainfall. However, the WTDR at depth C shows constant values for extended intervals. This may be due to cracks, body movement near the instrumentation, water zones near sensors, lack of sensor maintenance, etc. As a result, integrating the VWC at spatially varied depths is considered to overcome such uncertainties. Figure 5.36.a proves that considering surface VWC can overcome the uncertainties that may be found at deeper sensors due to the presence of a perched groundwater level, which affects the sensor reading. Moreover, considering only VWC can reasonably predict the earth pressure with an R2 of 0.89 and an RMSE of around 2.07 kPa. The prediction accuracy is comparable with that of Tung Chung, highlighting the general predictivity of the DL model. However, the time series between March

and May 2010 shows lower prediction accuracy. The integration of nearby time series responses on the EP1P is utilized with DL to train the model well. Figure 5.36.b shows that a better prediction is possible, which indicates the capabilities of the LSTM model to understand the spatiotemporal correlation among different features. The improvement is around 17% regarding RMSE, while the R2 value reached 0.93. The same finding presented in Figures 5.35 and 5.36 is found to be generalized among six different earth pressure sensors at various locations and two different case studies. A comparison is visualized in Figure 5.37, where the LSTM model predictions are comparable to the naïve basic model. The RMSE ranges between 0.97 and 2.15 kPa with an average value of 1.62 kPa, while the naïve basic model ranges between 1.03 and 2.44 kPa with an average value of 1.63 kPa. Consequently, it is concluded that the LSTM can predict the mechanical response of a landslide by learning from the hydrological response (VWC).




(a) Actual versus predicted EP2P and the scatter plot with the evaluation metrics.




(b) Actual versus predicted EP2P considering the neighboring response of EP1P

*Figure 5.36. Prediction analysis of the EP2P at Tsing Shan.*

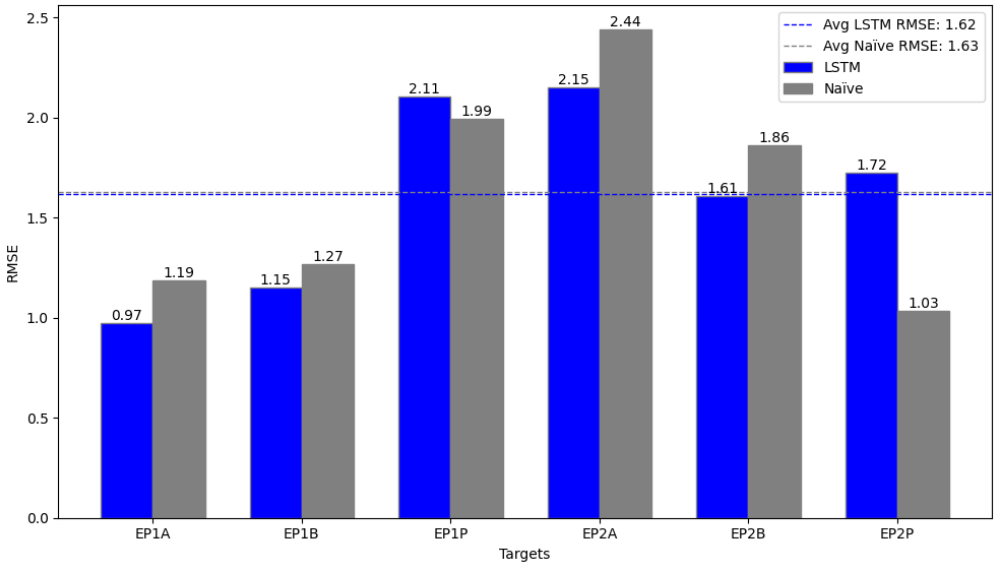


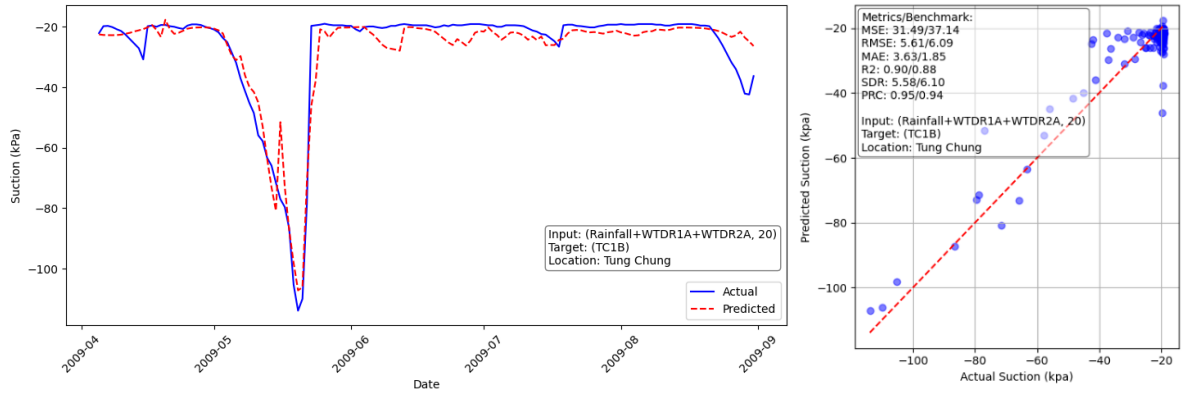

*Figure 5.37. Comparative analysis between the predictions of the six locations at the Tung Chung and Tsing Shan case studies.*

### c) Thermal conductivity (TC) suction response

Figures 5.38.a, b show the prediction results of the LSTM model for a thermal conductivity suction sensor TC1B. The predictions indicate that the DL model can learn from the spatially related features of the soil moisture and provide accurate predictions of R2 of 0.90, MAE of 3.63, RMSE of 5.61, and PRC of 0.95. The same finding is concluded for the three arrays TC1A, TC1B, and TC1C. To check the model's applicability of the model, the model is examined on an extended data set. In this extended dataset, three factors were evaluated: the first is to explore the generability of the LSTM with another data set, the second is to investigate how a more extensive data set can improve the model performance, and finally, to compare the LSTM predictions with the well-known Fredlund-Xing.

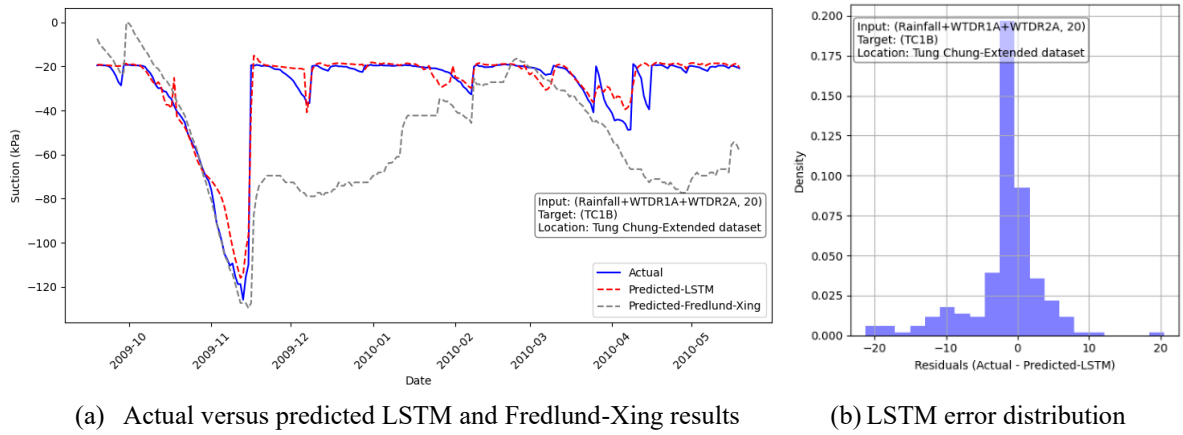
Figure 5.39 compares the LSTM predictions and the well-known Fredlund-Xing equation. The LSTM model outperforms the Fredlund-Xing equation, which illustrates how DL has high learning capabilities as visualized by the error distribution plot in Figure 5.39.b, d. To illustrate, Figure 5.39.b shows the best-fitted SWCC for the TC1B using the VWC of WTDR1A. This figure shows that the Fredlund-Xing equation cannot account for surface-

runoff complex response when considering spatially varied VWC response, where numerous points show the same suction values with different VWC. In contrast, LSTM can learn this transient response from rainfall as a primary trigger and overcome uncertainties by integrating the VWC of the spatially nearby sensors.

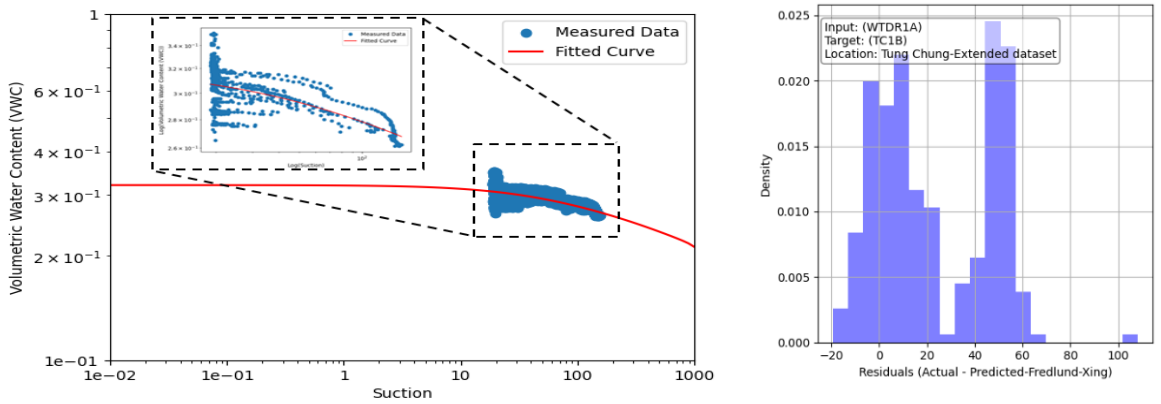



(a) Actual versus predicted LSTM results (b) Scatter plot

*Figure 5.38. Prediction analysis of the TC1B at Tung Chung.*




(a) Actual versus predicted LSTM and Fredlund-Xing results (b) LSTM error distribution

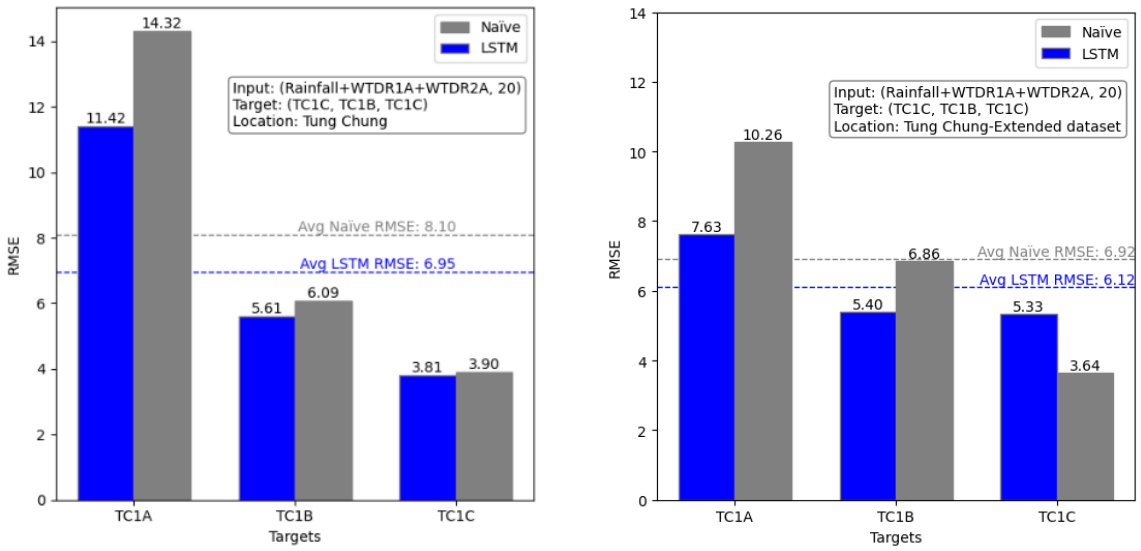



(c) Best-fitted soil water characteristic curve for the TC1B (d) Fredlund-Xing error distribution

*Figure 5.39. Prediction analysis of the TC1B at the Tung Chung extended case study.*

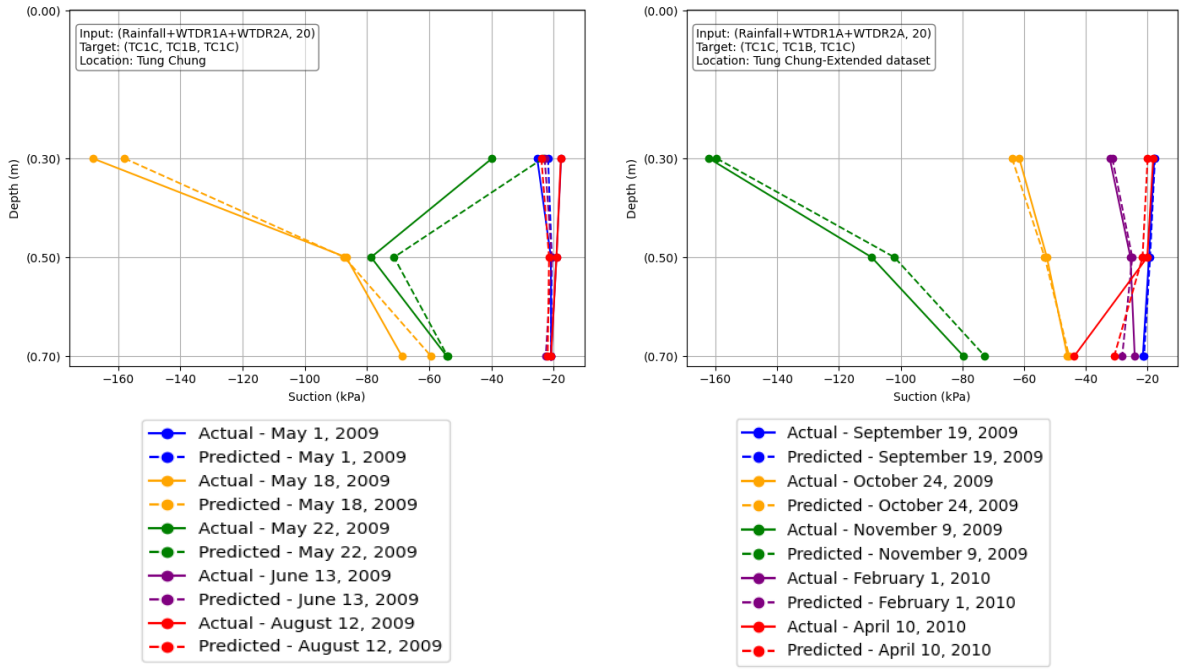
Figure 5.40 compares the overall performance between the three arrays in terms of RMSE for two different lengths of the dataset. The average RMSE of the LSTM for the three arrays is lower than that of the Naïve model. The performance of the deeper sensors is better than that of shallower sensors. This behavior is because deeper sensors are less sensitive to surface-runoff response than the shallower locations. The extended dataset proves the model's generability to predict suction; additionally, it helps the LSTM better understand the spatiotemporal dependencies among the features, providing improved performance by a ratio of 12%. Figure 5.41.a shows the predictions of suction at different depths for two different triggering conditions (i.e., dry and wet seasons) in the two datasets. Minor errors were found in predicting the exact values in the dry periods. During the wet seasons on May 1, June 13, and August 12, the LSTM accurately captures the response from the three locations, highlighting the capability of DL to predict spatially varied responses.

The LSTM accurately captured the matric suction's transient response (wetting front process dynamics), accounting for the temporal and spatial dependences. Zhang et al. (2011) illustrated the typical responses of the negative pore pressure to the change in rainfall conditions. The first response is the sharp wetting front at the three arrays, as visualized in Figure 5.41.a on May 1, June 13, and August 12. The second response is the ongoing saturation process (surface flux is zero); on May 22, TC1A showed lower suction as a result of the start of the infiltration process; at TC1B, the effect of the infiltration has not started yet, recording higher suction values, while at TC1C, the suction values were low as a result of the nearby groundwater level. Thirdly, as predicted, the hydrostatic pressure profile (full dry) as of May 18 shows the same trend as a well-known profile for dry soil (Rahardjo et al., 1995). The same finding is found with the extended dataset in Figure 5.41.b, highlighting the generability of the LSTM model in predicting the suction with outstanding accuracy.




(a) 549 daily time steps (b) 810 daily time steps

*Figure 5.40. Prediction results of TC at Tung Chung with two different dataset lengths.*




(a) 549 daily time steps (b) 810 daily time steps

*Figure 5.41. TC1 comparative analysis at different depths and different triggering conditions.*

### d) Jet Fill Tensometers (JFT) response

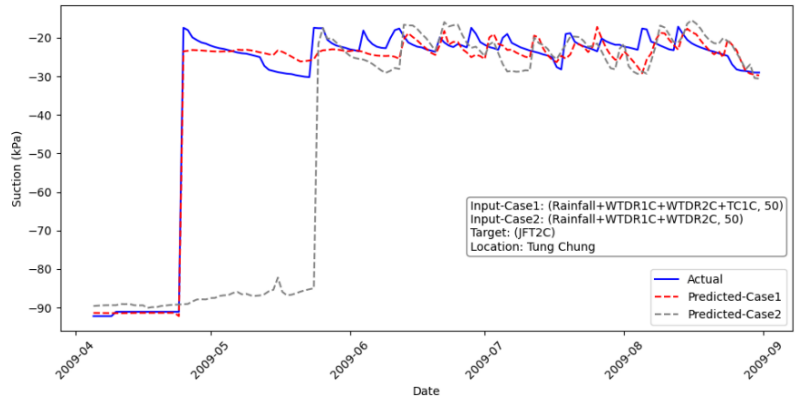
Figure 5.42 shows the prediction results for the two JFT sensors at Tung Chung and compares model performance among different sensors, case studies, and prediction approaches. Figure 5.42.a shows the LSTM prediction for two input features: the first integrated VWC and TC responses, while the second considers only the VWC. Unlike the shallower sensors at TC,

which show no delayed response between the suction and the corresponding rainfall and VWC, the deeper sensors at a depth of 2.50 show a temporal lag between the infiltration and the suction. As a result, VWC and rainfall cannot capture such temporal dependences, necessitating considering spatially nearby responses to let the LSTM learn such transient interaction. The TC values at a depth of 0.70 m at different locations to the JFT2C help the LSTM model learn better. The temporal sequence at deeper sensors, like the JFT2C case, was 50 days, while it was around 20 days for shallower sensors, such as the case of TC1A. The LSTM shows accurate predictions with RMSE of 3.18 kPa, MAE of 2.35 kPa, and R2 of 0.98 (Figure 5.42.b). The same accuracy is obtained in the Tsing Shan case study as visualized in Figures 5.42.c, d. The temporal dependences are accurately captured by integrating the shallower VWC, which gives an RMSE of 4.22 kPa, MAE of 3.00 kPa, and R2 of 0.89. The performance of other sensors is visualized in Figure 5.43 in terms of RMSE. The LSTM provides an accurate prediction compared to the naïve model by a ratio of 2.3% in the Tung Chung case study and 8.7% in the Tsing Shing case study. Figure 5.44 visualizes the capability of the LSTM to accurately capture the transient response (wetting front dynamics), similar to the TC1 sensor.

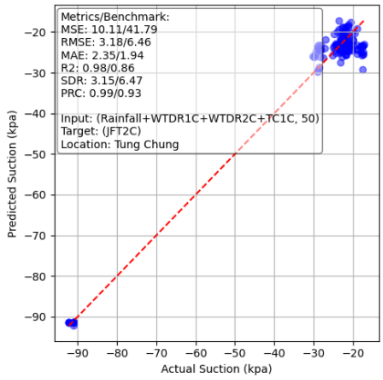
e) VWC response

In this section, additional features were investigated, such as the capabilities of the LSTM model to learn from suction, earth pressure, and groundwater level, predicting VWC. Figures 5.45.a, b visualize the time series predictions versus the actual validation set and the scatter plot between actual and predicted values, respectively. The best prediction metrics were 3.27%, 1.37%, and 0.89 for the RMSE, MAE, and R2. The features were selected at the same depth, while spatially different. It is found that these features provide reasonable accuracy except for some minor errors. Figure 5.45.c compares different model performances regarding RMSE. It is found that all features can account for the spatiotemporal uncertainties of rainfall. The best features are the spatially nearby VWC, as proved in Section 5.2.2. Suction with VWC has no

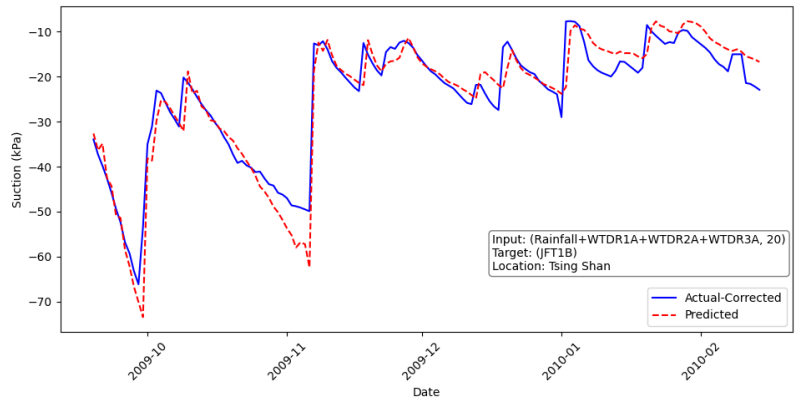
visible influence on prediction accuracy over VWC. The GWL variation, without considering VWC-related features in a location near the target, provides comparable predictions to those of case 2. The mechanical response of the slope in terms of the earth pressure and the hydrological response in terms of suction gives reasonable accuracy in the case of the unavailability of the other best-related features.



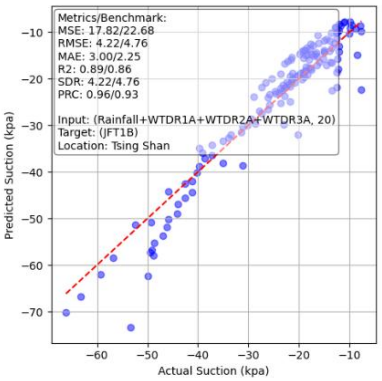

(a) Actual versus predicted JFT2C LSTM results




(b) Scatter plot of JFT2C




(c) Actual versus predicted JFT1B LSTM results




(d) Scatter plot of JFT1B

*Figure 5.42. Prediction analysis of the JFT: a, b) at Tung Chung; c, d) at Tsing Shan.*

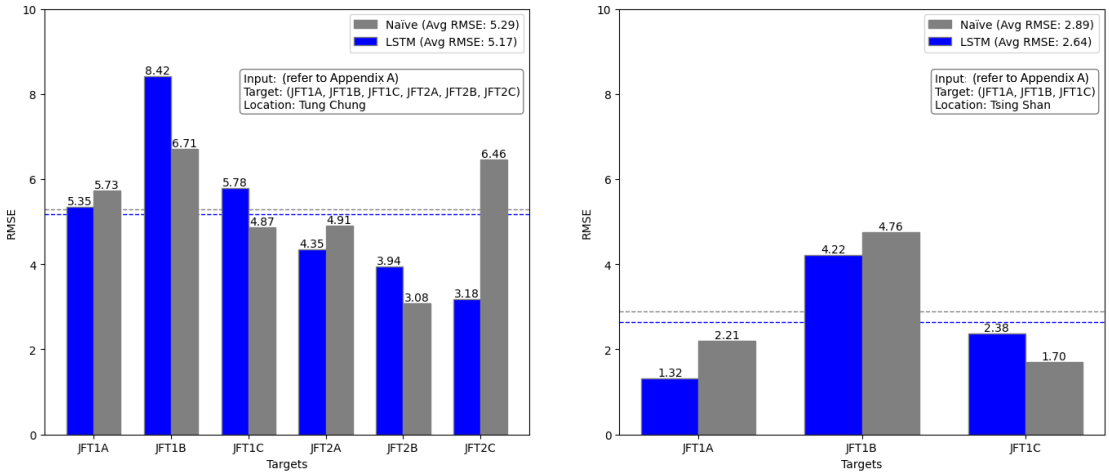
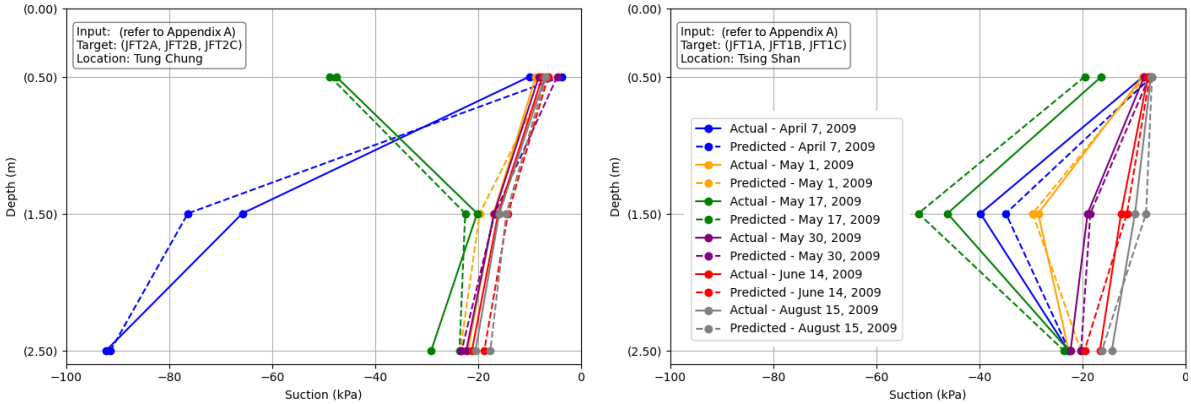



*Figure 5.43. Performance comparison of various sensors at two different case studies.*



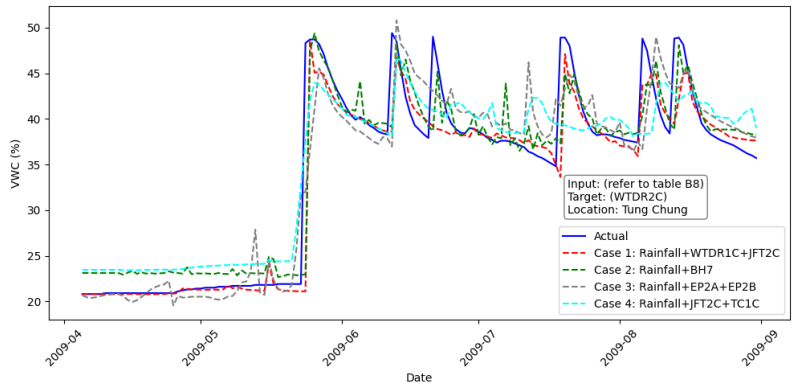

(a) Tung Chung (b) Tsing Shan

*Figure 5.44. JFT comparative analysis at different depths, triggering conditions, and two cases.*

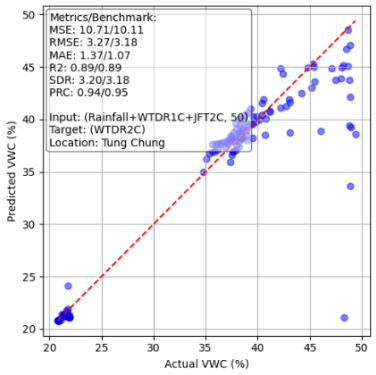
## f) Discussion of results

DL framework challenges: Predicting rainfall-induced landslides is fundamentally challenging due to the intricate spatiotemporal dependencies governing subsurface hydrological and mechanical responses. Traditional models relying on simplified proxies like cumulative rainfall often fail to capture these dynamic mechanisms (Melillo et al., 2018). This study introduces a tailored DL framework, centered on LSTM networks, to directly learn these complex dependencies from transient subsurface data (e.g., earth pressure, suction, and VWC). A critical component of the methodology was the systematic tuning of hyperparameters, particularly

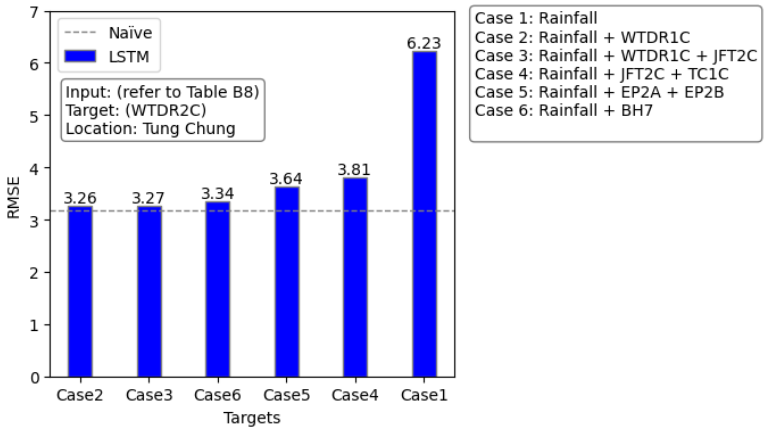
window size and feature sets (refer to Tables A9 to A15), to effectively capture antecedent conditions, a known challenge in hydrological modeling (Sit et al., 2020).




(a) Actual versus predicted LSTM results




(b) Scatter plot of WTDR2C




(c) Comparative analysis between different input features

*Figure 5.45. Prediction analysis of the WTDR2C at Tung Chung.*

Methodological implementation and validation: The modeling process employed a structured optimization strategy to ensure robustness. Hyperparameters were tuned using established regularization techniques like dropout (Srivastava et al., 2014) and early stopping to reduce overfitting, and a sensitivity analysis of window size was conducted to capture temporal dependencies for each target. Feature selection combined statistical methods with domain expertise to ensure the inclusion of key drivers such as rainfall, while accounting for unmonitored influences (e.g., surface runoff, transpiration) by incorporating neighboring subsurface hydromechanical responses. Training used a strict holdout validation strategy with

sliding windows to preserve temporal order. Candidate configurations from coarse searches were retrained for fine-tuning, and final evaluation was performed exclusively on the saved optimal weights without changing model topology, a practice critical for valid sensitivity analysis (Ghaderi et al., 2022). The loss curves showed stable convergence, with validation loss closely matching training loss, indicating limited overfitting. Model performance was benchmarked against a naïve persistence baseline. The study reports results from 38 independently trained models, each with distinct input configurations (Tables A9–A15); sensitivity analyses and model selection were conducted by comparing these fully fine-tuned models on the held-out validation set.

For these models (i.e., 38 models in Tables A9-A15), a dedicated sensitivity analysis was performed to evaluate input feature importance, a complex task for multi-dimensional time series where standard measures like Pearson correlation can under-represent antecedent effects (Sit et al., 2020). The analysis integrated DL attribution with time-resolved sensitivity, tuning for both antecedent window size and model architecture. Crucially, this analysis was applied to the saved optimal weights of each model (Ghaderi et al., 2022), ensuring the resulting feature sensitivities reflect the models' actual learned mapping rather than initial assumptions. While permutation importance offers one valid approach (Molnar, 2020), it is emphasized that visual comparison of time-resolved outputs is essential for accurate judgment, as small performance differences can be misleading without inspecting temporal fidelity.

Key findings, model performance, and scientific insight: This study demonstrates the significant potential of DL for interpreting complex subsurface landslide processes. A strong capability to capture spatiotemporal dependencies between hydrological and mechanical responses was demonstrated, with high predictive accuracy achieved for key variables including matric suction and earth pressure. A key finding is that the concept of a single "maximally activating" input is inherently multivariate and spatiotemporally constrained.

While rainfall is the essential trigger, its capacity to produce a critical subsurface response is entirely governed by the antecedent state of the slope, which is encoded in the historical sequences of spatially distributed sensors. The constraints are both temporal (requiring sufficient sequence length to capture initial saturation) and spatial (requiring input from multiple locations to define internal boundary conditions). This explains why models using only rainfall were significantly outperformed; the "maximizing" condition is not a rainfall total, but a specific interaction between that rainfall and a vulnerable slope state, consistent with geotechnical principles (Lu & Godt, 2013).

This multi-feature approach is a key strength of the proposed DL framework. By integrating a multitude of input features from surrounding sensor locations, the model does not rely solely on rainfall data, which alone exhibits a weak correlation with slope failure. Instead, it learns from the integrated response of the entire slope system. The surrounding hydrological and mechanical features effectively serve as dynamic boundary conditions, allowing the model to discern why similar rainfall patterns can produce drastically different subsurface responses. This provides a more robust and physically grounded prediction than models based on a single trigger. Spatiotemporal dynamics of subsurface systems were successfully captured by the proposed framework. The LSTM models demonstrated high predictive accuracy for key variables, significantly outperforming naive persistence baselines. Key mechanistic insights include: a) Spatial Integration: The model learned that a single “maximally activating” condition is not just rainfall, but the interaction between rainfall and the slope’s antecedent state. Integrating features from multiple sensor locations allowed the model to use surrounding hydrological and mechanical responses as dynamic boundary conditions, dramatically improving prediction. For example, using nearby subsurface responses improved VWC prediction RMSE from 6.23% to 3.26%; b) Temporal Understanding: The LSTM accurately captured temporal lags at different depths. Shallow sensors (0.30-0.70 m) showed immediate

responses to rainfall, while deeper sensors (1.50-2.50 m) exhibited delays, which the model learned by considering shallow responses from nearby locations; and c) Transportability: The model's robustness was validated across two distinct case studies (Tung Chung and Tsing Shan) and datasets of 549 and 810 daily steps, proving its generalizability.

The critical role of subsurface monitoring: This core finding addresses a fundamental preference in the risk management community for remote sensing techniques, which, while excellent for surface deformation monitoring (Wasowski & Bovenga, 2014), cannot directly observe the internal mechanics preceding a failure. Subsurface systems face challenges like energy consumption and harsh environments (Mucchi et al., 2018; Kumar et al., 2020), often leading decision-makers to favor surface-based methods. However, this study underscores the indispensable value of in-situ subsurface sensors. They are not only more cost-effective but are also uniquely capable of capturing the internal mechanisms that precede slope failure. By demonstrating that DL can effectively "see inside the slope" and model its complex inner workings, this work provides a compelling justification for the continued value and integration of subsurface in-situ monitoring networks.

## 5.4. SLOPE STABILITY ANALYSIS AND AUTOMATION FRAMEWORK (TIER 3)

### *5.4.1. Automated Parametrization Model (APHs)*

This research introduces a powerful tool, the Automatic Parametrization of HS Model (APHs), designed to make the HS model parameterization process easier and faster than conventional methods while maintaining high accuracy. Traditional parameterizations rely on oedometer tests, unloading-reloading data, or domain-specific assumptions. Existing optimization-based models often assume uniform parameter weighting, potentially overlooking the distinct sensitivity of each parameter. APHs address these limitations as a standalone tool that relies exclusively on conventional triaxial loading test data. To achieve this goal and address the scarcity of labeled datasets, this study integrates numerical modeling with DL. The study

focuses on a typical shallow Hong Kong soil with parameter ranges derived from field data and relevant literature. Latin Hypercube Sampling generated diverse parameters within theoretical bounds for reliable input, while a 2D axisymmetric finite element model (SIGMA/W) simulated laboratory tests to create a comprehensive labeled dataset.

### a) APHs tuning

In this study, the Bayesian optimization technique (Snoek et al., 2012) was employed to optimize the parameters for each DL model. A maximum of 100 trials were conducted, with 250 training epochs per trial, to initially explore the parameter space effectively. Figure 5.46 presents the density map of each variable against the loss. Some variables exhibit a consistently lower loss for specific models, while others do not show a clear pattern. It is also observed that multiple combinations of hyperparameters can result in similar loss values. Consequently, the optimal parameter set identified through Bayesian optimization was subsequently used to train the models, following a coarse-to-fine tuning approach (refer to Figure A4). Early stopping criteria were applied, and the training and validation processes were closely monitored by visualizing the loss curves to refine parameters iteratively. For instance, this process helped identify and address overfitting or underfitting by adjusting regularization values (i.e., L2) accordingly. The error benchmark for model evaluation was set to an NMAE of less than 1%, with the best-performing model either selected from the tuning process or manually adjusted based on learning curves, being the one that met this criterion. This methodology helps mitigate the risk of being trapped in local minima, and it also helps balance overfitting and underfitting. As a result, computational efficiency was improved, and higher accuracy was achieved. The final optimized parameter sets are presented in Table 5.14.

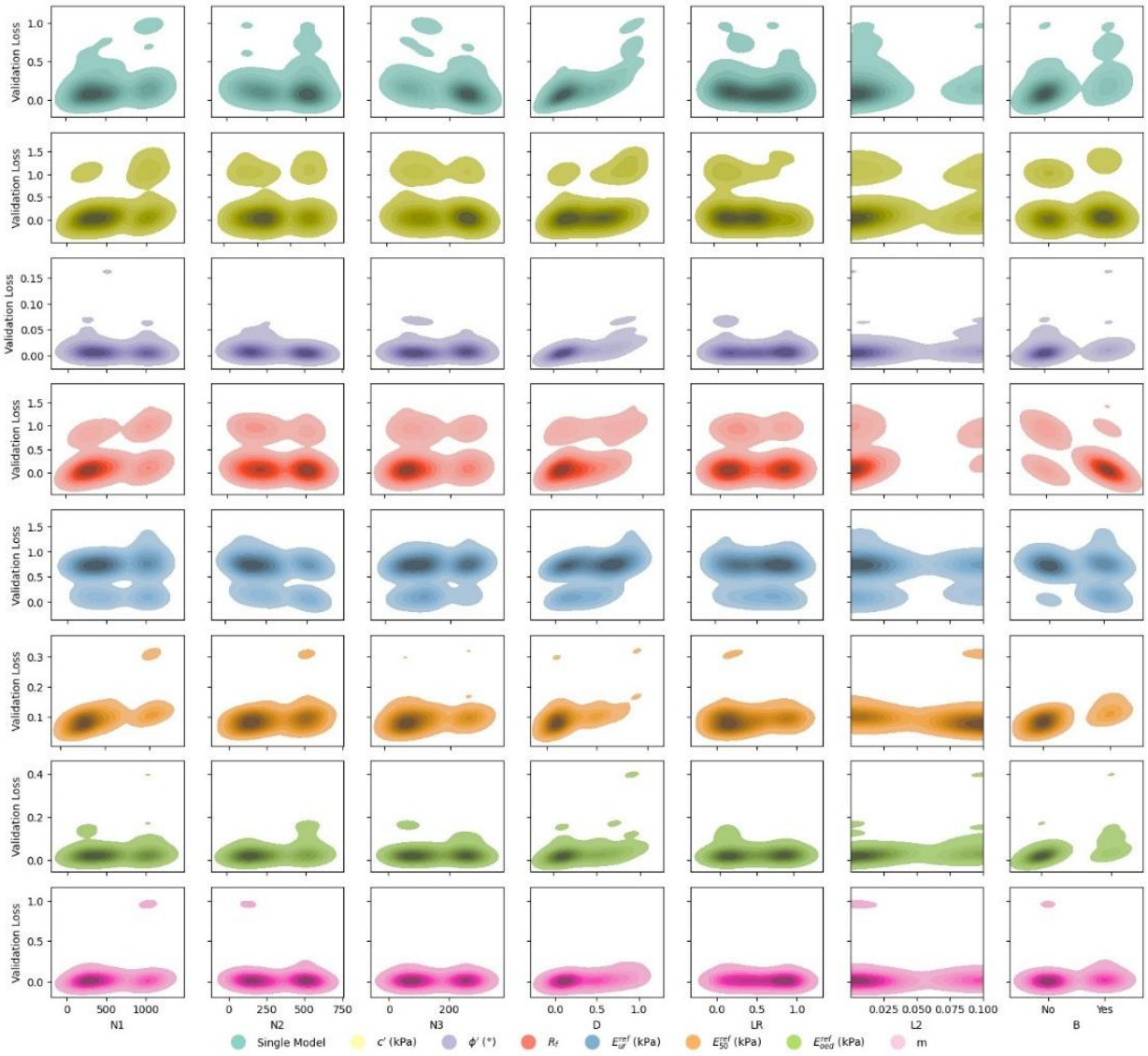



*Figure 5.46. Density maps of hyperparameters vs. validation loss (APHs).*

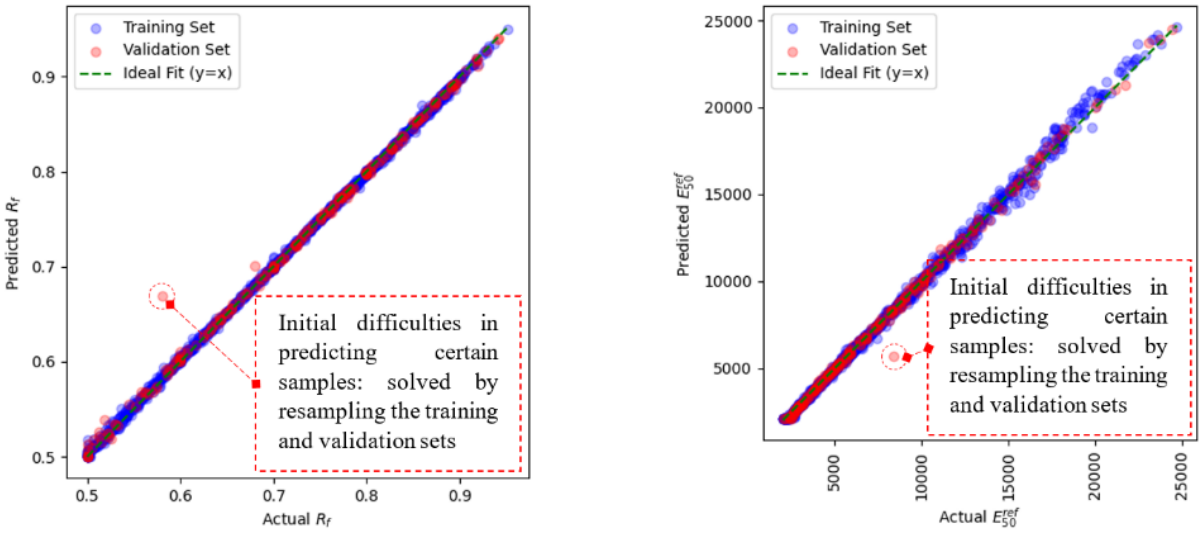
*Table 5.14. The APHs final multi-LSTM model parameters.*

| Parameter | Value (target) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Single LSTM** | Multi LSTM | | | | | | |
| | | ($c'$) | ($\phi'$) | ($R_f$) | ($E_{ur}^{ref}$) | ($E_{50}^{ref}$) | ($E_{oed}^{ref}$) | ($m$) |
| N1 | 512 | 256 | 256 | 512 | 512 | 512 | 256 | 512 |
| N2 | 512 | 256 | 128 | 256 | 256 | 256 | 128 | 256 |
| N3 | 256 | 128 | 64 | 128 | 128 | 128 | 64 | 128 |
| D | 0.1 | 0.1 | 0.1 | 0.3 | 0.3 | 0.3 | 0.1 | 0.1 |
| LR | 0.7 | 0.9 | 0.9 | 0.5 | 0.5 | 0.5 | 0.9 | 0.9 |
| L2 | 0.001 | 0.0001 | 0.0001 | 0.001 | 0.0001 | 0.001 | 0.0001 | 0.0001 |
| B | False | False | False | True | False | True | False | False |
| Initial tuning durations* | 06h 04m 41s | 05h 59m 02s | 07h 14m 46s | 07h 58m 28s | 08h 09m 02s | 06h 20m 15s | 06h 31m 50s | 06h 05m 32s |

* These durations are based on the high computational power provided by the HPC server.
** This refers to a single model that predicts all seven parameters simultaneously.

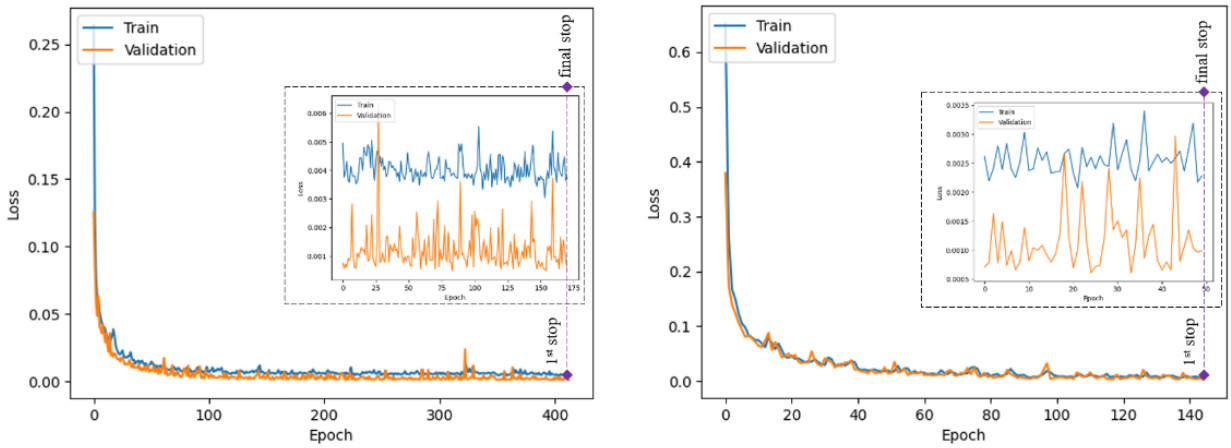
### b) APHs training and evaluation

A hold-out validation strategy was employed to ensure the robustness and reliability of the LSTM models. The dataset, comprising 1200 unique samples with no overlapping responses, was randomized and split into 1000 training and 200 validation samples. The model was trained using early stopping to prevent overfitting, and the optimized weights were saved. This process was repeated with different randomized splits, using the previously saved weights as initialization for further optimization. Given the uniqueness of each sample in the dataset, this strategy allowed the model to generalize effectively by exposing it to diverse training-validation splits while incorporating dropout variability. To ensure model consistency, L2 regularization stabilizes weight updates and reduces sensitivity to data variations (Srivastava et al., 2014). For instance, Figure 5.47 illustrates the training process for some parameters, such as ( $E_{50}^{ref}$ & $R_f$ ), where initial difficulties in predicting specific samples were resolved through resampling. Figure 5.48 presents the learning curves for the optimized LSTM model (for $\phi'$& $m$ as an example). The curves highlight the model's strong performance, as evidenced by the declining loss values, particularly during the initial epochs, followed by stabilization at a consistently low loss level. The close alignment between the training and validation loss curves indicates minimal risk of overfitting. However, minor noise and oscillations are observed during the training process, which can be attributed to the Adam optimization algorithm and random dropout techniques.




(a) $R_f$ (b) $E_{50}^{ref}$

*Figure 5.47. Model evaluation process (1st sampling dataset).*




(a) Internal friction angle ( $\phi'$ ) (b) Power for the stress level dependency ( $m$ )

*Figure 5.48. APHs model training learning curves.*

Figures 5.49 and 5.50 present scatter plots of predicted versus actual values for the strength and stiffness parameters, respectively, demonstrating the robust performance of the LSTM models in predicting the HS parameters. Each subplot evaluates the model's accuracy for a specific parameter, with both training and validation sets clearly delineated (training samples as blue circles and validation samples as red circles). The model achieved exceptional accuracy with an NMAE of 0.66% and 0.74% for $c'$ (Figure 5.49.a). For $\phi'$ (Figure 5.49.b), the NMAE was 0.24% and 0.22%. The model also demonstrated high accuracy in predicting $R_f$ (Figure 5.49.c) with an NMAE of 0.83% and 0.84%. $E_{ur}^{ref}$ (Figure 5.50.a) achieved NMAEs

of 0.9% and 1.23%, while $E_{50}^{ref}$ (Figure 5.50.b) reached 0.70% and 0.83%. $E_{oed}^{ref}$ (Figure 5.50.c) showed NMAEs of 0.39% and 0.42%, and the m (Figure 5.50.d) demonstrated NMAEs of 0.25% and 0.24% for the training and validation sets, respectively. Notably, the proximity of training and validation errors across all parameters highlights the absence of overfitting or underfitting, emphasizing the model's robustness. Figures 5.49 and 5.50 also demonstrate that the validation samples are well-distributed across the parameter space, ensuring comprehensive evaluation and consistent performance for all HS parameters. Table 5.15 presents the evaluation metrics for all models, including the MAE, MAPE, NMAE, and the $R^2$. The models demonstrated exceptional performance, with average NMAE, MAPE, and $R^2$ values of approximately 0.606%, 0.975%, and 0.9987, respectively, underscoring their accuracy and reliability, indicating an accuracy of roughly 99.40%.

In scenarios characterized by complex or noisy data, MC dropout was integrated to quantify predictive uncertainty using an unseen test set of 100 samples. Figures 5.49 and 5.50 illustrate the actual versus predicted values (Magenta), the scatter plots of MAE values (gray), and the corresponding histograms of residuals for each parameter (Blue & Magenta). The analysis shows that residuals exhibit minimal error across parameters, with approximately bell-shaped distributions, except for a slight skew in the stiffness parameters. Notably, the estimated uncertainties ($\sigma$) are minor relative to the range of each parameter. For instance, $\sigma_{E_{ur}^{ref}}$ recorded around 0.76 (ranging from 10 to 50 MPa). Moreover, the model performance is robust, with $R^2$ values exceeding 0.99 and MAE values comparable to those from the deterministic predictions (see Table 5.15). The proximity of the performance could be illustrated by the following: a) extracting axial strain in small increments helps represent each curve with enough points, minimizing the impact of any potentially noisy data; b) adopting a multi-approach that assigns distinct weights for each parameter; c) applying a structured tuning process, from coarse to fine, with integrated dropout and regularization. Given the similar performance of

deterministic and stochastic predictions, using the standard deviations of residuals from training and evaluation offers a time-efficient way to estimate uncertainty without requiring additional sampling or computationally expensive techniques at prediction time.

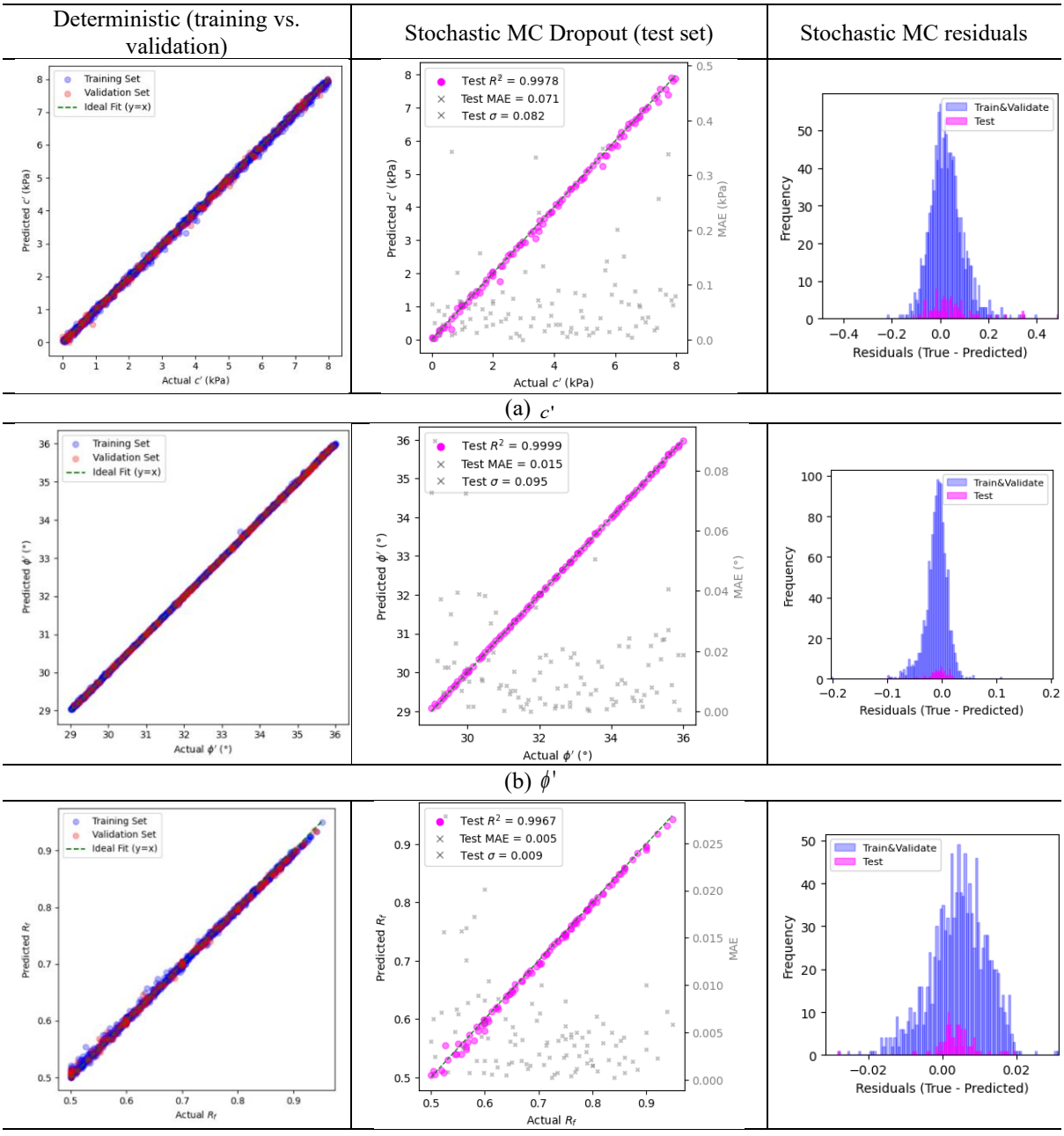



*Figure 5.49. Deterministic (training and validation sets) and stochastic (test set) predicted strength parameters versus actual HS parameters with stochastic test MAE and residuals.*

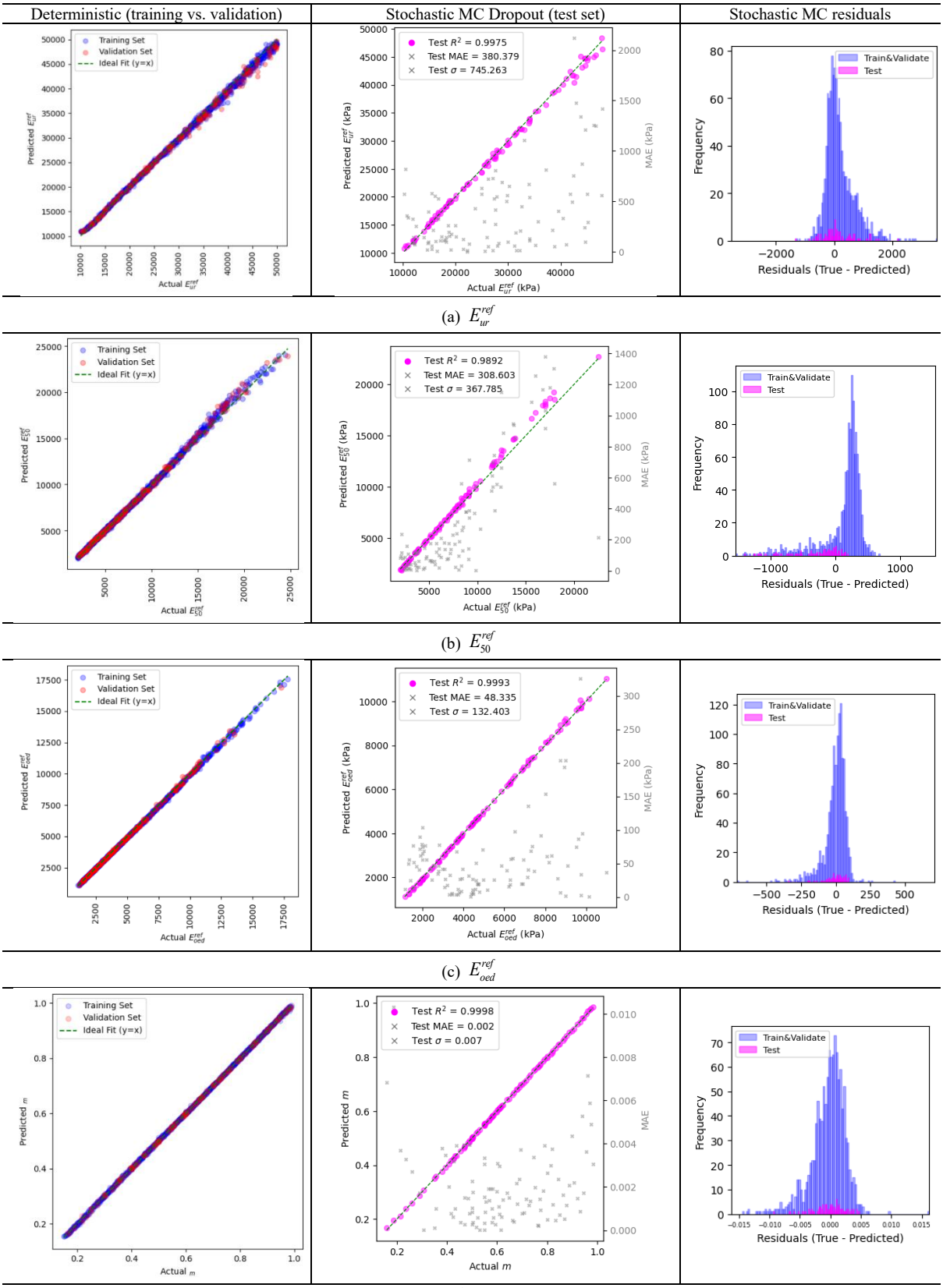



*Figure 5.50. Deterministic (training and validation sets) and stochastic (test set) predicted stiffness parameters versus actual HS parameters with stochastic test MAE and residuals*

*Table 5.15. The APHs model evaluation metrics: NMAE, MAPE, R2, and MAE.*

| Case | Training (T) and Validation (V) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Metrics | NMAE | | MAPE | | $R^2$ | | MAE | |
| Parameters | T | V | T | V | T | V | T | V |
| $c'$ (kPa) | 0.66 | 0.74 | - | - | 0.9991 | 0.9987 | 0.052 | 0.085 |
| $\phi'$(°) | 0.24 | 0.22 | 0.05 | 0.05 | 0.9999 | 0.9999 | 0.017 | 0.015 |
| $R_f$ | 0.83 | 0.84 | 0.56 | 0.55 | 0.9984 | 0.9985 | 0.004 | 0.004 |
| $E_{ur}^{ref}$ (kPa) | 0.9 | 1.23 | 1.28 | 1.59 | 0.9982 | 0.9958 | 361.512 | 490.832 |
| $E_{50}^{ref}$ (kPa) | 0.7 | 0.83 | 2.09 | 2.18 | 0.9976 | 0.9969 | 156.208 | 188.001 |
| $E_{oed}^{ref}$ (kPa) | 0.39 | 0.42 | 1.25 | 1.27 | 0.9994 | 0.9991 | 65.006 | 67.919 |
| $m$ | 0.25 | 0.24 | 0.43 | 0.4 | 0.9998 | 0.9999 | 0.002 | 0.002 |
| Average | 0.567 | 0.646 | 0.943 | 1.007 | 0.9989 | 0.9984 | - | - |

### c) APHs testing, validation, and deployment

After the training phase, the model was tested using unseen data. The first test was numerically sample parameterized by GEOSLOPE International Ltd. (2021), representing Ottawa sand (Krachtoudi, 2001). This test served two purposes: to evaluate the model's ability to parameterize the HS model for unseen data accurately and to validate the numerical model used to generate the training dataset. Since this soil sample falls within the parameter range used during model training, it provided a suitable benchmark for comparison. Table 5.16 presents the HS model parameters calculated by GEOSLOPE International Ltd. (2021) alongside the parameters retrieved using the APHs tool developed in this study. The comparison demonstrates that the APHs tool achieved an accuracy of approximately 99.6%, with minimal absolute error for each parameter. Furthermore, Figure 5.51 visualizes the simulated results obtained using parameters from both the APHs tool and the conventional parameterization method. The results confirm the precision of the APHs tool, validating its reliability and the accuracy achieved during training and validation when applied to unseen data.

*Table 5.16. APHs vs conventional parameterization (GEOSLOPE International Ltd., 2021).*

| Parameter | Value | | | | | | |
|---|---|---|---|---|---|---|---|
| | ($c'$) (kPa) | ($\phi'$) (deg) | ($R_f$) | ($E_{ur}^{ref}$) (kPa) | ($E_{50}^{ref}$) (kPa) | ($E_{oed}^{ref}$) (kPa) | ($m$) |
| GEOSLOPE | 0 | 29.6 | 0.941 | 45000 | 17745 | 11500 | 0.68 |
| APHs | 0.041 | 29.589 | 0.9359 | 45361.88 | 17849.41 | 11528.095 | 0.682 |
| MAE | 0.041 | 0.011 | 0.0051 | 361.88 | 104.41 | 28.095 | 0.002 |
| MAPE (%) | - | 0.037 | 0.54 | 0.80 | 0.588 | 0.24 | 0.294 |

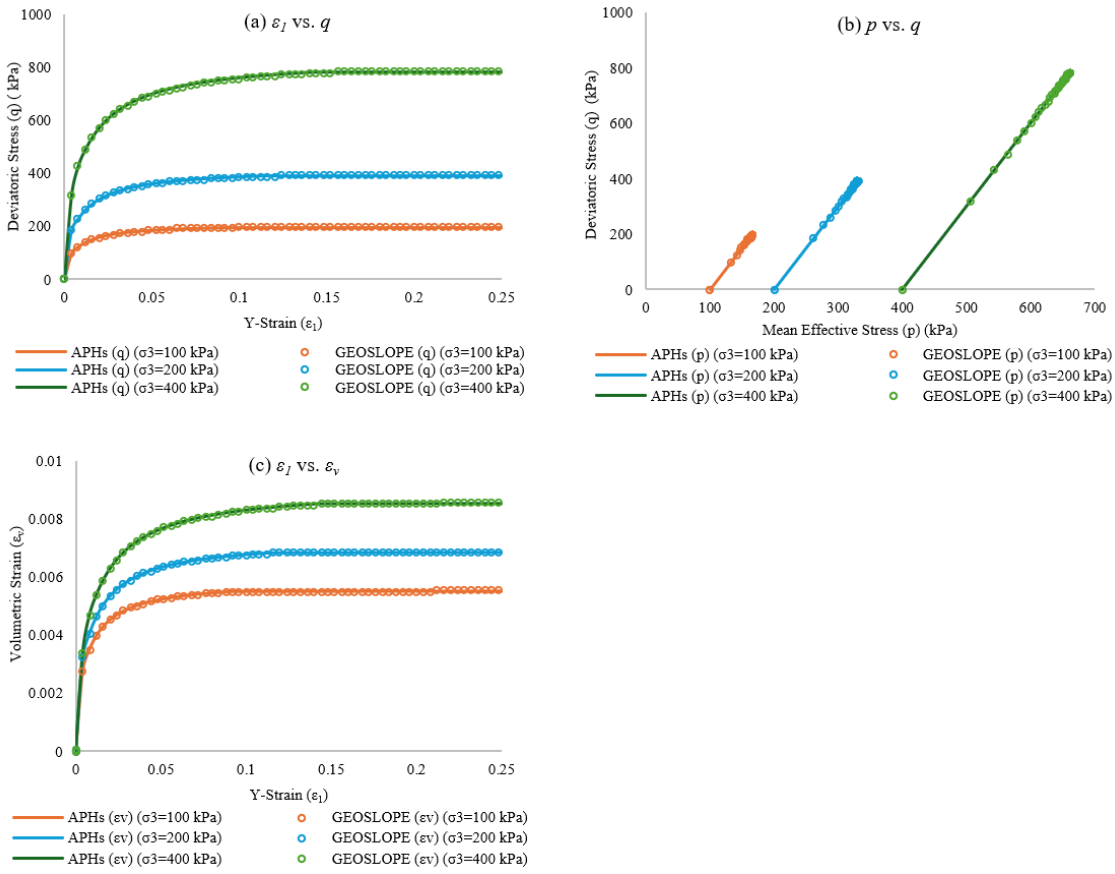



*Figure 5.51. Comparison of simulated results: APHs vs. conventional parameterization (GEOSLOPE International Ltd., 2021).*

The APHs tool was further applied to another real-world laboratory case study to validate its reliability. Zhou (2008) conducted conventional consolidated drained triaxial tests on recompacted saturated CDG soil. These tests aimed to determine its shear strength under $\sigma_3$ of 100 kPa, 200 kPa, and 400 kPa as part of a soil nailing research project. The laboratory data, along with $\upsilon_{ur}$ of 0.3 (Zhou, 2008), were used as inputs for the APHs tool. The HS parameters extracted by the tool were then simulated using an axisymmetric finite element model (SIGMA/W) (GeoStudio, 2023c), and the simulation results were compared against the original laboratory data for further validation, as shown in Figure 5.52. From the laboratory tests at 15% and 20% strain, the calculated shear strength parameters were $c' = 0$ and $\phi' =$ 35° and 35.1°, respectively. The APHs tool predicted $c' = 0.49$ kPa and an $\phi'$ of 34.96°, showing minimal differences. These minor differences did not significantly influence the

simulation results, as demonstrated in Figure 5.52. The differences in $\phi'$ align with the uncertainties shown in Figures 5.49 and 5.50 and the confidence intervals (CIs) in Table 5.17. The discrepancies in $c'$ may stem from inherent model uncertainties or from measurement noise. Measurement noise could also have been introduced during the digitization of laboratory data using the WebPlotDigitizer tool (Rohatgi, 2022). To illustrate, the original data from Zhou (2008) were converted from Portable Document Format (PDF) to numerical format for analysis, which might have contributed to these discrepancies. The extracted parameters and their corresponding 99% CI are presented in Table 5.17, providing users with a prediction range to account for any possible deviations. The CIs are calculated using the standard deviations of residuals during the training and evaluation process. The results in Table 5.17 align with the MC dropout uncertainty estimates, further validating the rapid interface predictions.

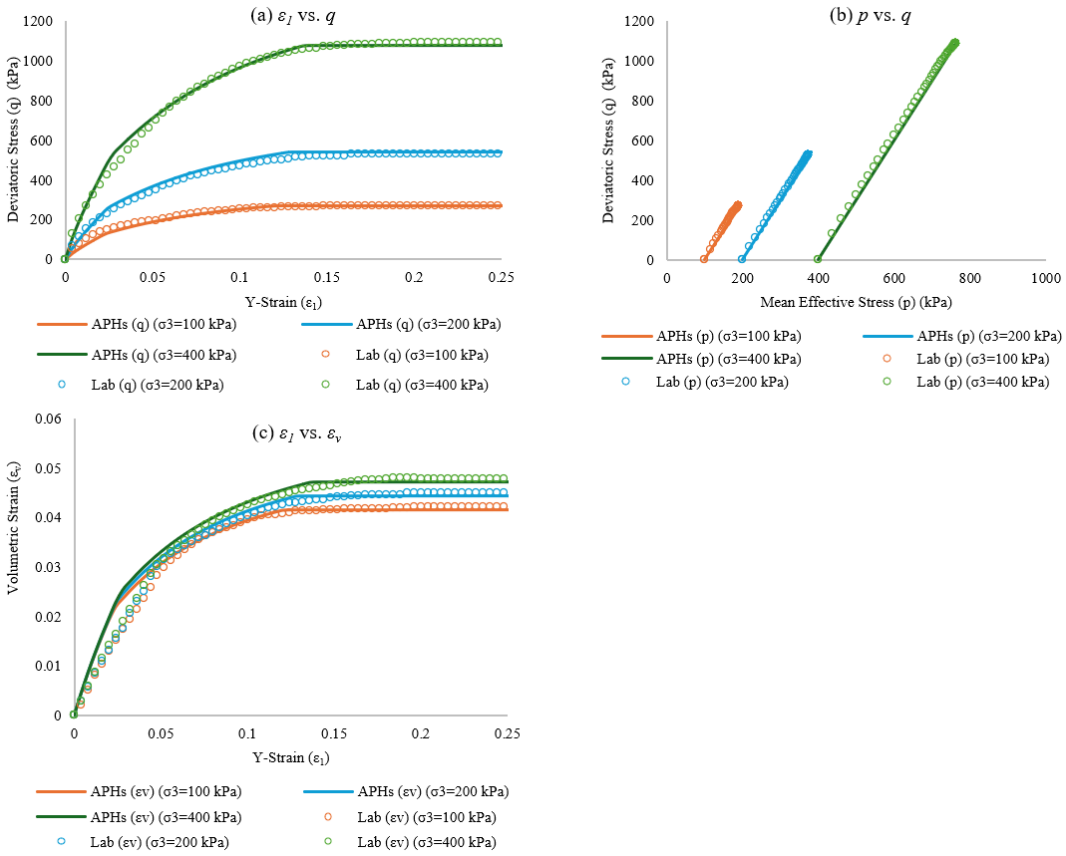



*Figure 5.52. Comparison of simulated results: APHs vs. laboratory test results (Zhou, 2008).*

*Table 5.17. HS model parameters with 99% CI lower and upper bounds using the APHs tool.*

| Parameter | Value | | | | | | |
|---|---|---|---|---|---|---|---|
| | ($c'$) (kPa) | ($\phi'$) (deg) | ($R_f$) | ($E_{ur}^{ref}$) (kPa) | ($E_{50}^{ref}$) (kPa) | ($E_{oed}^{ref}$) (kPa) | ($m$) |
| APHs | 0.493 | 34.960 | 0.715 | 12145.493 | 6041.435 | 1807.876 | 0.908 |
| 99% CI lower | 0.306 | 34.907 | 0.702 | 10815.303 | 5326.814 | 1515.862 | 0.902 |
| 99% CI upper | 0.682 | 35.014 | 0.727 | 13475.684 | 6756.055 | 2099.89 | 0.915 |

### *5.4.2. Automated Slope Stability Analysis Model (A Hybrid DL–Physics Model)*

Several deterministic numerical models reported in the literature provide accurate simulations but require extensive geotechnical investigation and are time-consuming. As noted in Section 3.2.2, landslides in Hong Kong are predominantly shallow and occur within the colluvium layer; therefore, the infinite-slope assumption described in Section 3.4.3 is adopted here. The remaining challenge is monitoring the saturation level of shallow infinite slopes. Sensor-based monitoring also requires extensive investigation, whereas data-driven approaches have demonstrated success in spatiotemporal modeling as presented in Section 5.2. Accordingly, in this section, the saturation level is automated using the DL algorithms, with inputs that account for rainfall characteristics and the spatial features that control how much rain infiltrates into the slope. The resulting saturation estimates are combined with the infinite-slope stability model to compute the FOS.

a) Model tuning, feature analysis, and evaluation

Following the procedure illustrated earlier, Table 5.18 shows the final model tuning parameters. Figure 5.53 compares the CNN and the LSTM models, displaying their performance metrics, ROC curves, and overall confusion matrices. Both LSTM and CNN models demonstrate strong predictive capabilities, with CNN showing a slight overall advantage. The CNN achieves higher accuracy (86.15% vs 85.15%) and precision (73.45% vs 69.26%), indicating better overall classification performance and fewer false positives. This is supported by the CNN's confusion matrix showing fewer false negatives (343 vs 440 for class 0) and comparable true

positive detection. The LSTM exhibits marginally better recall (76.61% vs 73.09%), suggesting it is slightly more sensitive in detecting positive cases, though this comes at the cost of lower precision. Both models show similar balanced accuracy (82.37% vs 81.91%), F1 scores (72.73% vs 73.28%), and correlation metrics (MCC and Kappa around 0.63), indicating comparable overall performance. The AUC values reinforce this pattern, with CNN showing slightly better training performance (0.91 vs 0.90) while both maintain strong validation and testing AUCs (0.87-0.88), demonstrating good generalization without overfitting. The consistent high AUC across training, validation, and testing phases for both models indicates robust learning and reliable predictive capability.

*Table 5.18. The optimized SSI model parameters.*

| Parameter | Value (target) | |
|---|---|---|
| | LSTM | CNN |
| N1 | 512 | 512 |
| N2 | 128 | 128 |
| N3 | 64 | 64 |
| D | 0.3 | 0.5 |
| LR | 0.1 | 0.1 |
| L2 | 0.0001 | 0.01 |
| Kernel size | - | 3 |
| Initial tuning durations* | 23h 02m 42s | 11h 38m 09s |

The feature importance analysis (Figure 5.54) validates the fundamental assumption that rainfall serves as the primary triggering mechanism for landslides in the study area, with both LSTM and CNN models assigning overwhelmingly dominant importance to precipitation variables. However, the positive importance scores of spatial features demonstrate that topographic characteristics critically modify how rainfall influences saturation across different landscape configurations: The LSTM model emphasizes slope gradient as the primary modifying factor. Whereas the CNN model shows more balanced weighting across multiple topographic dimensions (Aspect, Slope, Elevation), indicating spatial convolution captures complementary terrain influences.

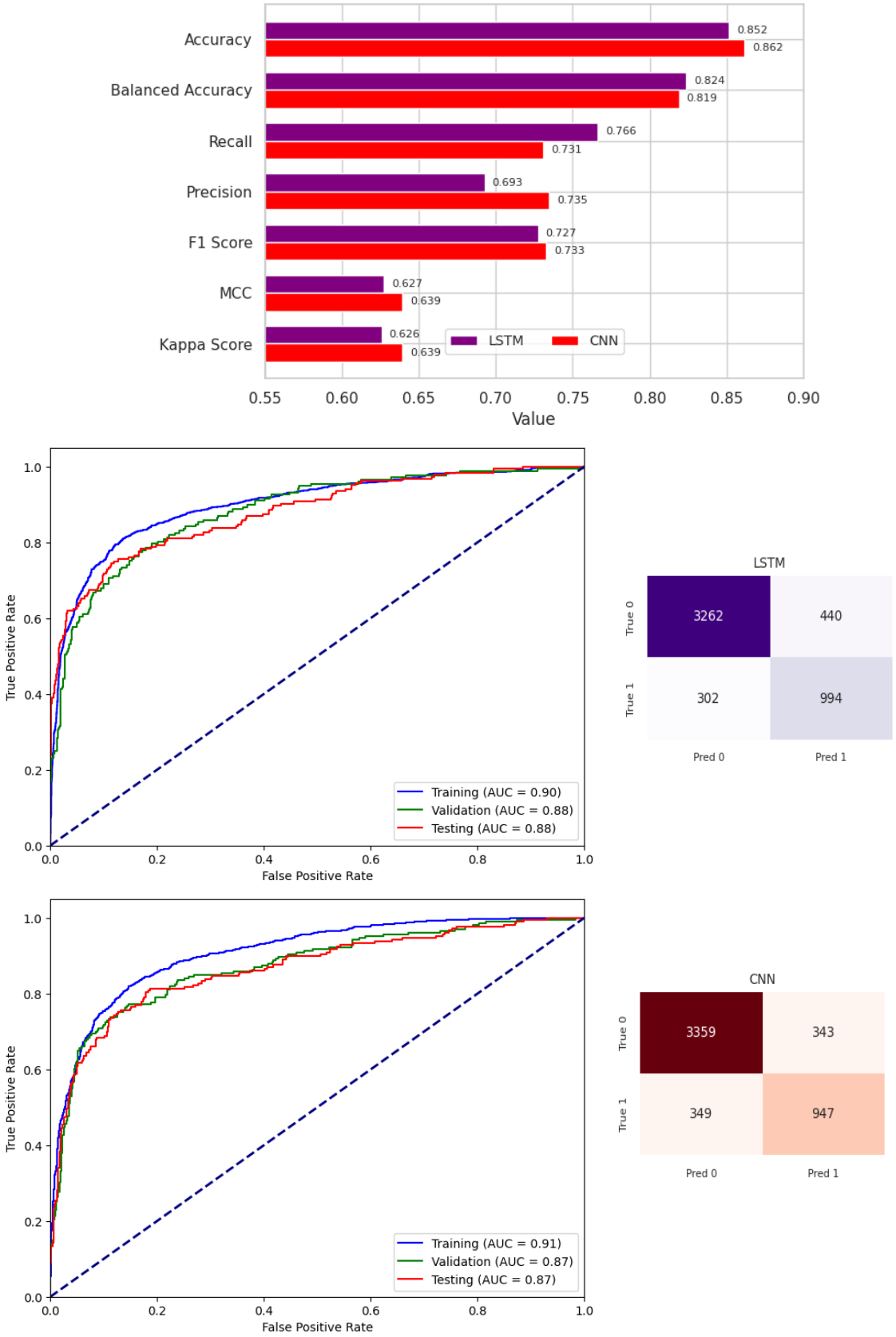
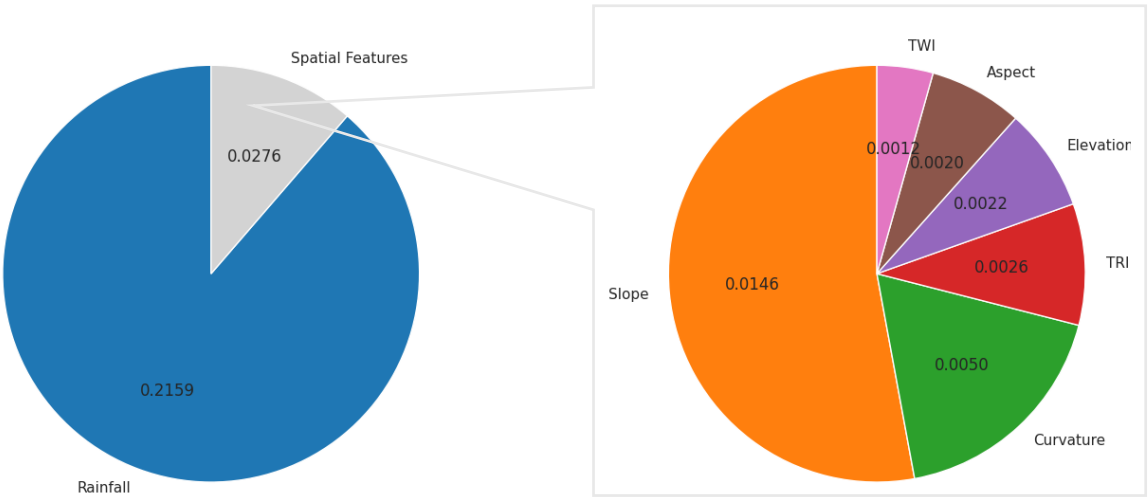
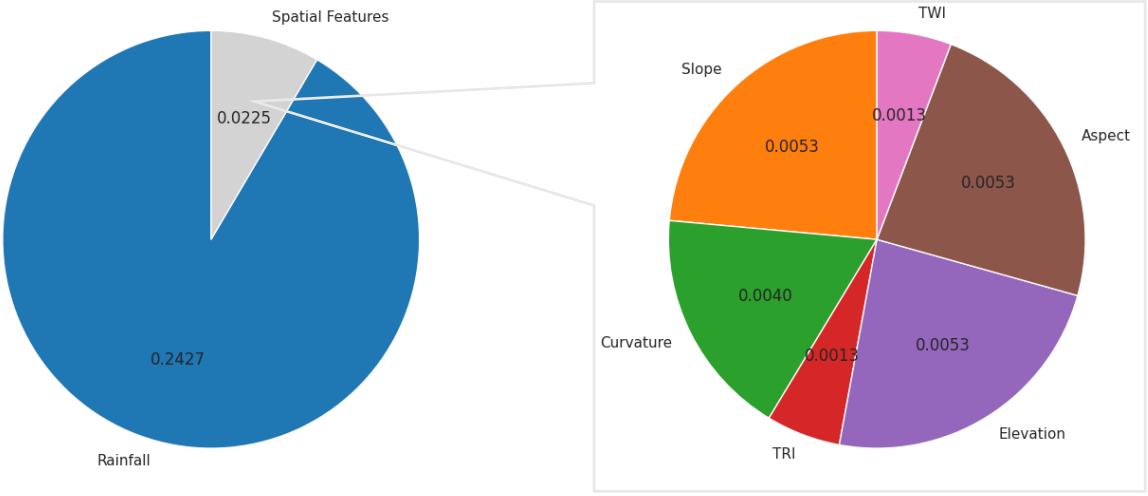



*Figure 5.53. Comparative Performance Evaluation of LSTM and CNN Models.*




a) LSTM



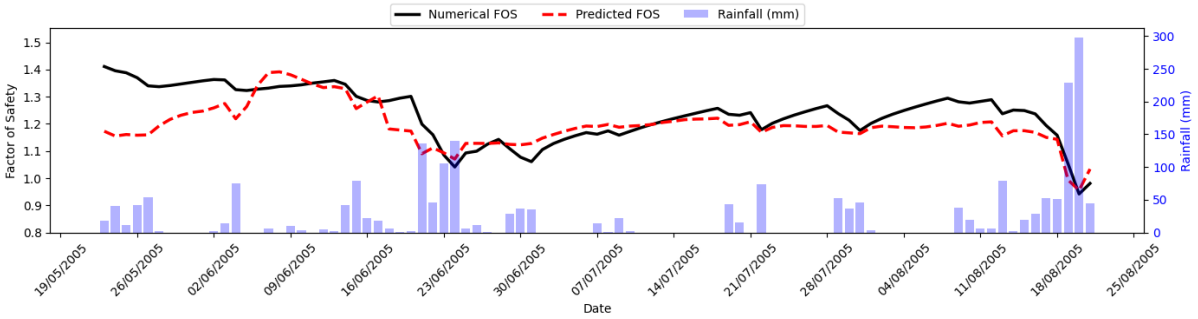

b) CNN

*Figure 5.54. Permutation-based feature importance scores.*

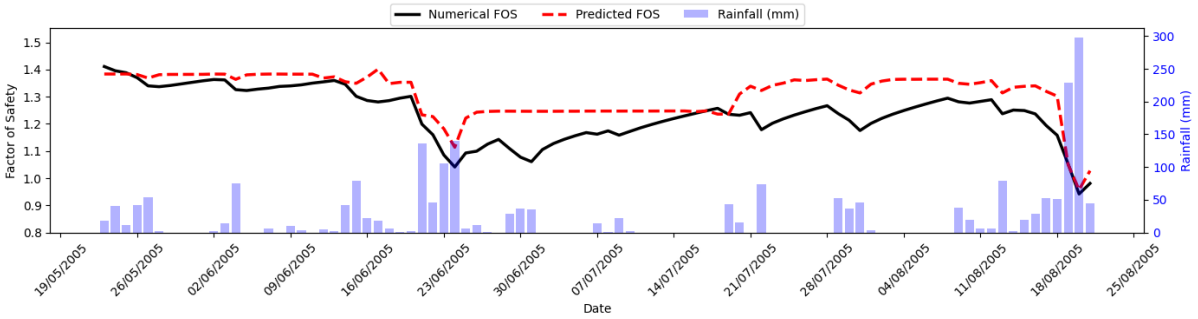
### b) Model validation and testing

The external validation of the developed DL models against an independent case study demonstrates robust predictive performance and physical consistency. The close alignment between data-driven predictions and numerically simulated FOS values, evidenced by low MAE scores (LSTM: 0.06, CNN: 0.07), validates the models' capability to capture complex slope stability dynamics (Figure 5.55). The models successfully reproduced the hydrological response to evolving rainfall patterns, with the integrated spatiotemporal framework proving particularly effective in minimizing false alarms. Notably, during the intensive rainfall period from 21-24 June 2005 (accumulating 426.5 mm over four days), both DL models correctly

maintained stability predictions despite the extreme precipitation, aligning with numerical simulations and field observations. A significant finding concerns the temporal dynamics of slope recovery. The CNN model demonstrated superior capability in capturing the delayed drainage processes observed in the physically-based numerical model, maintaining elevated probability estimates during the post-event recovery phase. This suggests the convolutional architecture's enhanced ability to represent the hysteretic hydrological behavior characteristic of slope stability systems. The Monte Carlo dropout analysis (Figure 5.56) revealed consistent predictive uncertainty, with the LSTM exhibiting marginally better stability (average standard deviation: 0.06 vs 0.07). Both models successfully identified the actual failure event, demonstrating reliable warning capability while maintaining computational efficiency compared to traditional numerical approaches. This validation confirms that the integration of data-driven methodologies with physically based equations produces models that not only achieve statistical accuracy but also maintain geotechnical consistency.




a) LSTM.




b) CNN.

*Figure 5.55. Comparison of predicted vs the numerical FOS for the Fei Ngo Shan Landslide.*

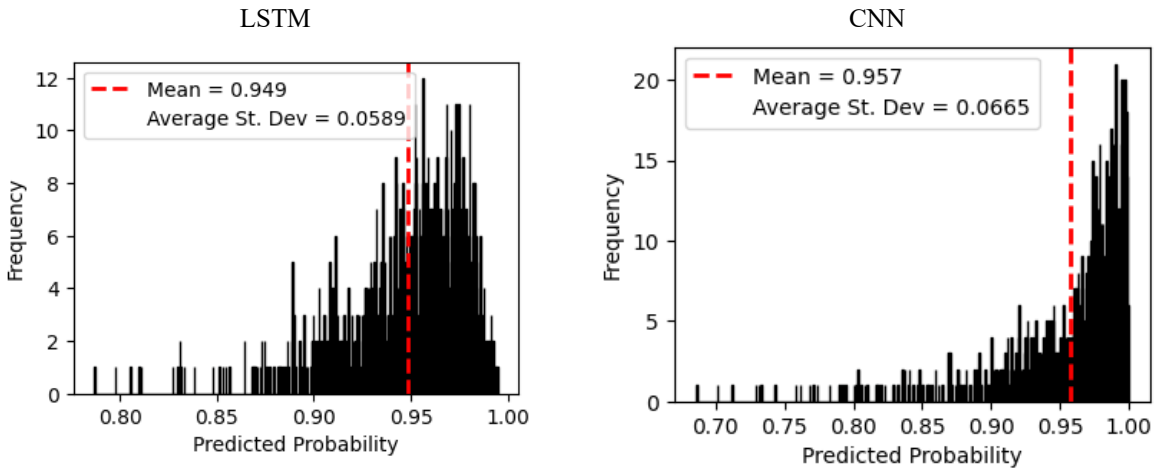



*Figure 5.56. Monte Carlo predictions and mean standard deviations for the August event.*

### *5.4.3. Factors Controlling Rainfall-Induced Landslides: A Parametric Study*

Empirical statistical thresholds rely on the rainfall intensity (I) and rainfall duration (D) using historical records for a given regional area (Caine, 1980; Gariano et al., 2019; Hong et al., 2006; Lee et al., 2014). For example, considering Hong Kong as a case study, if the rainfall in Hong Kong exceeds 70 mm/hr (heavy rainfall), there is a very high probability of a landslide (Vargas, 1971; Brand, 1984). Lee et al. (2014), Huang et al. (2015), Hong et al. (2018), Chhorn et al. (2016), Uwihirwe et al. (2020), Zhao et al. (2019b), and Brunetti et al. (2010) considered the antecedent rainfall to overcome the uncertainties of the empirical thresholds. For a specific amount of time, the cumulative rainfall is regarded as antecedent rainfall; before this time, the rainfall can be disregarded. However, it is still unknown what length of antecedent rainfall is most relevant for landslide prediction (Chhorn et al., 2016). Above all, cumulative rainfall may not be a reliable indicator of soil healing because it ignores the effect of non-rainfall duration, which is thought to influence soil healing behavior. In other words, a small non-rainfall timing might inhibit soil healing and vice versa (Bontemps et al., 2020).

Acknowledging that the physical deterministic models provide higher accuracy compared to empirical thresholds, as they account for the slope hydrological, mechanical, and topographical characteristics. There are numerous examples of physical deterministic research

on the role of rainfall intensity, pattern, duration, and topographic, mechanical, and hydraulic characteristics regarding rainfall-induced landslides (Oliveira et al., 2022; Ng & Shi, 1998; Ng et al., 2001; Rahimi et al., 2011; Bai et al., 2023; Sitarenios et al., 2021; Cheng et al., 2007; Feng et al., 2023; Nguyen et al., 2019). Nevertheless, further understanding of the triggering feature's randomness is required, as rainfall is an entirely random event characterized by intensity, duration, non-rainfall duration, and pattern.

This section aims to examine the factors that influence triggering randomness in rainfall-induced landslides to overcome the limitations stated above. This is accomplished by transient seepage analysis using the physical deterministic analysis approach through the commercial software SEEP/W and SLOPE/W (Geo-Studio, 2023a, b). A case study in Hong Kong validates the numerical model (Section 5.2.2), and then a typical Hong Kong natural slope is developed for the parametric analysis. The effect of rainfall patterns is studied by modeling four patterns: delayed, advanced, normal, and average. This study investigated how rainfall intensities, durations, initial conditions, and soil hydraulic conductivities impact soil healing and recovery behavior. This research offers novel insights for enhancing accuracy and lowering the false positive rates of empirical statistical thresholds. It does this by physically examining the features controlling the healing or recovery time (effective non-rainfall duration). This new characteristic (healing time) may be defined as the shortest time necessary for the slope to return to its earlier state. This investigation provides the cross-correlation between the slope and rain features and the recovery time, highlighting when the antecedent effect should be incorporated to increase the accuracy of empirical thresholds. This work offers novel insights into a recurrent question: When should the antecedent effect be incorporated to increase the accuracy of the empirical threshold?

### a) Assumptions and boundary conditions

Hong Kong's humid climate allows rainfall to be simplified as infiltration and runoff (Morbidelli et al., 2018; Liu et al., 2022). Location, vegetation area, rainfall duration, rainfall volume, slope geometry, and soil qualities all impact drainage. The runoff on natural Hong Kong slopes can be considered to be a constant percentage of the rainfall. According to Premchitt et al. (1992), GCO (1982), and Ng et al. (2001), this ratio might be considered to be 40%. Tropical cyclonic storms or slow-moving low-pressure troughs are the two primary sources of rainfall in Hong Kong (Lumb, 1975). Frequently, there will be days of nonstop rain, followed by another trough, followed by a few days of clear weather. Strong winds, which can even approach typhoon strength, and heavy rainfall spread over several hours are invariably linked to these storms. It is unusual for these rainstorms to produce more than two or three days of nonstop rain. According to Wisner's calculation, the average rainfall intensity in Hong Kong during 36 hours with a return period of 10 years is 25 mm/hr (GEO, 2023).

For the parametric study, the precise rainfall quantity is unimportant; therefore, any minor differences will not affect the results. Different rainfall patterns were investigated for the same commutative rainfall, as presented in Figure 5.57. The rainfall duration of all patterns was 36 hours, followed by a similar non-rainfall duration. As shown in Figure 5.58, the effective non-rainfall duration (healing or recovery time) was examined under the assumption of subsequent average rainfall incidences. The rainfall intensities over the same 36-hour period were 25, 50, and 75 mm/hr. Three distinct time spans of 25, 36, and 48 hours were examined for the same intensity (25 mm/hr). In reference to the soil in Hong Kong, where the saturated hydraulic conductivity varies between $5x10^{-6}$ and $3.5x10^{-4}$ m $s^{-1}$, three different hydraulic conductivities were simulated: $9x10^{-6}$, $1.5x10^{-5}$, and $3x10^{-5}$ (GCO, 1982). Different initial suctions were investigated to emphasize the influence of the antecedent effect; these values were 50, 75, and 100 kPa, as indicated by (Rahardjo et al. 2007).

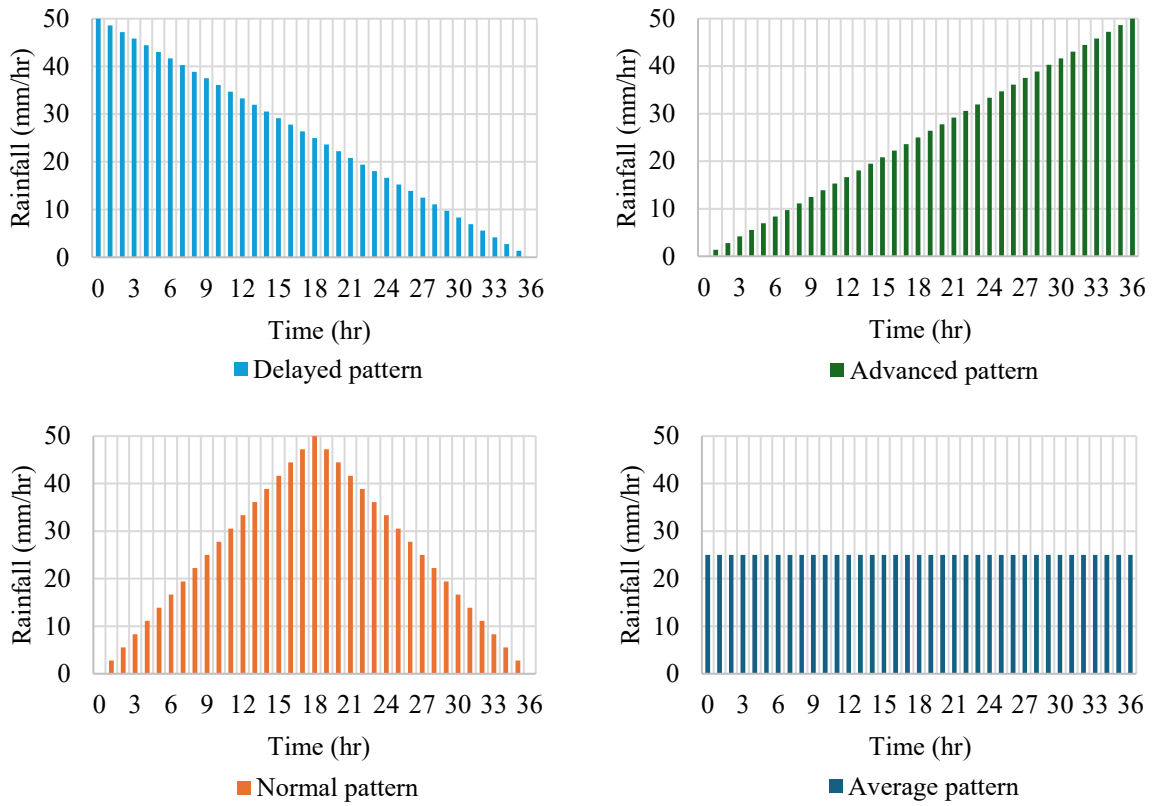



*Figure 5.57. Schematic illustration of different rainfall patterns investigated.*

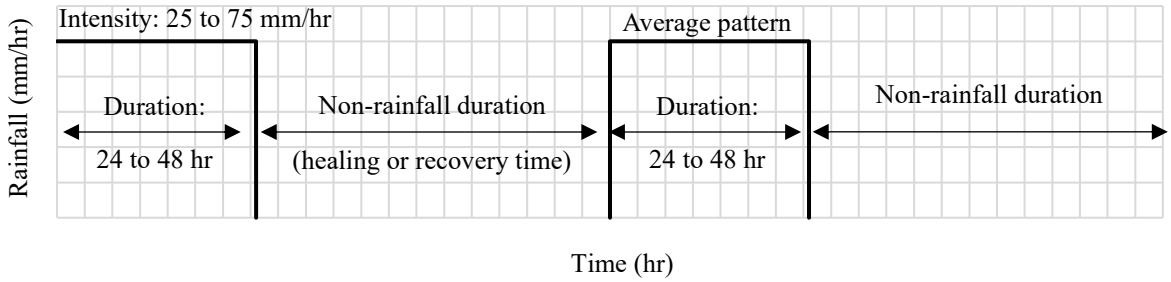



*Figure 5.58. Rainfall parameters studied: intensity, duration, and recovery time.*

The presence of colluvium and completely decomposed granitic, volcanic, and sedimentary rocks, which the USDA classifies as loam or silt loam (GCO 1982, Maunsell Geotechnical Services Ltd., 2007), is a significant problem related to slope instability in Hong Kong. The thickness of this layer ranges from 1 to 36 m (GCO, 1982; Ng & Shi, 1998). The average slope angle ranges from 15 ° to 40 ° (Liu et al., 2022). Most rain-caused landslide failures are generally shallow, with slip surfaces parallel to the slope surface (Das et al., 2022; Saadatkhah et al., 2015; Thang et al., 2022). Because the thickness and slope of each soil layer change depending on location, it was assumed for parametric simulations to idealize the

complicated hillside geology as a single infinite layer with an impermeable rock stratum, as depicted in Figure 5.59. The mechanical and hydrological properties are allocated based on existing literature, as shown in Table 5.19.

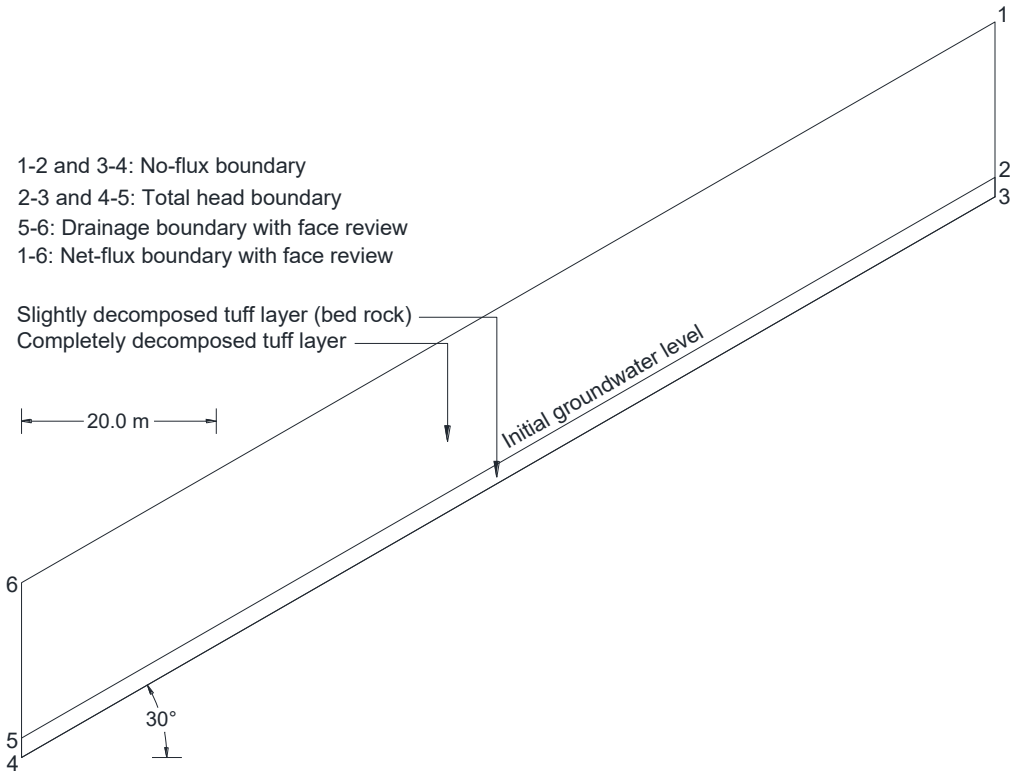



*Figure 5.59. Slope dimensions and boundary conditions.*

*Table 5.19. Slope mechanical and hydrological characteristics.*

| Material | | Completely decomposed tuff | Slightly decomposed tuff |
|---|---|---|---|
| Effective cohesion (kPa)[1,2] | | 4 | 10 |
| Effective friction angle (degree) [1,2] | | 35 | 45 |
| $\phi^b$ (degree) [3,4] | | 16 | 13.8 |
| Unit weight (kN/m$^3$) [1,2,5,6] | | 18 | 23 |
| Saturated hydraulic conductivity (m/s)[1,7] | | 1.5 x 10$^{-5}$ | 2 x 10$^{-7}$ |
| *Fredlund – Xing Parameters[8] | a (kPa) | 63.14 | Saturated |
| | n | 2.118 | |
| | m | 0.665 | |

*The solver from the VWC function estimated the hydraulic conductivity function. [1](Maunsell Geotechnical Services Ltd., 2007); [2](Wong, 2020); [3](Ho & Fredlund, 1982); [4](Gan & Fredlund, 1996); [5](Liu et al., 2022); [6](Cuomo & Di Perna, 2020); [7](Premchitt et al., 1986); [8](Carsel & Parrish, 1988).

### b) Results

Rainfall pattern effect: Figure 5.60 depicts the relationship between time and the FOS with a 50 kPa initial suction. This study focuses on the temporal effect of varied rainfall patterns with

the same cumulative rainfall. The delayed pattern recorded the lowest FOS one hour before the rainfall event. At the end of the rainy event, the normal pattern recorded the lowest FOS. The advanced and average patterns reported the lowest FOS two hours after the rainfall event ended. The time gap between these events is three hours. The Lowest FOS was observed due to the advanced pattern (1.107), followed by the normal pattern (1.115) and then the average pattern (1.117), while the delayed pattern recorded the highest FOS among all patterns (1.125). This is demonstrated by the fact that water continues to seep after the rainstorm event. For the delayed pattern, a massive amount of water penetrated the slope initially, giving the water time to seep. The massive amount of water at the end of a rainy event demands more time to seep for the advanced pattern, resulting in a low FOS and a delay in the failure time.

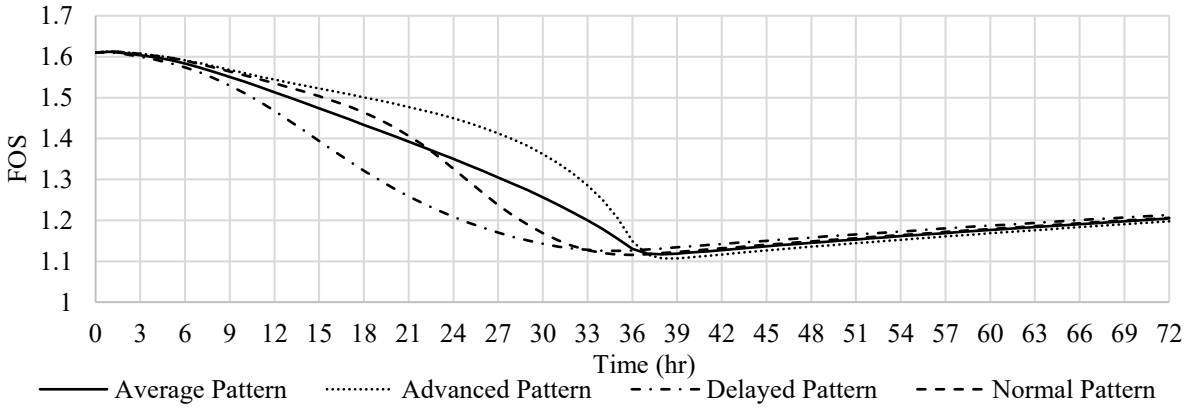



*Figure 5.60. FOS for different rainfall patterns.*

Initial conditions effect: There were four simulations: no initial suction, 50 kPa, 75 kPa, and 100 kPa (Figure 5.61). These initial conditions replicate the antecedent effect, where the two subsequent rainfall patterns were average patterns with an intensity of 25 mm/hr. The lower the initial suction, the lower the FOS. For initial suction of 50 kPa, 75 kPa, 100 kPa, and non-kPa, the FOS reported 1.61, 1.65, 1.68, and 1.70, respectively. The decrease in shear strength ($\tau$) demonstrates this, where the lower the suction, the lower the shear strength. For initial suctions of 50 kPa, 75 kPa, 100 kPa, and non-kPa, the healing time necessary for the slope to return to its original FOS (1.70) was determined to be 31.5 d, 29 d, 26.5 d, and 24.5 d,

respectively. The greater the initial suction, the less the slope is impacted by infiltration and shear strength loss, and the faster the slope recovers.

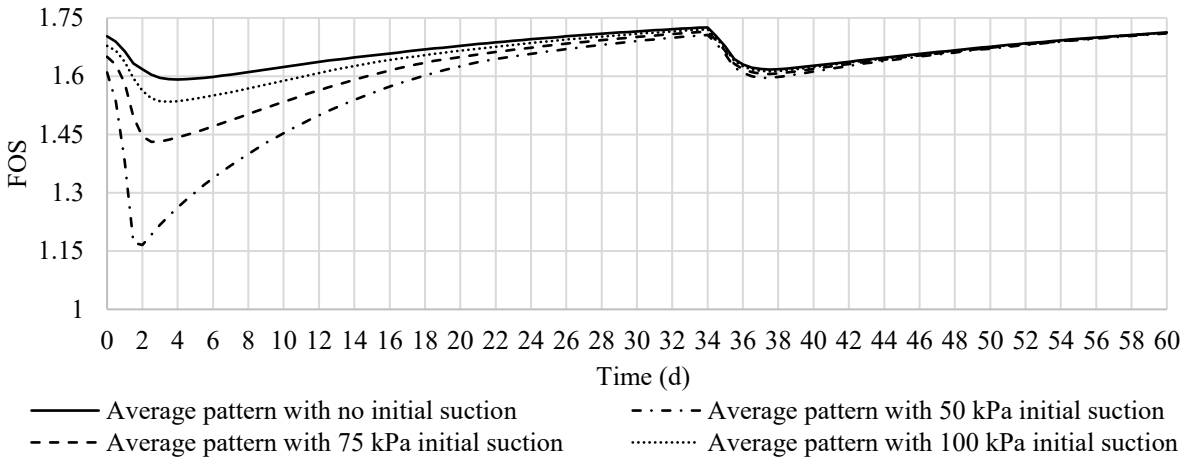



*Figure 5.61. FOS for different initial suctions versus non-rainfall duration.*

Rainfall intensity and duration effect: Figure 5.62 depicts three intensities with no initial suction assumptions: 25 mm/hr, 50 mm/hr, and 75 mm/hr. The higher the intensity, the lower the FOS, as the minimum FOS of 25 mm/hr, 50 mm/hr, and 75 mm/hr was 1.59, 1.35, and 0.90 at 4 d, 2 d, and 1.5 d, respectively. It indicated that the stronger the intensity, the faster the slope response. The time needed for the slope to return to its original (FOS of 1.70) value was determined to be 24.5 days, 30.5 days, and 32.5 days, respectively. According to this study, the higher the intensity for the same duration, the longer the healing time. As shown in Figure 5.63, the same conclusion may be expressed for the same rainfall intensities with varied durations: 24 hr, 36 hr, and 48 hr, where the recovery time was 21 d, 24.5 d, and 27 d, respectively. This may be represented by the quantity of water that infiltrates the soil; the more water that infiltrates the soil, the longer the recovery period.

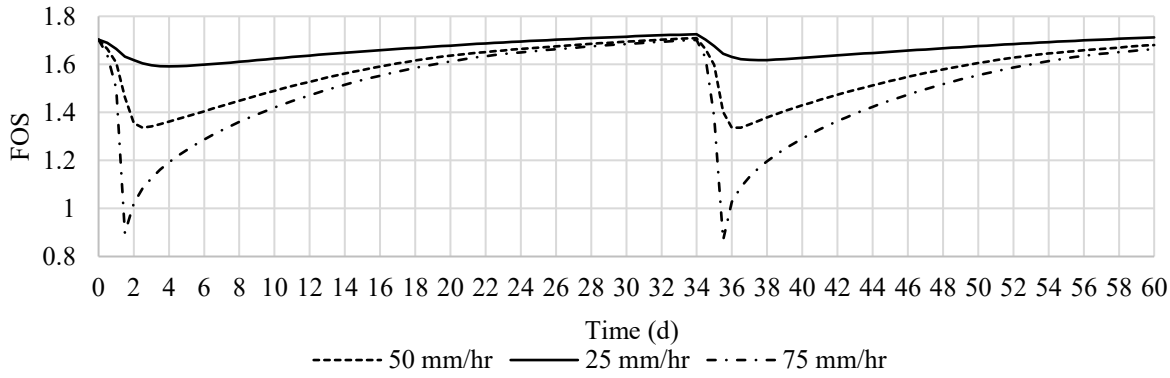



*Figure 5.62. FOS for different rainfall intensities versus non-rainfall duration.*

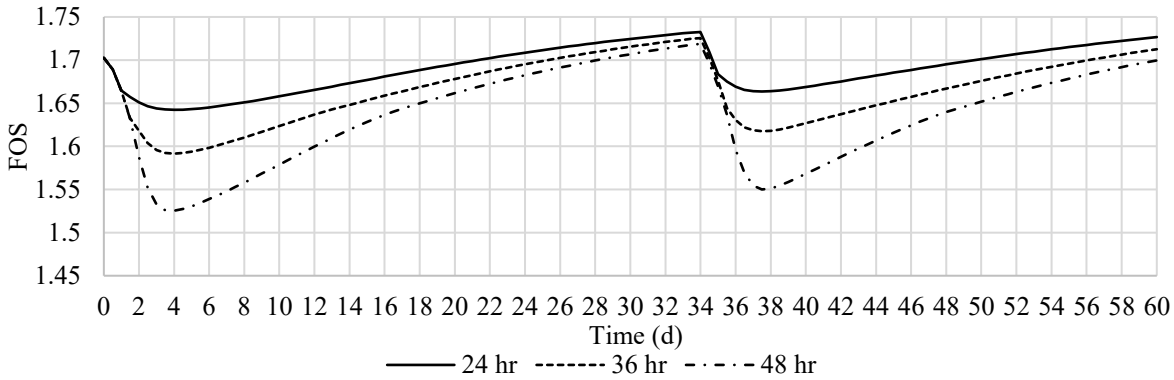



*Figure 5.63. FOS for different rainfall durations versus non-rainfall durations.*

Hydraulic conductivity effect: Figure 5.64 depicts a parametric study that considers three distinct hydraulic conductivities: $1.5x10^{-5}$ m $s^{-1}$, $3x10^{-5}$ m $s^{-1}$, and $9x10^{-6}$ m $s^{-1}$. The hydraulic conductivity strongly influences the recovery time; the higher the hydraulic conductivity, the shorter the recovery time. This is demonstrated by the soil's high conductivity, which allows water to percolate more quickly. Non-duration rainfall occurred after 15 days, 30.5 days, and more than 32.5 days for hydraulic conductivities of $1.5x10^{-5}$ m $s^{-1}$, $3x10^{-5}$ m $s^{-1}$, and $9x10^{-6}$ m $s^{-1}$, respectively. This figure shows that if the slope has not fully recovered from the initial rainfall event, its stability will be substantially affected by the second rainfall event. It emphasizes the significance of knowing the factors that influence the selection of effective rainfall. Effective rainfall is the duration of rainfall that occurs after the recovery period and is unaffected by any event that occurred before the recovery time.

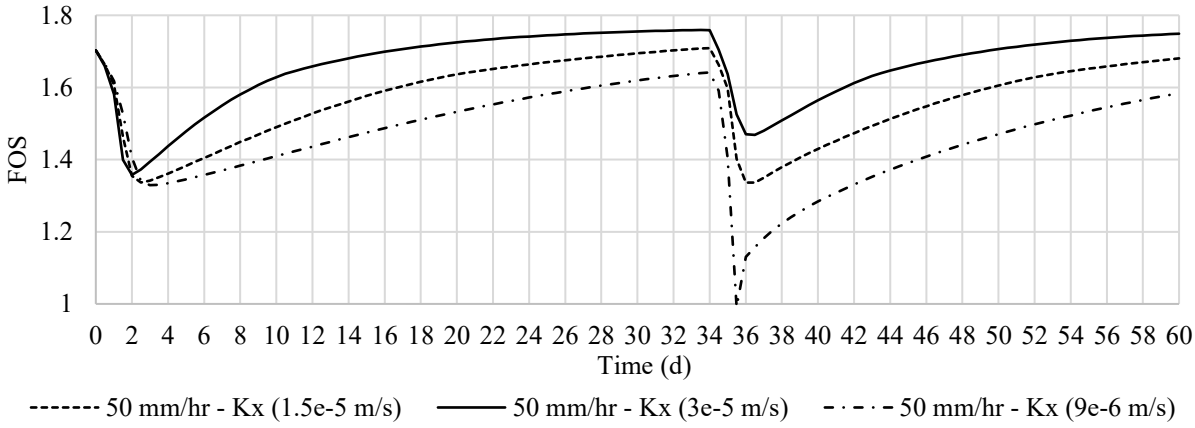



*Figure 5.64. FOS for different hydraulic conductivities versus non-rainfall duration.*

### *5.4.4. Geometry Effects on Rainfall-Induced Landslides: A Parametric Study*

Acknowledging that many studies have adopted simplified approaches by using hypothetical rainfall patterns and idealized slope conditions, thus overlooking the critical influence of actual hillslope geometric boundaries and real-world hydrologic processes, is a shortcoming that is increasingly significant in the context of climate change. In light of this gap, the section evaluates the impact of realistic hillslope geometries and bedrock formations on rainfall-induced landslide vulnerability. By integrating actual rainfall patterns, including their random intensity and duration, into a deterministic modeling framework using Geo-Studio (2023), and by modeling the landslide at the Fai Ngo Shan Service Reservoir (Maunsell Geotechnical Services Ltd., 2007), this research addresses two key questions: Why do landslides tend to concentrate in specific locations, and how can monitoring sites be optimally distributed to enhance early warning systems and risk mitigation? This research is designed to advance the understanding of landslide dynamics and to support more resilient hazard management strategies under evolving climatic conditions.

#### a) Results

The parametric numerical study was designed to isolate the influence of geometric boundary conditions on slope behavior, with the assumptions stated in Section 3.4.2 (refer to Figures 3.18 and 3.19). The variables for the parametric study are summarized in Table 5.20. Three

subsurface curvature profiles, planar, converging, and diverging, were each evaluated with a spectrum of inflow-to-outflow area ratios (R/r) to quantify curvature impacts on subsurface flow dynamics. Five observed rainfall events representing a range of antecedent moisture conditions were defined, coupled with intensity–antecedent windows. Figure 5.65 illustrates the selected rainfall series. Black bars mark the focused analysis time steps, and blue bars denote the antecedent moisture windows. These windows, combined with the detailed landslide cross-section (Figure 4.6), integrate realistic rainfall forcing and terrain geometry to enhance the prediction and mitigation of rainfall-induced landslides under a changing climate. XY displacement at the slope toe (BH3; Figure 4.6), where soil thickness is minimal, and vulnerability is highest, was numerically monitored. For the parametric study, values were min–max normalized (0 = no movement; 1 = maximum observed displacement). Results are summarized in Table 5.20 and plotted in Figure 5.66.

Figure 5.66 displays five displacement curves, one for each rainfall intensity and antecedent moisture window. Within each curve, red segments denote diverging curvature scenarios and black segments denote converging curvature scenarios. These deformation trends differentiate the diverging and converging curvature cases across the five intensity–antecedent windows: Windows 1-2 (low intensity, low antecedent): diverging slopes show negligible toe displacement at all R/r ratios, whereas converging slopes exhibit notable movement due to initial flow boundaries and antecedent moisture. Windows 3–4 (high intensity, low vs. high antecedent): diverging slopes undergo increasing deformation, but still minor, driven primarily by rainfall intensity, while converging slopes display amplified displacements from the combined antecedent and intensity effects. Window 5 (low intensity, high antecedent): both geometries register significant displacements, highlighting the delayed but dominant influence of elevated antecedent moisture. For the converging cases, the larger the R/r ratio, the greater

the effect of rainfall, whereas for the diverging cases, the greater the R/r, the less the rainfall effect on the slope.

*Table 5.20. The parametric study variables and results.*

| Condition | Code | R/r | Surface normalized XY- surface displacement at BH3 (refer to Figure 4.6) | | | | |
|---|---|---|---|---|---|---|---|
| | | | Window 1 (T1-T10) | Window 2 (T1-T20) | Window 3 (T1-T30) | Window 4 (T1-T31) | Window 5 (T1-T32) |
| Planer (benchmark) | 0-0 | 1 | 0.0018 | 0.0028 | 0.0072 | 0.0862 | 0.5010 |
| Diverging | 1-1 | 2 | 0.0012 | 0.0011 | 0.0010 | 0.0009 | 0.0010 |
| | 1-2 | 1.4 | 0.0007 | 0.0007 | 0.0007 | 0.0013 | 0.0022 |
| | 1-3 | 1.2 | 0.0010 | 0.0012 | 0.0018 | 0.0026 | 0.0063 |
| | 1-4 | 1.1 | 0.0012 | 0.0017 | 0.0026 | 0.0041 | 0.0653 |
| | 1-5 | 1.05 | 0.0013 | 0.0020 | 0.0033 | 0.0072 | 0.1554 |
| Converging | 10-1 | 2 | 0.1068 | 0.7067 | 0.9047 | 0.9481 | 1.0000 |
| | 10-2 | 1.4 | 0.0028 | 0.0717 | 0.6825 | 0.7702 | 0.8665 |
| | 10-3 | 1.2 | 0.0021 | 0.0046 | 0.3989 | 0.6336 | 0.7990 |
| | 10-4 | 1.1 | 0.0018 | 0.0032 | 0.0298 | 0.2824 | 0.5909 |
| | 10-5 | 1.05 | 0.0016 | 0.0027 | 0.0081 | 0.0937 | 0.5231 |

- R/r is the inflow-to-outflow area ratio for constant soil thickness, enforcing continuity (Figure 3.19).
- Windows 1–2: low intensity, low antecedent moisture
- Window 3: high intensity, low antecedent moisture
- Window 4: high intensity, high antecedent moisture
- Window 5: low intensity, high antecedent moisture

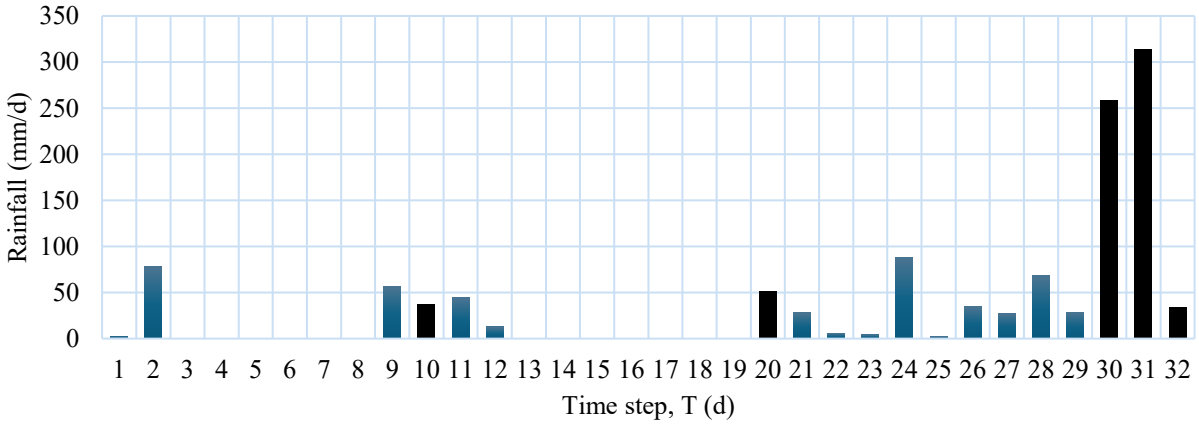



*Figure 5.65. Rainfall pattern with target event and antecedent moisture windows.*

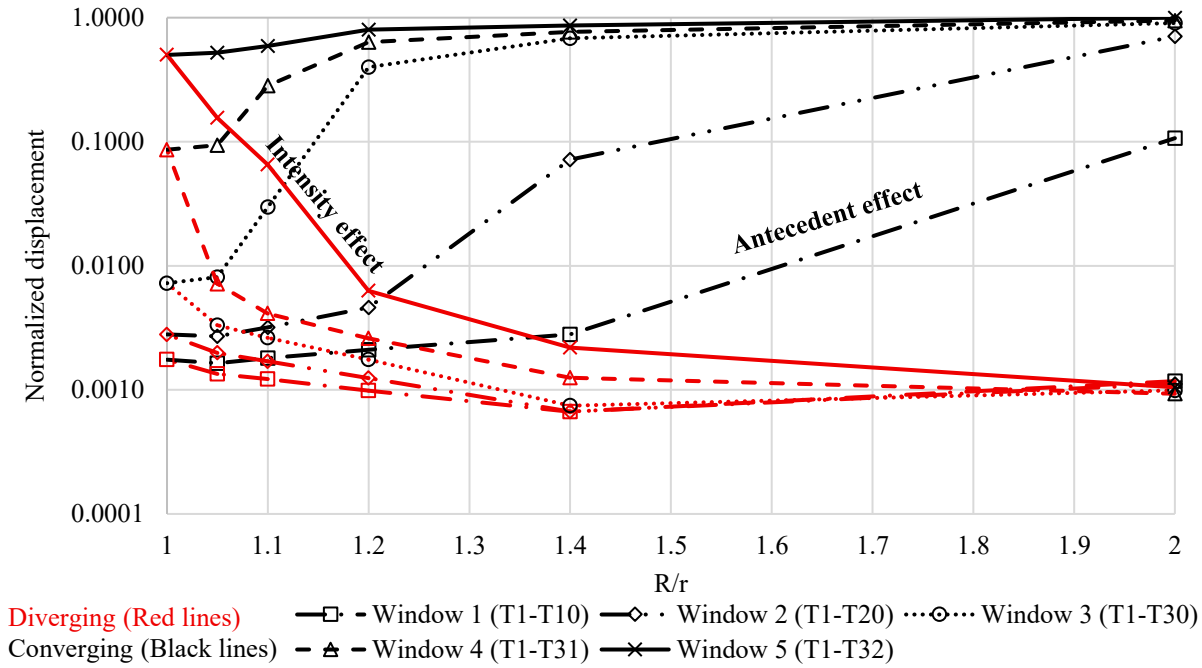



*Figure 5.66. Normalized slope-toe displacement across subsurface curvature and rainfall intensity–antecedent moisture scenarios.*

These trends reflect radial groundwater flow mechanics. In diverging geometries, inflow spreads into an expanding cross-section, damping pore-pressure buildup and limiting perched groundwater rise. As R/r increases, sensitivity to rainfall intensity decreases accordingly. In converging geometries, inflow is funneled into a shrinking cross-section, concentrating pore pressures and elevating the groundwater table even under modest rainfall. Higher R/r (smaller outflow area) further intensifies this effect, making converging slopes particularly sensitive to antecedent moisture. Figures 5.67.a and 5.67.b illustrate the temporal variation of perched groundwater levels under two geometric boundary conditions: diverging and converging slopes, both with an R/r ratio of 1.4. These figures reflect the response of each configuration to varying rainfall inputs over time. The converging geometry (Figure 5.67.a) exhibits a pronounced sensitivity to both rainfall intensity and initial perched groundwater levels, resulting in significant groundwater rise. In contrast, the diverging geometry (Figure 5.67.b) shows minimal response to precipitation, indicating limited groundwater accumulation.

These results underscore the critical role of subsurface flow directionality in modulating rainfall-induced pore pressure buildup.

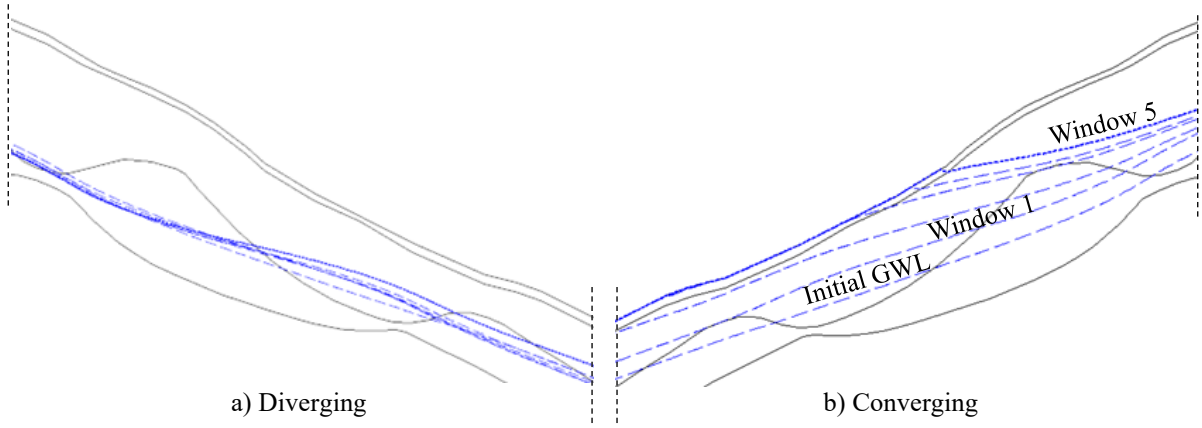



*Figure 5.67. The perched GWL variation with time (Windows 1-5): a) for the diverging geometry (R/r = 1.4); b) for the converging geometry (R/r = 1.4).*

### *5.4.5. Numerical Slope Stability Modeling (Focused Site Specific Analysis)*

This study section investigates the influence of bedrock (i.e., the variation of the rock layer thickness along the slope and its effect on the local vulnerability of the slope). This employs a deterministic approach using the commercial packages Geo-Studio (2023) and MPM (Anura3D, 2023). The landslide at the Fai Ngo Shan service reservoir was modeled to validate the numerical model framework. While the report by Maunsell Geotechnical Services Ltd. (2007) thoroughly investigates the case study, this study elucidates several inquiries. For example, how does the data frequency affect the temporal prediction of the landslide? How do Hong Kong's natural landslides collapse shallowly?

#### a) Boundary conditions

The case study described in Sections 4.3 and 5.2.2 provides the boundary conditions adapted for the small-deformation analysis. The HCF was calculated by the solver based on the SWCC function listed in Table 5.1. An initial pressure head of (suction = 75 kPa) was assigned to account for the antecedent effect before 7/20/2005, as suggested by Rahardjo et al. (2007). For

the large deformation analysis, Table 5.21 summarizes the material properties. To reduce computational expense, several simplifications were applied: permeability and infiltration rate used in the simulations were set several orders of magnitude higher than field values (Soga et al., 2016); the base material was treated as rigid; and frictional boundary conditions were imposed in the momentum balance following the approach of Du et al. (2023).

*Table 5.21. Material characteristics utilized in Anura3D.*

| Parameter | | Value (CDT) | Parameter | Value (CDT) |
|---|---|---|---|---|
| Material type | | Unsaturated material-2-phase with suction effect | Initial porosity[1] | 0.3 |
| Constitutive model | | Mohr-Coulomb [1] | Density solid (kg/m³) [1] | 2650 |
| Effective cohesion (kPa) [3] | | 4 | Density liquid (kg/m³) [1] | 1000 |
| Effective internal friction angle (°) [3] | | 34 | K0-value (calculated by the solver) | 0.4408 |
| Effective Poisson ratio (υ) [4] | | 0.3 | Intrinsic permeability liquid ($m^2$) [3] | 5.52e-9 |
| Effective young modulus (MPa) [4] | | 20 | Bulk modulus liquid (MPa) | 60 |
| Minimum degree of saturation [5] | | 0.25 | Liquid dynamic viscosity (Pa · s) | 1e-6 |
| Maximum degree of saturation [5] | | 0.80 | Number of material points | 3 |
| Van Genuchten and Mualem parameters | Po (kPa) | 76 | Infiltration rate (m/s) [3] | -9.3e-3 |
| | λ | 0.4735 | | |

[1](Cuomo et al., 2021); [2](Du et al., 2023); [3](Maunsell Geotechnical Services Ltd., 2007); [4](GCO, 2000); [5](Liu et al., 2020).

Figure 5.68 shows the boundary conditions assigned for the post-failure analysis. The friction between the CDT layer was assigned to be 0.75, while the friction between the CDT and the ground surface was set to be 0.20 till the convex break in the slope and 0.5 afterward to simulate the field conditions (Maunsell Geotechnical Services Ltd., 2007; Yang et al., 2014). The mass scaling of 100, the homogenous local damping of 0.1, and the Courant number of 0.8 were adequate to keep the stability of the model (Anura3D, 2023; Soga et al., 2016). The purple points (i.e., $P_{6251}$, ..., $P_{9176}$) represent the displacement and velocity monitoring points after failure initiation. The ground surface was assigned with seepage and infiltration faces. The water table level was assigned as the level obtained from numerical analysis before the failure. The rapture surface is assigned from the validated LEA and FEA analysis. This model was to investigate how shallow failure initiates (not to predict failure).

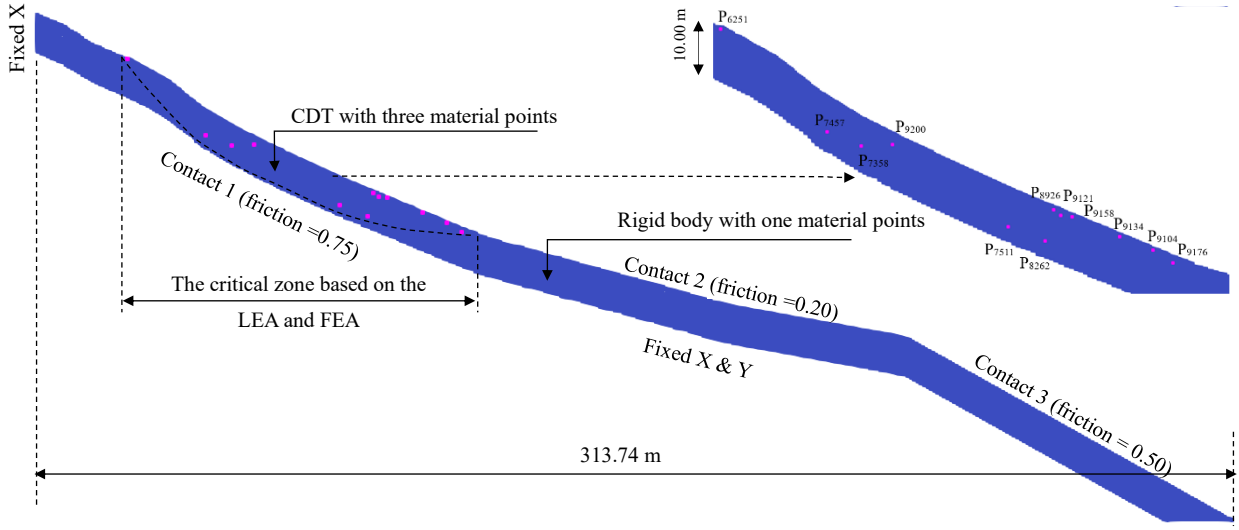



*Figure 5.68. The model boundary condition and meshing for MPM.*

### b) Analysis results and discussion

The rainfall from 19 to 21 August 2005 was simulated using an hourly time step (Figure 5.69. a). Figure 5.69.b presents the sensitivity analysis with the parameters identified with the legend with hourly rainfall. Furthermore, some parameters were modified to investigate their effects, such as the saturated hydraulic conductivity, the shear strength parameters, and the value of Phi-B. This Figure illustrates that the failure time is not affected by the factors above, while these factors affect the value of the FOS. Figure 5.69.c presents three cases concerning the rainfall pattern (daily, hourly, and average). Significantly, the rainfall pattern affects the temporal prediction, as the change between the FOS using daily and hourly rainfall can be understood by the rainfall pattern effect. The hourly precipitation predicts failure at 16:00 20/8/2005, the daily rain predicts failure at 16:00 21/8/2005, and the average rainfall indicates failure at 11:00 21/8/2025. It should be noted that the reported failure is assumed to be at 6:00, 21/8/2005.

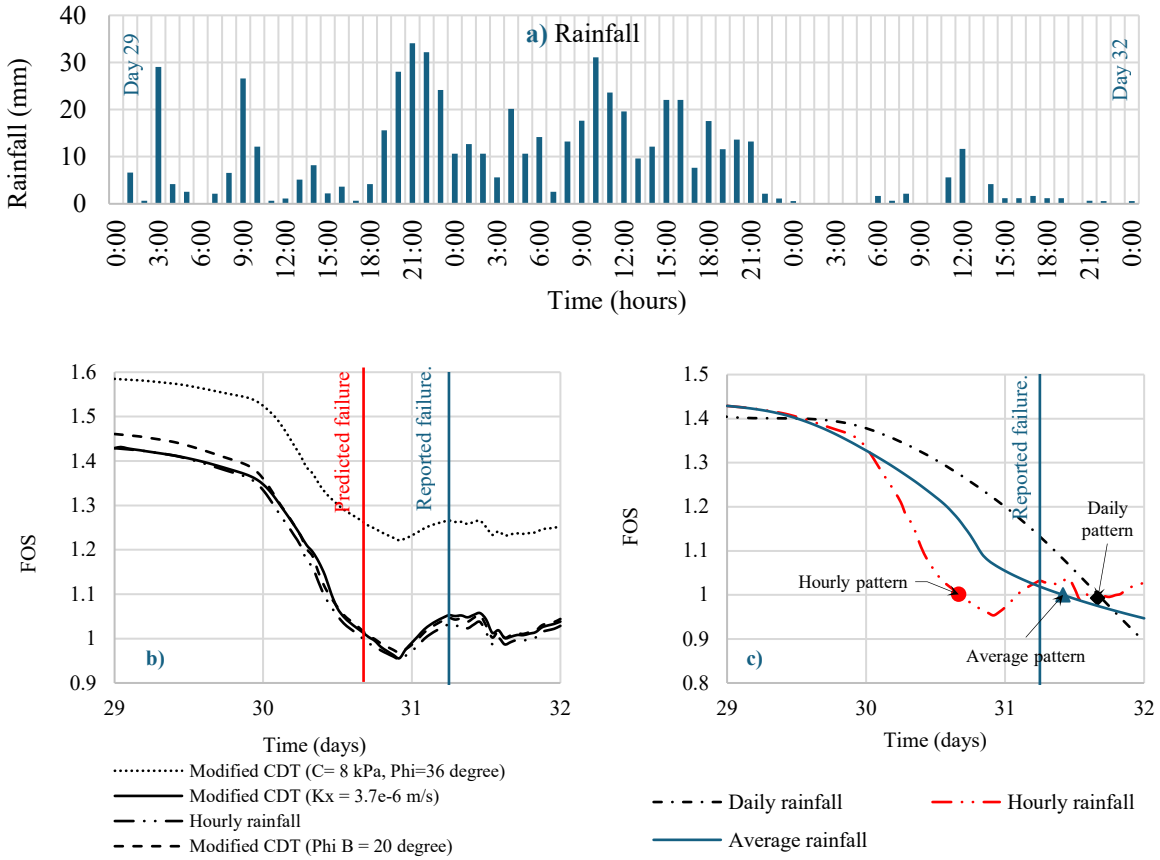



*Figure 5.69. The temporal variation of the FOS: a) rainfall between 19 and 21 August 2005; b) LEA (Phi-B)-sensitivity analysis; c) LEA (Phi-B)-different rainfall.*

According to Cheng et al. (2007), it is challenging to use optimization techniques to estimate the precise slip surface due to the effect of the local minima safety factor. Adopting the LEA's critical FOS alone might be deceptive when local minima FOS have comparable FOS but distinct locations. A safety map approach that uses LEA to depict both the local and global minima was developed by Baker & Leshchinsky (2001). For example, the safety map is shown in Figure 5.70. a, b utilizing the Phi-B approach through LEA and FEA. In Figure 5.70, the color map emphasizes the critical zone in red while omitting the presentation of zones with greater FOS to reduce confusion and streamline clarity. The predicted location (Figure 5.70) of the failure surface is in line with field observations in considering the following facts: a) The main scarp location, b) the side scarp that was noticed between chainage 42 m to Chainage 95 m (refer to Maunsell Geotechnical Services Ltd., 2007 – pp.16 “On the southern side of the

debris trail, small side scars, up to 1 m high, were observed, although these were less well-defined owing to the thicker debris deposited there."), c) the exact position of the failure surface at TP7 (the only trial pit at this location), d) the highly saturated soil following chainage 42 because of the increased GWL, and e) the existence of unsaturated soil from the upper zone above the failure surface (refer to Maunsell Geotechnical Services Ltd., 2007).

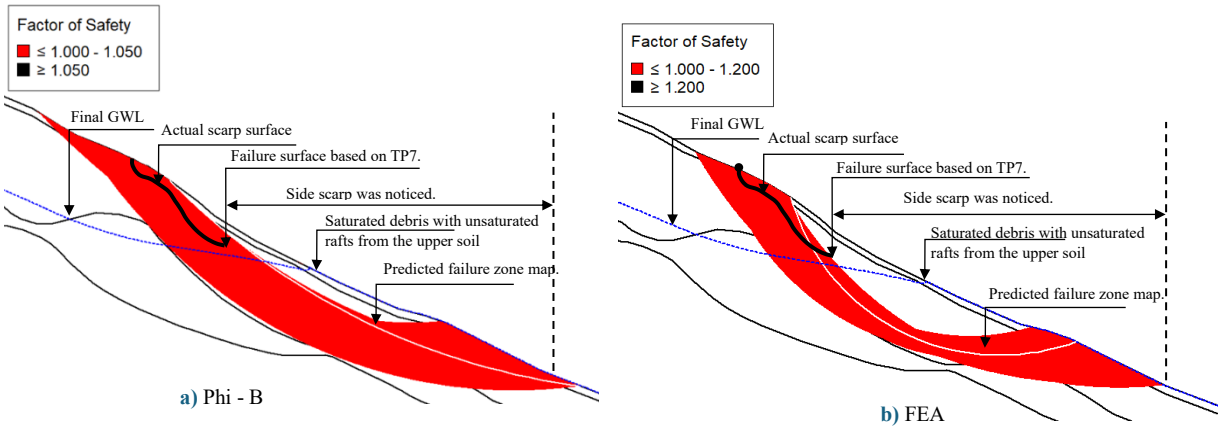



*Figure 5.70. The critical zone map: a) LEA-Phi-B; b) FEA.*

Recognizing that each calibration and simulation involves uncertainty and unavoidable numerical assumptions. In this section, the results obtained from the FEA and LEA analysis were used in the large deformation analysis to compare the results with the field work: the measured debris velocity and travel distance, along with other fieldwork observations, which are reported in the original report by Maunsell Geotechnical Services Ltd. (2007). Additionally, this section is to validate that the slip surface identified from the FEA is reasonably accepted. In other words, the ground stresses are difficult to assess for natural slopes (Fredlund et al., 1999; Krahn, 2003), where it is challenging to estimate the precise slip surface (Cheng et al., 2007). According to Soga et al. (2016), there are three different scales involved in the dynamics of landslides: (a) the microscopic scale, which is characterized by the contact between grains; (b) the mesoscale, which represents microstructural effects like grain rearrangement; and (c) the macroscopic scale, which is essential for soil layering, faults, and other geological and geomorphologic features. The macroscopic scale is simplified in MPM due to the

computational effect: the model dimensions are large enough to simulate the exact soil characteristics, necessitating massive computational and time efforts (i.e., months to solve the model). As a result, the critical slip surface from the LEA and FEA was simulated to be a contact surface in large deformation analysis (Anura3D, 2023). This simplification was utilized to study the post-failure and debris flow. Additionally, this model investigates how shallow failures initiate (not to predict failure).

The actual failure time (T) (i.e., the post-failure and debris flow time) was assumed to be 55 seconds as reported in (Maunsell Geotechnical Services Ltd., 2007). In the MPM, it took 2012 seconds. Comparing the observed data to the numerical findings, time ($t_{numerical}$) and velocity ($v_{numerical}$) were normalized by the failure time (55 s) as suggested by Soga et al. (2016), as presented in Equations 5.1 and 5.2. Figure 5.71 highlights the travel distance at the end of the analysis ($t_{numerical}$ = 2012 s) with the actual final observed ground level (i.e., after debris flow). Figure 5.72 presents the travel horizontal distance versus the normalized velocity solid magnitude.

$$t' = t_{numerical} / T \qquad (Equation\ 5.1)$$

$$v' = v_{numerical} * t_{numerical} / T \qquad (Equation\ 5.2)$$

It should be noted that due to the significant debris flow distance, which is around 300 meters, the monitoring points (purple points), besides the legend scale, help visualize the process over time. Twelve points were selected to monitor the travel distance and the velocity along the slip failure (Figure 5.68). These figures illustrate how a local failure in a natural slope initiates: the points near the toe ($P_{9176}$, $P_{9140}$) start to move due to the loss of support and strength loss as indicated by the LEA and FEA. A movement of the shallow layer follows the trend of such soil due to the loss of support ($P_{9134}$, $P_{9158}$, $P_{9121}$, $P_{8926}$). This causes a slight movement and a slide along the critical slip surface ($P_{6251}$, $P_{7457}$, $P_{7358}$, $P_{7511}$, $P_{8262}$, $P_{9200}$), causing the clear scarp at the crest that followed the main landslide. This finding is supported by field

investigation and tests. Figure 5.72 supports this and quantifies the displacement and the velocity for different points. The points near the toe traveled longer distances with an average normalized (velocity 1) of 5.78 m/s, which matches well with the estimated values from the field investigation program (Maunsell Geotechnical Services Ltd., 2007): the field estimated values are visualized by Black dots (Figure 5.72). Meanwhile, the points near the critical surface and inside the soil body record much smaller movement and a small average normalized (velocity 2) (Figure 5.72).

Figure 5.72 depicts the travel distance for the twelve points (P6251..., P9176) versus normalized time. It is clear that all of the points had reached their stability, and there was no further movement before the model came to an end at 55 seconds normalized. This is evidenced by the vertical line of travel distance, which shows no change over time for all points. The points drawn in black lines reached stability earlier than those represented in red lines. Furthermore, this graph depicts the average velocity for the indicated points: average velocity 1 and average velocity 2. Average velocity 1 is for the points that experienced large movements to validate with the actual estimated velocities in the field, which are represented by black scattered dots. It should be noted that there were six field estimations of the velocity at different travel distances, as shown in Figure 5.72. Average velocity 2 refers to the points that were mostly near the slip surface and had much lower velocity. The presented results show that numerical simulations and field observations agree well on debris velocities and estimated travel distances. These results give a good agreement with field observations for the following reasons: a) the most derbies was recorded before chainage no 224 (i.e., greater than 95%) while a small volume of debris flow was recorded after chainage 224 (i.e., around 5%); b) the simulated velocities are well consistent with the field estimated velocities as presented in Figure 5.72; c) The crest recorded low velocity and small travel distance (refer to Figure 5.72)

which match with field observation where unsaturated soil rafts were observed in the upper debris (Maunsell Geotechnical Services Ltd., 2007).

## 5.5. SUMMARY

This chapter presents the results of the DL and deterministic model implementations. It describes the training, validation, and testing procedures for each model and benchmarks their performance against real-world case studies. The findings provide scientific support for operationalizing three primary tools: a) landslide prediction; b) operational VWC prediction; and c) automated parameterization. The chapter also reports several parametric studies that elucidate the physical behavior of rainfall-induced landslides.

1

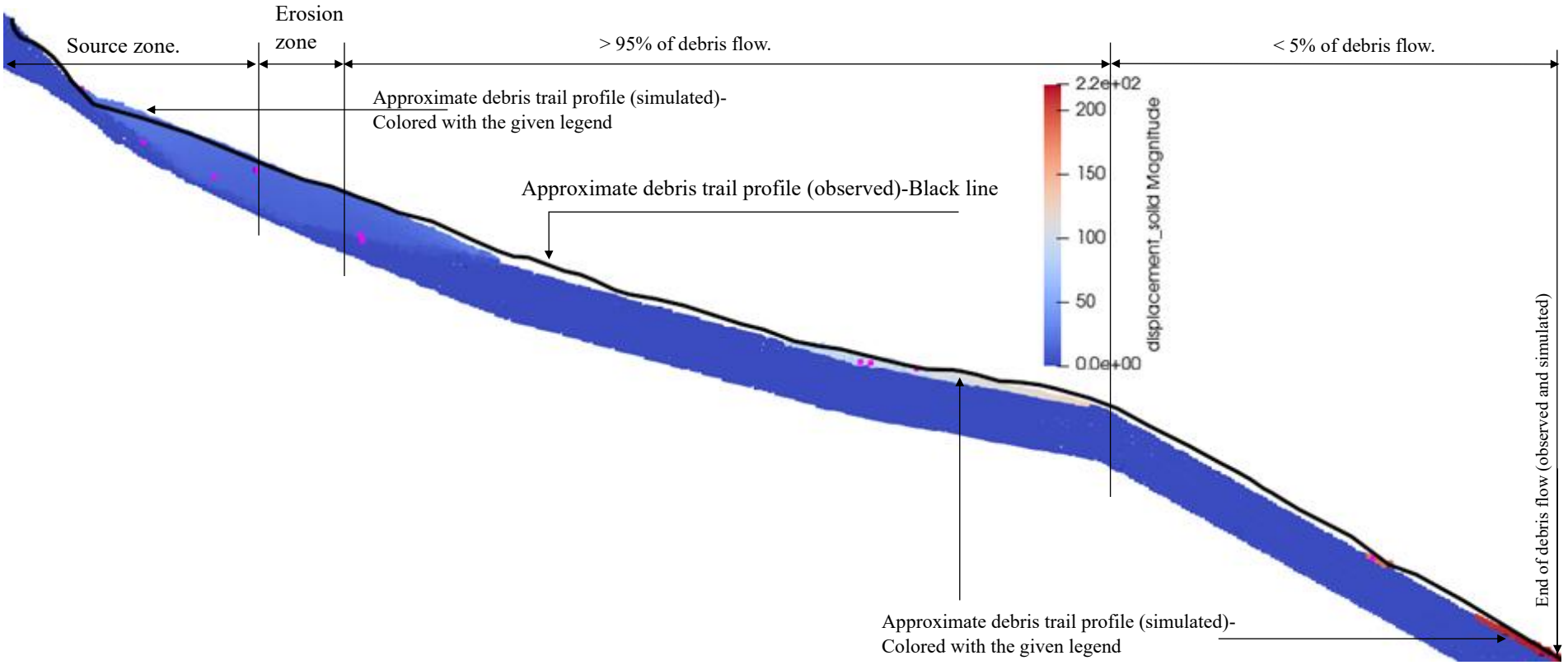



*Figure 5.71. Post-failure simulation of the debris flows (t numerical = 2012 s) with the maximum displacement.*

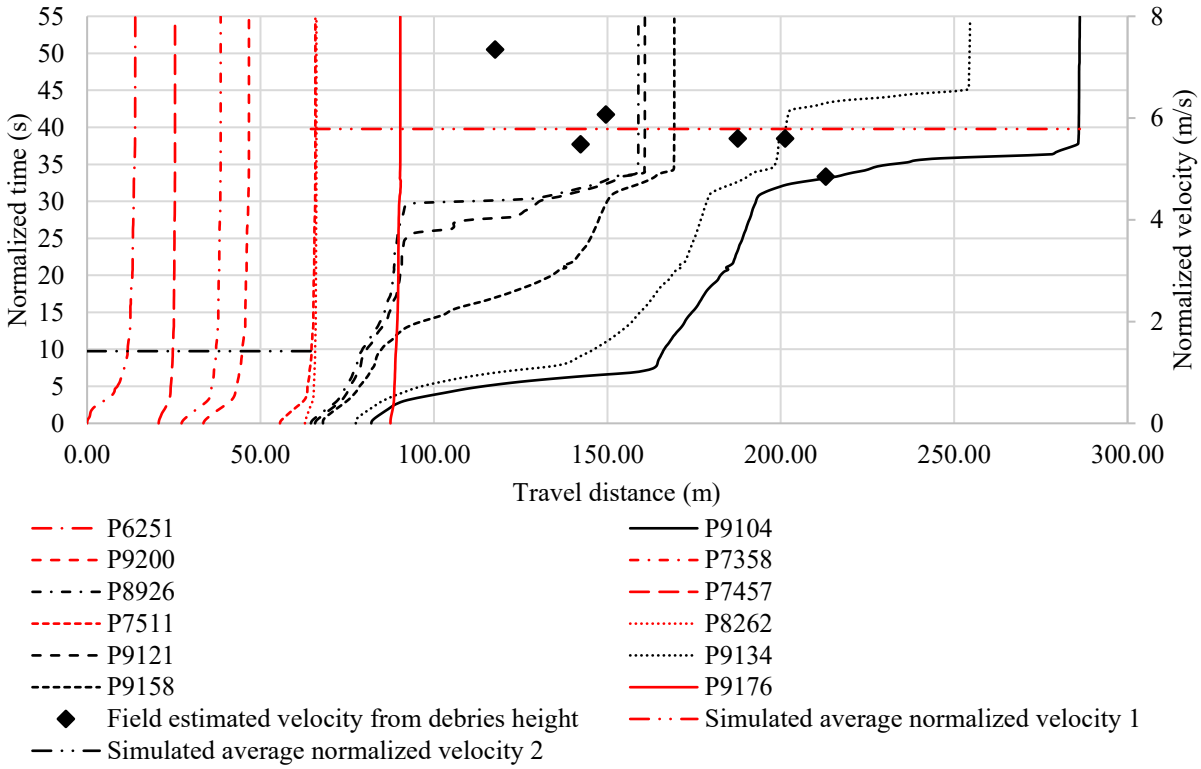



*Figure 5.72. Normalized average velocity versus travel horizontal distance.*

# CHAPTER 6: IMPLEMENTATION AND DEPLOYMENT OF WEB-BASED PREDICTIVE TOOLS

## 6.1. INTRODUCTION

This chapter details the development of four web-based applications designed to implement the methodologies proposed in this study: the landslide probability prediction tool, OWHK, APHs, and the integrated stability system. Built with Python for backend operations and HTML/CSS for frontend design, these applications ensure both functionality and user-friendly interfaces. The chapter discusses the development process of each application and provides an overview of their graphical user interfaces (GUIs), demonstrating how users can interact with the functionalities to apply the predictive models in real-world scenarios. Each interface incorporates intuitive visualizations and transparent feature importance indicators, translating complex model outputs into actionable information. Embedded validated case studies demonstrate reliability, while automated functionalities lower expertise barriers, enabling decision-makers to confidently apply these tools in practice.

## 6.2. LANDSLIDE PREDICTING TOOL

### a) Application interface and architecture:

The deployment process of the model is detailed in this section, following the evaluation of the model's performance outlined in Section 5.2. The developed ML models (CNN, GRU, and LSTM) were successfully deployed as an interactive web application using Streamlit (Khorasani et al., 2022), providing end-users with operational forecasting capabilities for landslide susceptibility assessment. The deployment architecture supports single-site predictions and batch processing of multiple cases, incorporating uncertainty quantification through Monte Carlo dropout analysis (refer to Figure 6.1). The deployed system features a user interface with two primary operational modes (Figure 6.1.a). The single prediction mode

provides a streamlined interface for individual site assessments, incorporating location coordinates that are automatically processed to determine the nearest rain gauge station and corresponding zone classification. The system includes input validation and provides multiple rainfall data entry methods, including manual input, CSV upload, and sample datasets. Site characteristics such as slope angle, elevation, curvature, material description, and nearby structures are configurable through an intuitive sidebar interface. Temporal parameters, including rainfall, initial boundaries (i.e., initial month), and landslide date (i.e., target date), accommodate various assessment scenarios, while Monte Carlo sampling parameters allow users to adjust uncertainty analysis precision. The batch prediction mode enables users to upload Excel files containing multiple test cases with the required geomorphological, meteorological, and temporal features (Figure 6.1.c). The system validates input data structure, ensuring compliance with the 28-day temporal window requirement and proper feature formatting. Users can select from available models and configure Monte Carlo simulation parameters for uncertainty analysis. Date range filtering capabilities allow for temporal subsetting of large datasets, enabling focused analysis on specific monitoring periods or extreme weather events.

b) Prediction results and visualization:

The application generates prediction outputs with model-specific risk thresholds and visual representations (Figure 6.1.b). Deterministic predictions are complemented by Monte Carlo analysis, providing uncertainty quantification through configurable simulation runs. The results interface displays a color-coded gauge chart visualization with model-specific warning and failure thresholds. Statistical metrics, including deterministic prediction values, Monte Carlo means, uncertainty estimates through standard deviation, and confidence intervals, provide comprehensive assessment information. Automated risk classification interprets results as "Low Risk," "Warning," or "High Risk" based on model-specific thresholds. Distribution

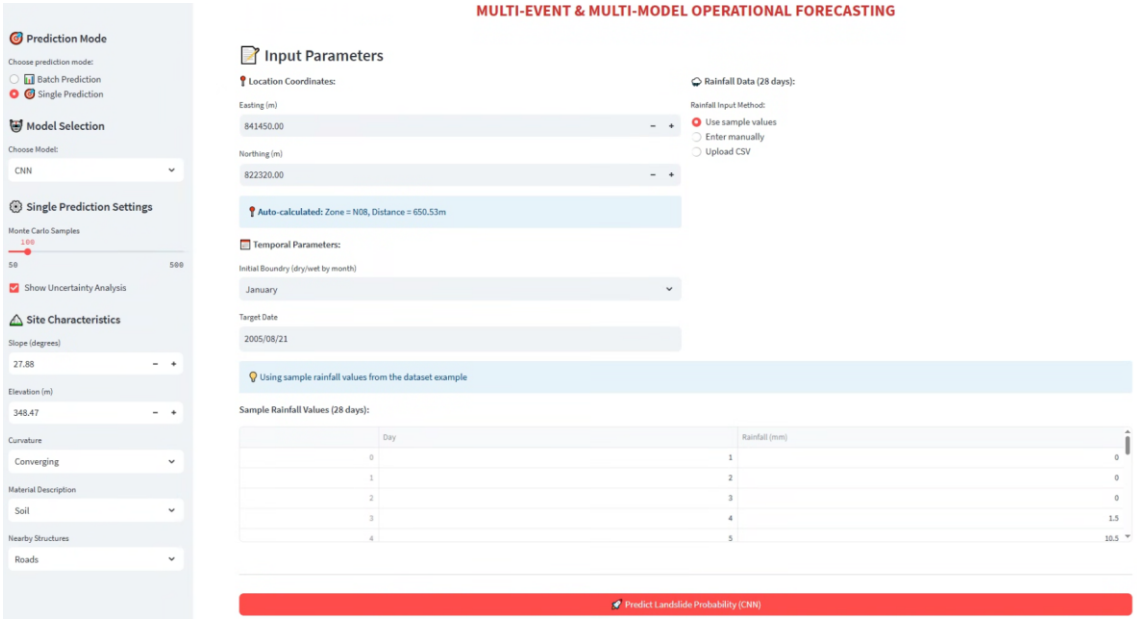
analysis through histogram visualization shows Monte Carlo prediction patterns, enabling users to assess prediction reliability and identify potential outliers or bimodal distributions that may indicate model uncertainty.

### c) Operational features and quality assurance

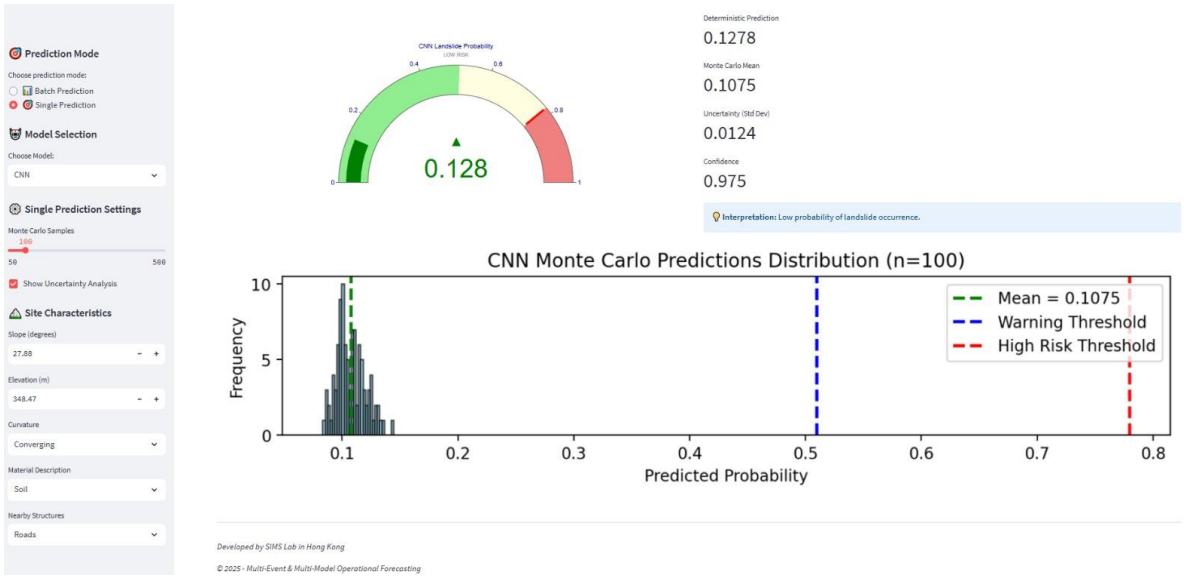
The deployed application incorporates several operational features ensuring reliability and usability in real-world scenarios. Data validation mechanisms verify input completeness, feature consistency, and temporal requirements before processing, preventing erroneous predictions from incomplete or malformed datasets. Model selection capabilities allow users to compare predictions across different architectures, enabling consensus-based decision-making and cross-validation of results across multiple modeling approaches. Uncertainty quantification through Monte Carlo dropout provides confidence estimates essential for risk-based decision making in geotechnical applications. The system includes comprehensive error handling with informative messages for data format issues, missing features, or processing failures, guiding users toward the resolution of input problems. Performance optimization through model caching and efficient data processing ensures a responsive user experience even with large datasets containing hundreds of landslide cases. Documentation integration provides in-application guidance on data requirements, feature specifications, and result interpretation, reducing the learning curve for new users and ensuring consistent application across different operational contexts. The system maintains backward compatibility with existing datasets while supporting extensibility for future model improvements or additional features. This deployment represents a complete transition from research prototype to operational tool, enabling practical application of the developed multi-event, multi-model forecasting framework for landslide assessment in Hong Kong and similar geological environments.

To demonstrate operational readiness, the developed web application was tested on a recent, independent case: the Yiu Hing Road hillside between October 2022 and September

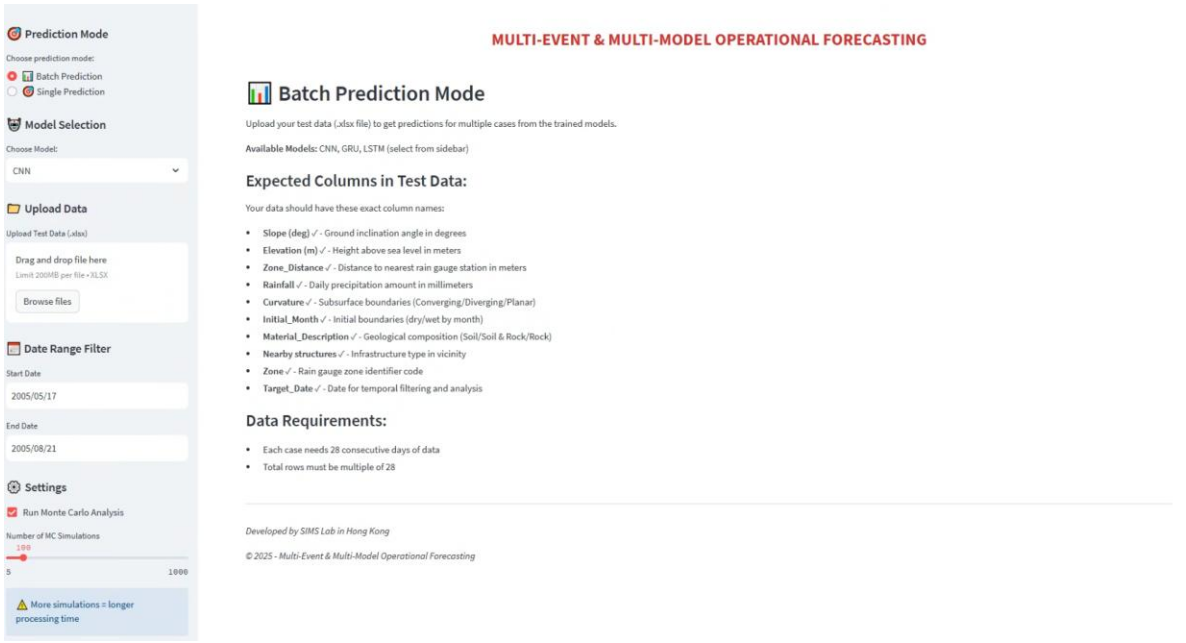
2023 (Figure 6.1.e). This period was outside the model's training timeline, providing a strict test of its generalization. Throughout this extended period, the model consistently showed no landslide warning, which correctly matched the official reports of no failures occurring. This successful real-world validation confirms the tool's reliability and its practical utility for ongoing slope monitoring.




(a) Single prediction mode interface displaying input parameter configuration



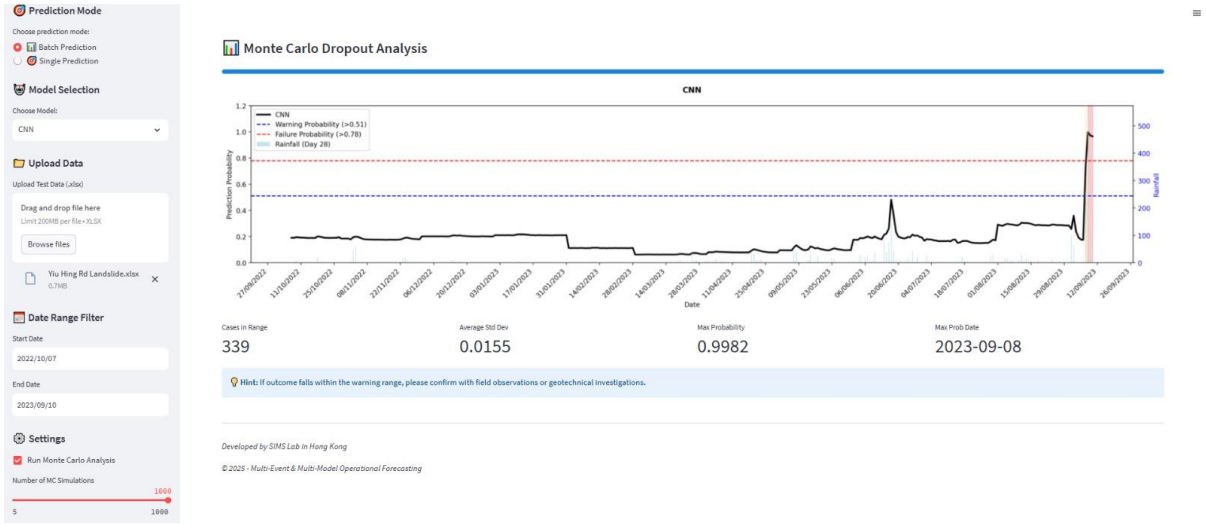

(b) Single prediction results interface featuring gauge chart visualization




(c) Batch prediction mode interface showing data upload requirements




(d) batch prediction results interface




(e) performance against recent events outside the trained temporal data

*Figure 6.1. Multi-event and multi-model operational forecasting tool*

### d) Tool capabilities and limitations

While the proposed tool is specifically calibrated for Hong Kong's natural slopes, its underlying methodological framework is architected for regional adaptability. Successful extension to new regions requires: a) Integration of region-specific dynamic features (e.g., rainfall temporal patterns) and static features (e.g., local geological formations) (Table 6.1); b) Customization based on dominant failure mechanisms; c) Systematic retraining of core model architectures (i.e., LSTM/GRU/CNN configurations) and hyperparameter optimization strategies; d) The standardized evaluation framework (AUC, precision-recall, uncertainty quantification) remains consistently applicable; d) Modular H5 file structure for model weights and datasets, with corresponding updates to interface design and core model integration. The generalizability of the model to other regions faces two primary challenges: (a) inherent aleatory and epistemic uncertainties, as discussed by Mondini et al. (2023), which manifest as model-related and data-related limitations including imprecise event timing and spatial rainfall variability; and (b) difficulties in acquiring high-resolution temporal data for some features such as vegetation cover, human activities, and detailed soil properties. These challenges warrant focused attention in future research to develop reliable data inventories, which will subsequently

enhance model performance and generalization capabilities. Future work should pursue two interconnected goals: first, to enhance the dataset by incorporating high-resolution temporal features such as precise vegetation indices and detailed soil properties; and second, to leverage this improved data to systematically investigate the optimal class ratio for modeling. The selection of this ratio is a critical methodological factor that is inherently dependent on the quality and completeness of the underlying spatiotemporal data. A more comprehensive dataset will enable a robust exploration of how dynamic slope changes influence model performance and the ideal balance between landslide and non-landslide cases

*Table 6.1. Landslide prediction tool target improvement.*

| Aspect | Current literature | This study | Future Directions |
|---|---|---|---|
| Rainfall representation & Temporal dependencies | Cumulative short-term or average rainfall; hypothetical scenarios (Qiu et al., 2024; Xiao et al., 2023; Zhao et al., 202; Mondini et al., 2023) | Time-dependent modeling capturing sequential rainfall patterns | Incorporate transformers or hybrid architectures for enhanced temporal modeling |
| Methodological framework | Single-method validation; Single threshold (0.5); Balanced 1:1 ratio common (Wang et al., 2020) | Multi-model validation (LSTM, GRU, CNN); Dual-threshold approach; 3:1 imbalance handling | Enhance the dataset by integrating high-resolution temporal features (e.g., vegetation and soil properties). Systematically investigating optimal class balancing ratios. |
| Model performance | AUC = 0.93 with low precision (Xiao et al., 2023); AUC ≈ 0.90, Balanced accuracy = 0.81 (Mondini et al., 2023) | AUC = 0.92, Balanced accuracy = 0.83, Precision = 0.80 (at 0.5 cutoff). Adopting the dual-threshold further improves reliability. | Develop hybrid approaches that integrate deterministic modeling with DL, leveraging DL for forecasting seepage responses from historical patterns while utilizing deterministic equations to accurately represent slope stability mechanisms. |
| Operational deployment | Research-focused applications without standalone tools | Interactive interface using rain gauge and terrain data | Develop a digital twin system: creating a closed-loop system where model predictions are continuously refined against observed slope behavior |

## 6.3. OPERATIONAL VWC FORECASTING TOOL (OWHK)

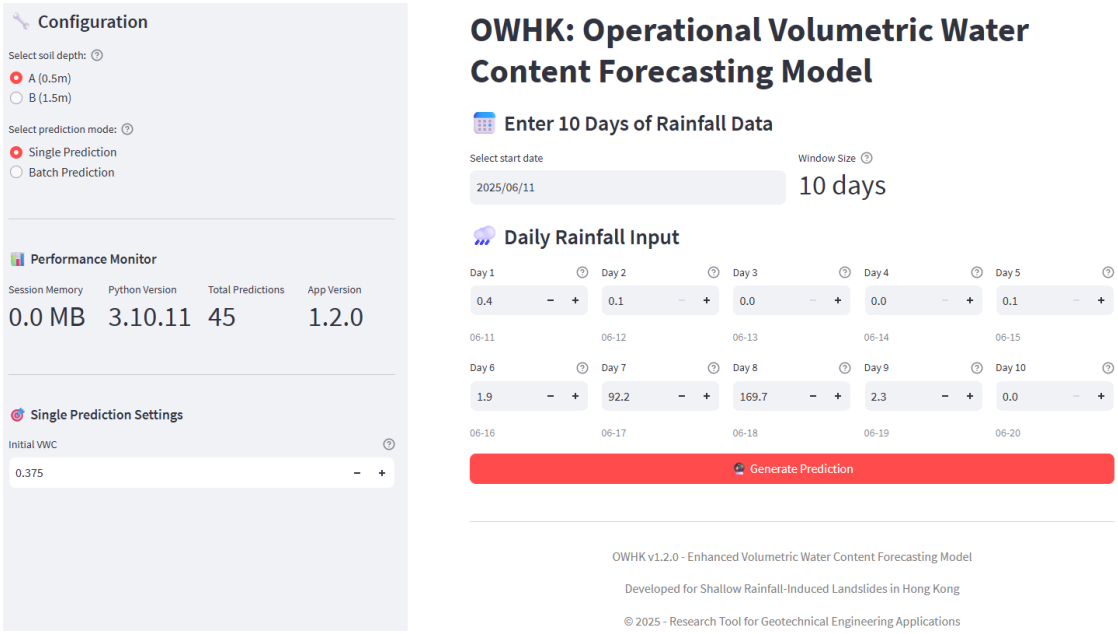
The deployment process of the model is detailed in this section, following the evaluation of the model's performance outlined in Section 5.3.1. This study develops an operational model with relatively broad applicability and spatiotemporal adaptability, distinguishing it from narrowly focused, site-specific models. The OWHK model has been deployed as a web-based

application using Streamlit (Khorasani et al., 2022), implementing a modular three-layer architecture: presentation layer (GUI), business logic layer (model inference engine), and data layer (trained LSTM models). The backend utilizes TensorFlow 2.12 for model inference, with all multi-trained LSTM models loaded at startup and cached in memory using a singleton pattern to minimize inference. As illustrated in Figure 6.2.a, the application features an intuitive interface with configuration options in the sidebar, including soil depth selection (0.5m or 1.5m) and prediction modes (single or batch). The intelligent model selection algorithm automatically determines the appropriate LSTM model based on the user input. Performance testing reveals efficient operation with average inference times of 0.40 seconds for single predictions and 1.02 seconds for batch processing of 10 windows (Figure 6.2.b-d). Memory usage remains stable at approximately 50MB during operation. Key optimizations include: (i) vectorized operations for batch processing, improving throughput by 60% compared to sequential processing; (ii) cached data transformations and normalization; and (iii) asynchronous file processing to maintain responsive interface performance.

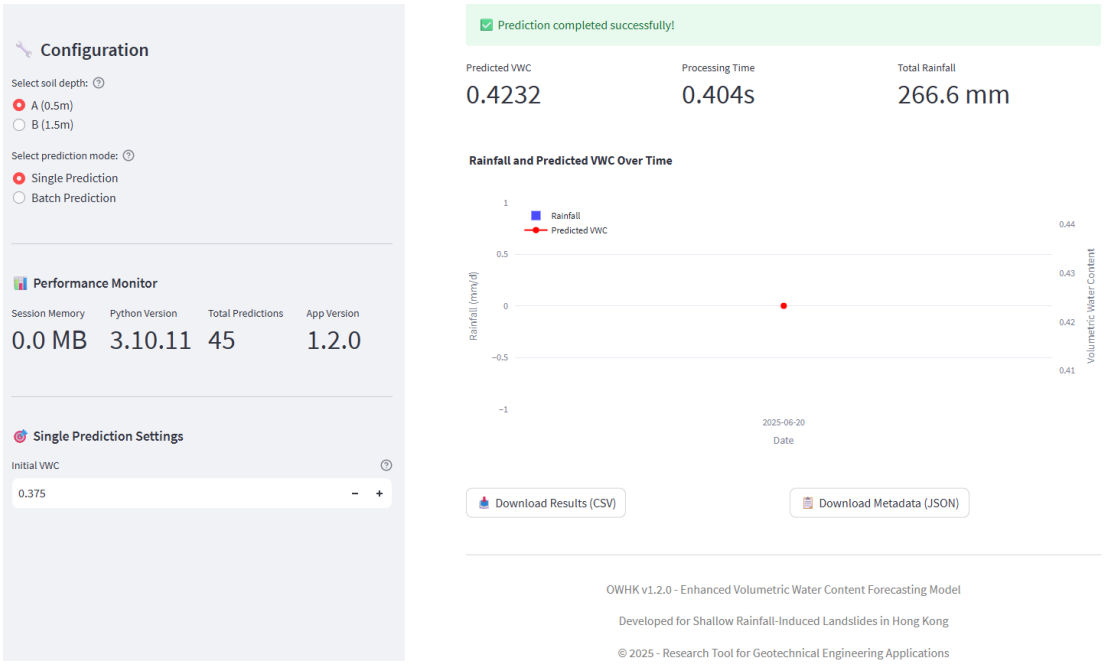
The application implements chunked processing for large datasets to prevent memory overflow, with automatic cleanup of temporary variables. The current configuration supports file uploads up to 200MB, accommodating typical monitoring datasets from monthly to multi-year time series. For datasets exceeding 1,000 prediction windows, the system automatically partitions data into chunks of 500 windows, displaying real-time progress tracking (see Figures 6.2.c-d). Built-in validation handles missing data points and irregular time series through intelligent window creation, only processing valid sequences with non-null initial VWC values, ensuring robust operation with real-world monitoring data that often contains gaps or inconsistencies. Validation results indicate that the deterministic approach processes 1000 subcases in 17–24 seconds, whereas the Monte Carlo (MC) stochastic approach, with 1000 stochastic paths, requires 3 hours, 36 minutes, and 51 seconds. Given the comparable

performance exhibited by both evaluation methods, the deterministic approach offers a computationally efficient means of generating predictions at the interface without the overhead of additional sampling or computationally intensive procedures during inference. The interface provides comprehensive data format guidance (Figure 6.2.c), a performance monitoring dashboard, and downloadable templates.

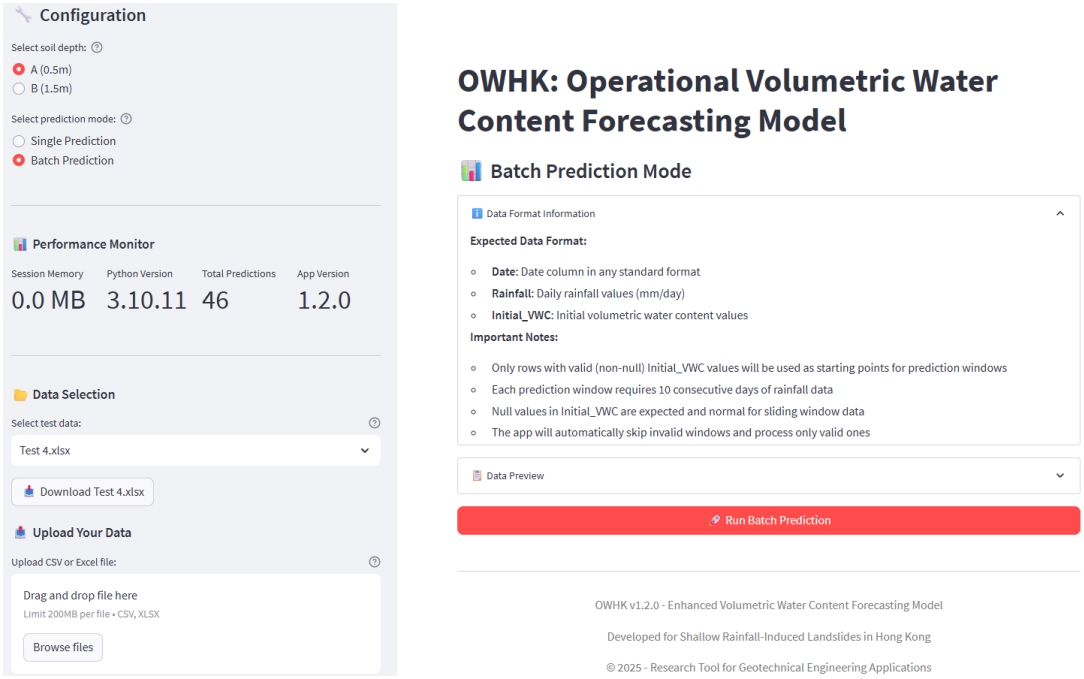
Future work will enhance the core models (i.e., the trained DL models) with additional data collection and improvements outlined in Table 6.2, with updates reflected in the App. The interface provides comprehensive metadata (processing time, model version, input parameters, timestamp) available in downloadable JSON format to ensure reproducibility. The app uses session-based, non-permanent data processing to guarantee privacy while maintaining transparency. Planned scalability enhancements include containerized cloud deployment, API endpoints for automated monitoring, and distributed processing for larger datasets. Its modular architecture facilitates seamless integration with real-time monitoring networks, establishing the application as a production-ready tool for operational landslide risk assessment.



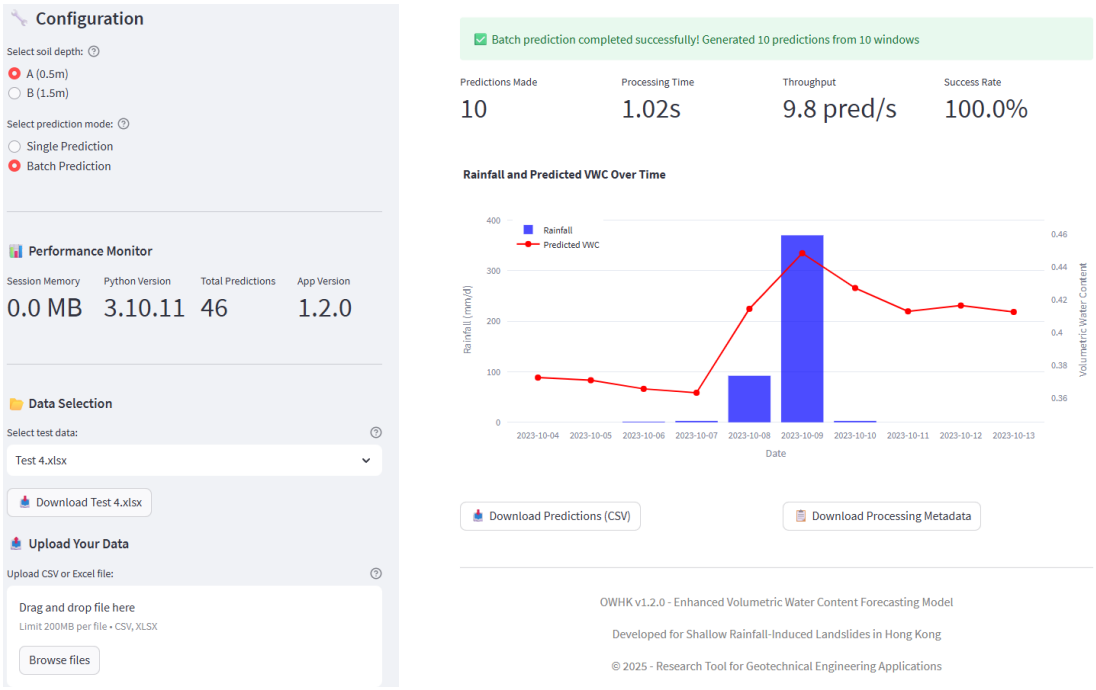

(a) Single prediction interface




(b) Prediction results from the single prediction




(c) Batch prediction interface




(d) Prediction result from the batch prediction

*Figure 6.2. Graphic user interface of the OWHK app*

The results demonstrate that OWHK yields accurate predictions attributable to a methodology that effectively reduces spatiotemporal uncertainties. First, rainfall is modeled as a time-dependent feature that explicitly accounts for its inherent characteristics (intensity, duration, and pattern), which are critical for reflecting hydrological variability. Temporally, an independent time series windowing approach that avoids inter-case dependencies while preserving the inherent intra-case dynamics was employed. In addition, numerically simulated patterns were incorporated to adapt the model to diverse real-world rainfall conditions. Spatially, the use of a multi-LSTM framework enables the model to reduce uncertainties arising from variability in soil properties across different locations and depths. Finally, a structured tuning process, from coarse to fine adjustments, with increased emphasis on fine-tuning. However, it is essential to note that the OWHK performance relies on both (a) the DL algorithm's capability, which influences computational requirements during training and

inference, and (b) the volume and quality of the data. Table 6.2 shows the current achievements and future directions.

This study addresses key aspects of practicability. First, it models VWC, a crucial parameter for mitigating landslide and flooding hazards. For instance, as shown in Figure 6.2, a slope triggered by nearly 400 mm/d of rainfall indicates high VWC, suggesting near saturation; this not only increases landslide risk but also raises the potential for flooding and elevated runoff ratios during subsequent rainfall events, necessitating further model calibration. Second, although wireless networks are widely used for subsurface monitoring, they face challenges such as energy consumption, memory limitations, and communication issues that can result in data loss. While existing literature discusses various strategies, from reducing data frequency and limiting active sensors, a critical gap remains in employing DL to capture the complex dependencies in subsurface responses. The proposed multi-LSTM methodology directly addresses this gap by modeling intricate spatiotemporal dynamics, thereby offering a practical tool to advance future analyses. Third, a budget-friendly monitoring network, approximately USD 1000 per sensor node, is modeled, whose VWC readings offer the most direct measure of rainfall infiltration and, thus, subsurface dynamics.

Although the proposed framework attains predictive accuracy, it remains computationally intensive. Each LSTM sub-model contains 248129 trainable weights, yielding 2233161 parameters for the entire ensemble; iterative fine-tuning further amplifies the cost. In addition, the present monitoring network may bias the representation of in situ soil variability. Expanding field measurements and evaluating several architectures, such as Transformers with sensitivity-analyzed grouping schemes, could lower the computational burden while maintaining or improving skill. The extensive field dataset, derived from 15 sensors over nearly two years of daily monitoring, initially validates the proposed operational tool. Nonetheless, further improvements could be achieved by: a) Extending the monitoring duration to capture

additional temporal variability, especially those that rarely temporarily occur; b) Incorporating additional features (e.g., evaporation, plant transpiration, air and soil temperature, humidity, wind speed, etc.) and experimentally simulating extreme future scenarios; c) Integrating a digital twin with real-time monitoring to facilitate dynamic model updates in response to possible deviations.

The validity of this tool is inherently limited by its reliance on data boundaries intrinsic to the data-driven approach. The model focuses on the 0.5 m and 1.5 m depths because those are the locations where reliable sensor data are available. The 0.5 m level captures near-surface moisture influenced by short-term atmospheric fluctuations, while the 1.5 m depth provides insights into more stable subsurface moisture dynamics. Future work aims to incorporate additional depths, which are expected to exhibit different responses. The primary objective of this study is to assess the performance of operational vs. site-specific models, paving a foundation for future work to expand the dataset and further enhance model performance.

*Table 6.2. Summary of OWHK's current achievements and future direction.*

| Aspect | Current setup | Current achievement | Future directions | Target improvement |
|---|---|---|---|---|
| Model | Multi-LSTM with emphasis on data preparation | Capture the complex spatiotemporal behavior | Explore Transformers /CNNs | Explore faster training with fewer trainable parameters and accurate predictions. |
| Dataset Volume | 15 sensor-based responses with roughly 2 years of daily monitoring | - Validate the initial hypotheses for developing operational models | Improve the dataset with a more extensive dataset for longer monitoring durations and several depths | Broaden applicability to different soils and conditions, especially those that rarely or temporarily occur |
| Triggering and features | Real-world rainfall patterns | - Simple usage<br>- Cost-effective | - Consider future scenarios through experimental-based simulations.<br>- Consider related features (i.e., evaporation, plant transpiration) | To accommodate future climatic challenges. |
| Tool versatility | Tolerated for the Hong Kong colluvium shallow soil | - Practicability<br>- User interface | Digital twin with a real-time monitoring system | Update model weights and performance in response to any possible deviations in model responses |

Regarding the subsurface multi-sensor predictions (see Section 5.3.3 and Table 6.3), this study has certain limitations that chart a course for future work. While the framework addresses uncertainties associated with rainfall and sensor data through its multi-feature approach, the inherent epistemic uncertainty of the model itself, a function of dataset volume and diversity, remains a consideration (Abbaszadeh Shahri et al., 2022). As this work is an initial step towards collecting more comprehensive subsurface data, a full quantification of model uncertainty is left for future research. Other limitations include: a) The responses from neighboring sensors were integrated to implicitly represent the spatial heterogeneity of the subsurface geomaterial, an approach supported by contemporary research (Saha et al., 2022). This strategy was necessitated by the common challenge of limited availability of high-resolution subsurface mapping, as noted by Abbaszadeh Shahri et al. (2024). Consequently, the hydrological and mechanical responses from spatially distributed sensors were used as dynamic proxy features, allowing the DL model to infer the complex spatial interactions and boundary conditions within the slope without explicit, continuous 3D geomaterial data. While this data-driven approach has proven effective, future work will focus on enhancing the explicit featuring of spatial soil characteristics by integrating geostatistical interpolation and 3D modeling techniques (Abbaszadeh Shahri et al., 2021; Shi & Wang, 2022; Kim et al., 2023; Xie et al., 2022) to further mitigate the constraints imposed by sparse subsurface investigation data.

b) The analysis is based on a specific field dataset, highlighting the general lack of available subsurface monitoring data. To improve generalizability, future work should develop a numerical model rigorously calibrated to the present dataset. This validated simulator can then produce scenario-based synthetic time-series (varying rainfall, initial states, soil parameters). These simulated responses, combined with realistic sensor noise, can augment training data and increase robustness across a wider set of field conditions. c) The transition from a high-performing historical model to a robust, real-time early warning system presents

challenges. The current work has not fully standardized the framework for a dynamic, continuously updating database. Therefore, future work will focus on: a) dynamic data normalization and streaming input; b) advanced handling of data heterogeneity and quality; c) model retraining and continual learning protocols (Abbaszadeh Shahri et al., 2024); and d) Model Optimization: There was limited exploration of alternative normalization techniques and activation functions, which may have the potential to further optimize model performance.

*Table 6.3. Summary of the subsurface multi-sensor prediction model achievements.*

| Aspect | Current Achievement | Future Recommendation |
|---|---|---|
| Spatial Representation & Data Integration | Implicitly modeled subsurface spatial heterogeneity by using neighboring sensor responses as dynamic proxy features (Saha et al., 2022), overcoming the limitation of scarce high-resolution geomaterial mapping (Abbaszadeh Shahri et al., 2024). | Explicit spatial characterization by integrating geostatistical interpolation and 3D geological modeling techniques (Abbaszadeh Shahri et al., 2021; Shi & Wang, 2022; Kim et al., 2023; Xie et al., 2022) to enhance feature engineering and mitigate sparse data constraints. |
| Framework & Methodology | Developed a tailored LSTM framework with rigorous hyperparameter tuning, holdout validation, and sliding windows. Used dropout/early stopping for robust training. | Explore alternative normalization techniques and activation functions to further optimize model performance. |
| Model Performance & Insight | Achieved high predictive accuracy, outperforming naive models. Demonstrated that the triggering condition is a multivariate spatiotemporal interaction, not just rainfall. | Further investigate and quantify model epistemic uncertainty (Abbaszadeh Shahri et al., 2022) |
| Validation & Analysis | Trained and saved several distinct models; performed a dedicated sensitivity analysis using DL attribution to evaluate feature importance in a temporal context. | Develop and use numerically calibrated simulators to create synthetic data for rare events, enhancing model generalizability and robustness. |
| Operational Application | Proved the value of integrating subsurface monitoring with DL as a cost-effective and accurate alternative to surface-based techniques. | Standardize the framework for real-time deployment: Implement dynamic data normalization, advanced data handling, and continual learning protocols (Abbaszadeh Shahri et al., 2024). |

## 6.4. AUTOMATED PARAMETRIZATION TOOL (APHs)

The deployment process of the model is detailed in this section, following the evaluation of the model's performance outlined in Section 5.4.1. This study introduces the APHs tool, a significant advancement in automating and streamlining the parameterization process of the HS model. Conventional methods typically rely on oedometer tests, unloading/reloading stress-strain data, empirical correlations, domain-specific assumptions, or expert judgment, all of which can be time-consuming, costly, and error-prone. In contrast, the APHs tool exclusively uses data from conventional loading triaxial tests, making it a versatile solution for geotechnical applications. The APHs tool employs DL, specifically LSTM networks, to capture the relationship between stress and strain under loading conditions. By modeling this relationship as a time series problem, the tool effectively leverages the sequential evolution of stress-strain responses, offering a more efficient approach (Figure 6.3).

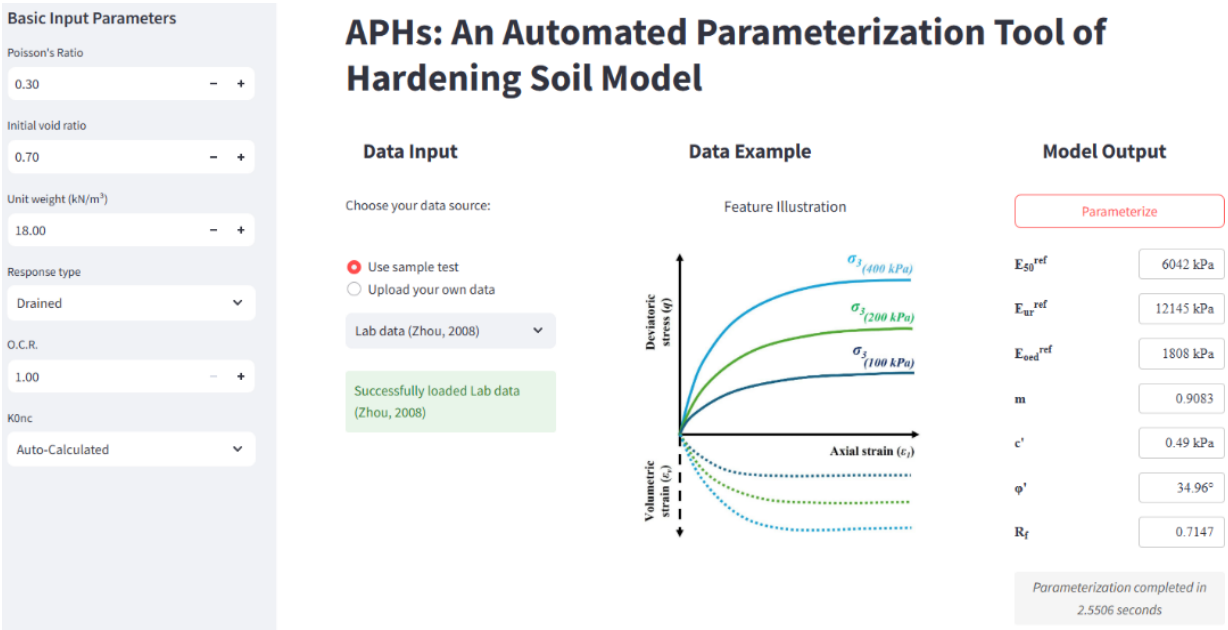



*Figure 6.3. The APHs tool.*

The development of the APHs tool systematically addressed key challenges related to data generation and model optimization. Given the limitation of insufficient triaxial laboratory data, the study focused on a specific case study in Hong Kong to establish a parameter space

suitable for numerical simulations. This parameter space was derived from field samples collected in Hong Kong and supported by existing literature, ensuring the tool's practical applicability to real-world conditions. To efficiently explore this parameter space, LHS was employed, facilitating the generation of diverse synthetic datasets that comprehensively cover the key parameters of the HS model. The numerical model SIGMA/W was then integrated to simulate soil responses under these defined conditions, producing stress-strain curves for DL training. For model optimization, a critical consideration was the variation in scale among input parameters, such as the stress vs. the strain values, and the target parameters, such as $c'$ vs. $E_{ur}^{ref}$. Data preparation played a significant role in this process, particularly through normalization techniques that ensured uniformity across all input features. An extensive coarse-to-fine-tuning strategy was employed. The process included early stopping, close monitoring of the loss curves, and reshuffling the training and validation sets to mitigate overfitting. Given the uniqueness of each sample, this approach enabled the model to generalize effectively.

Seven novel parallel LSTM models were implemented to enhance accuracy, each dedicated to predicting a specific parameter. This approach independently optimizes each parameter, reflecting the physical processes observed in conventional parameterization. For example, as shown in Figure 6.4, the multi-model assigns distinct weights to each parameter. It gives higher weight to features such as deviatoric stress for specific targets (i.e., $c'$). Meanwhile, volumetric strain has less impact, which is expected. This approach ensures the model successfully mimics the underlying physical processes. Similarly, for the $E_{oed}^{ref}$ which is found to be controlled with the volumetric strain features. Most significantly, as expected, $E_{ur}^{ref}$ is influenced mainly by deviatoric stress due to the constraints applied (i.e., $E_{50}^{ref} = 0.5E_{ur}^{ref}$) during dataset generation. Additionally, since $E_{50}^{ref}$ is correlated with $E_{oed}^{ref}$, $E_{ur}^{ref}$ also exhibits a relative correlation with volumetric strain, which governs $E_{oed}^{ref}$. Given the absence of

unloading/reloading data, the LSTM model derived correlations from a robust dataset, informed by literature-defined boundaries and validated LHS-generated parameter sets.

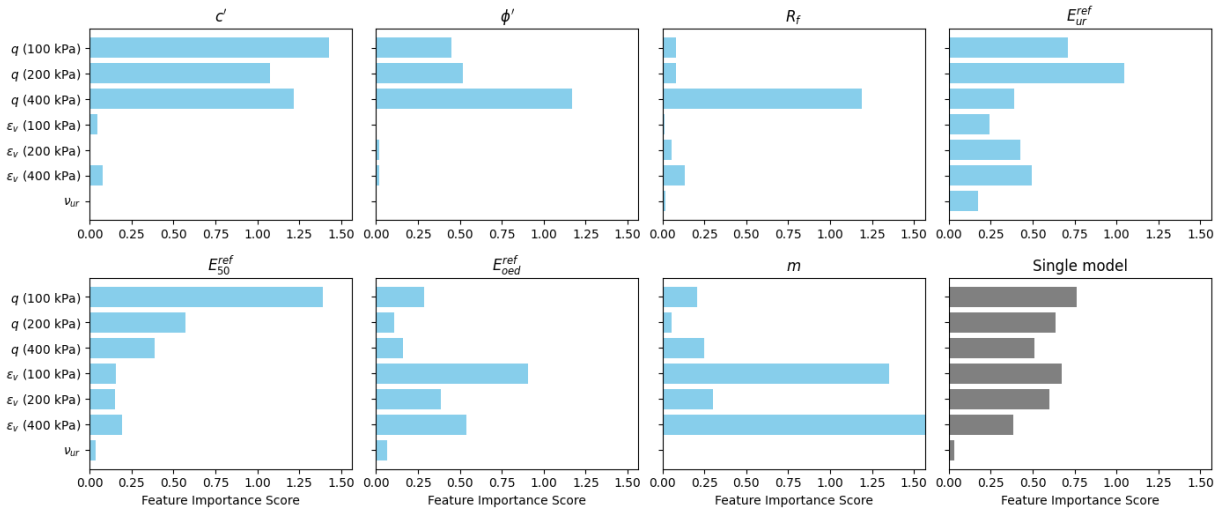



*Figure 6.4. Permutation feature importance: Blue (Multi-model) vs. Grey (Single model).*

In contrast, the single-model approach applies uniform loss weights based on the average predicted error of all seven parameters, which may increase the risk of overfitting and reduce generalization. This is evident in Table 6.4, where the multi-model achieves lower MAE values: for instance, $E_{ur}^{ref}$ sees a reduction from 1016.57 kPa (single model) to 380.38 kPa (multi-model) in the testing set. This trend is pronounced in the tested case study, where the multi-model shows an improvement, with an MAE of 104.41 kPa compared to 1477.93 kPa for the single model for $E_{50}^{ref}$. The multi-model approach delivers more accurate results on unseen data and shows robustness through strong alignment between its stochastic and deterministic predictions, as confirmed by the MC dropout simulation (Figures 5.49 and 5.50).

In contrast to studies that use optimization with physically informed models, this study adopts a DL approach. Rather than relying on established equations, the model learns underlying physical patterns directly from raw data, enhancing its flexibility and establishing it as a standalone tool with potentially high predictive accuracy. To illustrate its practical utility, a comparison was made between the proposed approach and that of Jin et al. (2016), who

employed a real-coded genetic algorithm to optimize parameters for Hostun sand. WebPlotDigitizer (Rohatgi, 2022) converted triaxial test plots into numerical data along usable strain sequences. Considering $\phi'$ as an example, Jin et al. (2016) employed a search range (10° to 50°) and obtained values between 27° and 29° for different constitutive models. Remarkably, the proposed approach predicts 28.24° for the Hostun sand despite using data outside the training scope (minimum training value of 29°). As shown in Figure 6.5, APHs closely match lab results with only minor errors.

*Table 6.4. The single model vs. multi-model evaluation metrics: MAE.*

| Parameters | MAE values [value in brackets corresponds to the single LSTM model] | | | | | | |
|---|---|---|---|---|---|---|---|
| | $c'$ (kPa) | $\phi'$(°) | $R_f$ | $E_{ur}^{ref}$ (kPa) | ($E_{50}^{ref}$) (kPa) | $E_{oed}^{ref}$ (kPa) | $m$ |
| Train set | 0.052 [0.151] | 0.017 [0.073] | 0.004 [0.005] | 361.512 [557.42] | 156.208 [204.77] | 65.006 [88.22] | 0.002 [0.009] |
| Validate set | 0.085 [0.152] | 0.015 [0.071] | 0.004 [0.005] | 490.832 [575.87] | 188.001 [206.83] | 67.919 [87.16] | 0.002 [0.009] |
| Test set (Stochastic Dropout) | 0.071 [0.174] | 0.015 [0.075] | 0.005 [0.011] | 380.38 [1016.57] | 308.6 [393.42] | 48.34 [96.73] | 0.002 [0.009] |
| Test case (GEOSLOPE International Ltd., 2021) | 0.041 [0.324] | 0.011 [0.026] | 0.005 [0.019] | 361.88 [768.02] | 104.41 [1477.93] | 28.095 [357.83] | 0.002 [0.017] |

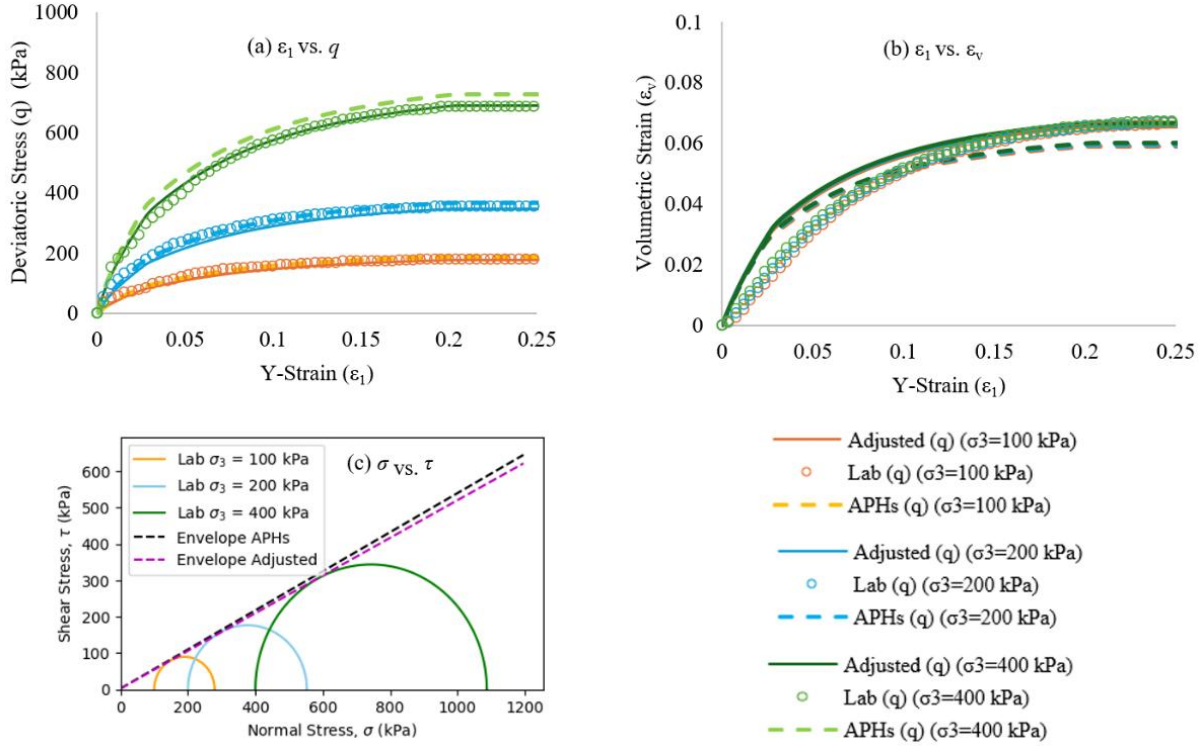



*Figure 6.5. Comparison of simulated results: (APHs) vs. lab test results (Jin et al., 2016).*

To further enhance the simulation, fine-tuning is enriched by integrating a feature importance analysis (Figure 6.4) to guide parameter adjustments. For example, deviations in volumetric strain indicate that $E_{oed}^{ref}$ needs adjusting. Remarkably, the accurate value of $E_{oed}^{ref}$ (adjusted to 1000 kPa) falls between 911 and 1337 kPa (Table 6.5) despite the training minimum being 1050 kPa. Similarly, deviations in the magnitude of deviatoric stress indicate that $\phi'$ needs adjusting (adjusted to 27.37°). The slight deviation in $\phi'$ may partly reflect laboratory bias. This is evident in the envelope plot (Figure 6.5), where the APHs envelope is tangent to the first two Mohr circles but deviates slightly from the third. Overall, APHs deliver accurate parameterizations even beyond the training range and effectively complement physically informed models. They refined the optimization search space from 10°-50° down to approximately 28°, and the feature importance in Figure 6.4 further guides model calibration.

*Table 6.5. HS model parameters for Hostun sand with 99% CI using the APHs tool.*

| Parameter | Value | | | | | | |
|---|---|---|---|---|---|---|---|
| | ($c'$) (kPa) | ($\phi'$) (deg) | ($R_f$) | ($E_{ur}^{ref}$) (kPa) | ($E_{50}^{ref}$) (kPa) | ($E_{oed}^{ref}$) (kPa) | ($m$) |
| APHs | 2.5026 | 28.2405 (27.3671)* | 0.8551 | 10692.4502 | 4591.1729 | 1124.2331 (1000)* | 0.9834 |
| 99% CI lower | 2.3297 | 28.191 | 0.8422 | 9483.7782 | 4057.8961 | 911.133 | 0.9769 |
| 99% CI upper | 2.6755 | 28.2901 | 0.8679 | 11901.1222 | 5124.4496 | 1337.3333 | 0.9899 |

Values in (brackets) are adjustments based on controlling features (Figure 6.4) and visualized deviations (Figure 6.5)
* Falls outside the trained boundaries of the APHs tool.

Analyzing soil samples beyond the trained parameter ranges while accounting for laboratory and numerical biases enables evaluation of how these errors propagate through simulation results. For Hostun sand, the absolute errors were approximately 0.87° and 0.126 MPa for $\phi'$ and $E_{oed}^{ref}$, resulting in only minor deviations in both deviatoric stress and volumetric strain. This suggests that when soil properties fall within the training range (0.015° and 0.068 MPa) indicate minimal expected deviations. However, experts should interpret

simulated results, considering the specific characteristics and sensitivity of the problem under investigation.

One of the most notable achievements of this approach is its ability to accurately estimate stiffness parameters without requiring unloading-reloading tests or oedometer data, making it an efficient tool. Notably, the parameters retrieved by the APHs tool for CDG-recompacted saturated soil were found to be as $E_{oed}^{ref} = 0.3E_{50}^{ref}$ which is found to fall outside the augmented samples for which the tool was developed and even beyond the ratios proposed in the literature. For example, Lu et al. (2018), Lee et al. (2008), and Lam (2018) suggested a ratio between 0.7 and 1.0, while GEOSLOPE International Ltd. (2021) recommended a range of 0.5 to 0.8 in cases where oedometer tests are unavailable. Comparable results were observed for Hostun sand. This demonstrates the robust performance of the APHs tool in accurately determining the HS parameters.

The results demonstrate the performance of the APHs tool, achieving an accuracy of 99.4%. Scatter plots of predicted versus actual values highlight the tool's ability to generalize effectively, with the predicted values aligning closely with the 1:1 line for both training and validation samples. The performance of the APHs tool was tested, validated, and deployed using unseen datasets to confirm its reliability and precision. The first validation reference is a soil sample previously parameterized by GEOSLOPE International Ltd. (2021). The comparison demonstrated that the APHs tool achieved an accuracy of approximately 99.6%, with minimal absolute error across all predicted parameters. The second validation case focused on a CDG-recompacted saturated soil. The simulated results using parameters retrieved from the APHs tool were favorably comparable to the original laboratory tests (Zhou, 2008). This study validates the initial hypotheses proposed at the outset. One of the key implications of this research is the development of a smart laboratory system. This work paves the way for transforming traditional laboratories into innovative laboratories, leveraging the

power of DL to enhance efficiency and accuracy. APHs can be locally installed on any device with the required libraries to deliver parameterizations within seconds.

While the results demonstrate that APHs yield accurate predictions, the reader needs to note the considerable computational power required (Table 6.6). Essentially, prediction performance relies on two factors: (a) the data's volume and quality, and (b) the DL algorithm's capability, both of which impact computational requirements during training and for end users. First, comprehensive laboratory datasets with accurately measured parameters are scarce; for instance, according to the authors' knowledge, one of the minor parametrized cases available is the one provided by GEOSLOPE. Conducting laboratory tests usually requires extensive manual parameterization, which is time-consuming, highlighting the advantage of the proposed method in handling the sheer volume of data needed. However, there remains potential to further reduce the required sample size by experimenting with different sample sizes and alternative sampling strategies such as Sobol sequences or MC methods. Secondly, it is acknowledged that laboratory tests may introduce bias regarding the dataset's quality. To mitigate this, it was ensured that 102 time steps represent each curve. However, this level of detail means that each sample (comprising 6 curves, 102-time steps per curve, and 7 parameters) demands significant computational resources during training. Therefore, exploring alternative DL architectures, such as Transformer networks and CNNs, could potentially reduce the computational burden without compromising performance. Thirdly, regarding the computational burden, the experiments (using a single model as an example) illustrate this challenge well. For instance, the first LSTM layer, which processes 7 input features, incorporates roughly 1,064,960 trainable parameters (calculated as $4 \times (7 + 512 + 1) \times 512$). Both the second and third LSTM layers require approximately 2,099,200 parameters each. In addition, a 256-unit dense layer adds about 131,328 parameters, and the final 7-unit output layer contributes another 1,799 parameters, resulting in a total of roughly 5.4 million trainable

parameters. The sequential processing across all 102 timesteps further amplifies memory and time requirements. To clarify, model tuning on the HPC server takes about 6 hours. In contrast, tuning on a regular PC (8 CPU cores, 128 GB memory, with no GPU) could extend the duration to several days; for example, 20 iterations with 250 epochs required approximately 2 days, reflecting the substantial computational resources required.

*Table 6.6. Summary of APHs current achievements and future direction.*

| Aspect | Current setup | Current achievement | Future directions | Target improvement |
|---|---|---|---|---|
| Dataset Volume | A comprehensive dataset of 1300 samples | - High accuracy | Exploring alternative sampling techniques (e.g., Sobol sequences, MC) | Reduce dataset size while maintaining high accuracy for further efficiency. |
| Dataset Quality | 102 timesteps represent each curve | - High accuracy<br>- Mitigate lab data bias | Explore Transformers /CNNs | Achieve faster training with fewer trainable parameters while capturing long-term behavior. |
| Dataset generation | Based on 2D-axisymmetric simulation assumptions | - Testing on real-world samples confirms the validity of the 2D-axisymmetric simulation assumptions. | Consider 3D simulations | Enhance realism and robustness with further training as needed |
| Tool versatility | Tolerated for Hong Kong soil, as outlined in the Boundary Condition Section | - High accuracy<br>- User interface<br>- Fast parametrizations<br>- Integrable with any finite element models or physically informed tools | Incorporate anisotropy, dynamic loading conditions, other constitutive models, and wider parameter ranges. | Broaden applicability and versatility to different soils and conditions. |

## 6.5. AUTOMATED STABILITY ANALYSIS TOOL

Figure 6.6 illustrates the deployment of the automated slope-assessment system. The system architecture begins with real-time rainfall data acquisition from Hong Kong's network of raingauges, processing precipitation sequences through a dual-model machine learning framework. The CNN and LSTM network processes temporal rainfall sequences to capture hysteresis effects and antecedent moisture conditions. Both models output predicted soil saturation ratios (SSI) that serve as critical inputs to the physics-based slope stability model. The architecture is designed to integrate with APHs, given the availability of raw laboratory data, after which the tool automatically predicts soil saturation and subsequently computes the

FOS. The geotechnical component employs an infinite slope stability formulation, calculating FOS through Equation 3.31. The automated FOS calculation is constrained by the model assumptions summarized in Section 3. The deployment includes real-time visualization through an interactive web dashboard that displays time-series FOS projections, rainfall patterns, and safety status indicators, enabling transportation authorities and geotechnical professionals to implement proactive measures when FOS trends approach critical thresholds below 1.2. This operational framework represents a significant advancement from conventional rainfall threshold methods by providing mechanistically based, site-specific stability assessments that account for both current conditions and forecasted rainfall impacts.

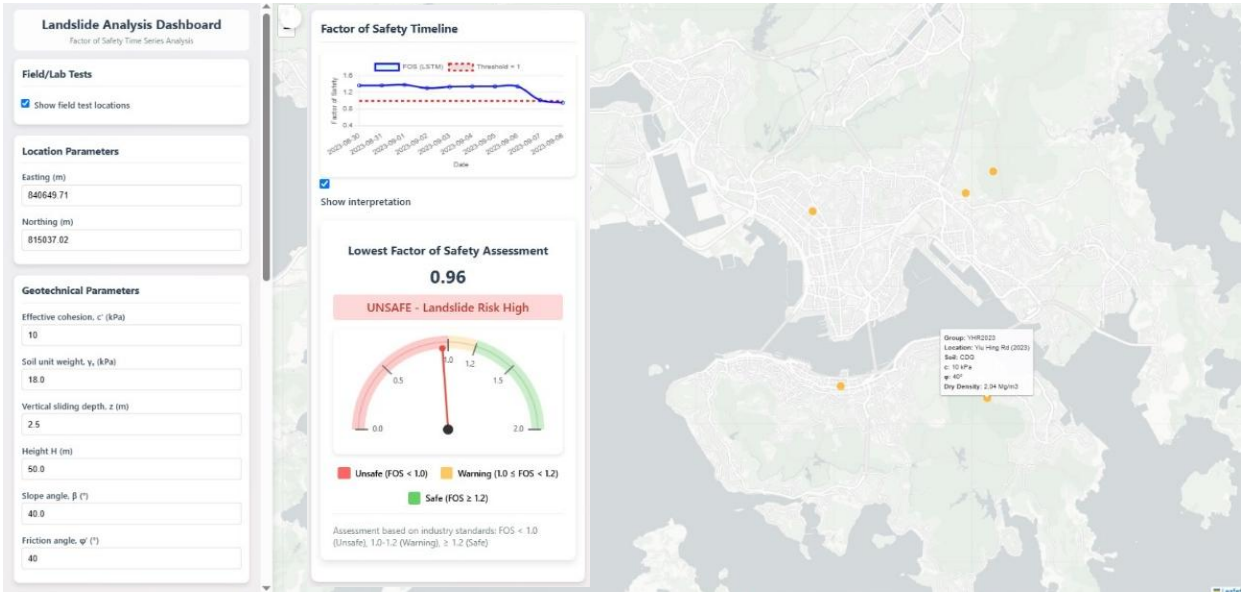



*Figure 6.6. The automated slope stability tool.*

This study demonstrates that the integration of data-driven approaches with physically based equations produces models that achieve both statistical accuracy and geotechnical consistency. The proposed framework provides a reliable foundation for landslide early warning systems, while the identified limitations and research directions outline a clear pathway for continued improvement in landslide susceptibility assessment. This study introduces several key methodological innovations that collectively enhance landslide predictions: (a) development of a spatiotemporally imbalanced dataset; (b) integration of real

rainfall patterns; (c) incorporation of multiple spatial features that modulate rainfall impact across different terrain configurations; (d) implementation of a coarse-to-fine architecture tuning approach; and (e) establishment of an integrated framework coupling DL-based Soil Saturation Index (SSI) predictions with physics-based FOS calculations.

The comparative assessment of LSTM and CNN architectures reveals distinct operational characteristics (Table 6.7). While both models achieve high precision, the CNN demonstrates superior computational efficiency (12 hours vs 23 hours tuning time) and marginally better predictive performance (74% vs 69% precision). Uncertainty quantification via Monte Carlo dropout revealed moderate and comparable stochastic variations (LSTM: 6%, CNN: 7%). The CNN model strikes an optimal balance between accuracy, efficiency, and operational reliability, though both models are integrated into the decision-support interface to accommodate scenario-specific requirements.

*Table 6.7. Comparative assessment of data-driven models.*

| Model | Performance (Precision) | Computational power (Tuning) | Uncertainties (St. deviation) |
|---|---|---|---|
| LSTM | High (69%) | High (23 hr) | Moderate (6%) |
| CNN | High (74%) | Moderate (12 hr) | Moderate (7%) |

To emphasize the developed model's reliability, it was tested on a recent massive landslide outside the training dataset, which is the Yiu Hing Road landslide (Figure 6.7). The model calculated an FOS of 0.96, which aligns closely with the professionally assessed value of 0.97 from the forensic investigation (Fugro (Hong Kong) Limited, 2025) (Figures 6.6 and 6.7). This independent validation across a different location and temporal context demonstrates the framework's transferability and readiness for practical application. This multi-faceted validation strategy confirms that the hybrid framework not only achieves high statistical accuracy but also maintains geotechnical consistency with physics-based models and operational reliability against professional benchmarks. The validation against two independent case studies (Yiu Hing Road and Fei Ngo Shan), encompassing 106 distinct

rainfall events, underscores the model's robustness for regional early warning. Comparative benchmarking against established studies reveals competitive performance, with this study achieving AUC (0.89) and balanced accuracy (0.819) values comparable to Mondini et al. (2023), while explicitly addressing temporal and spatial data imbalances as emphasized by Xiao et al. (2023).

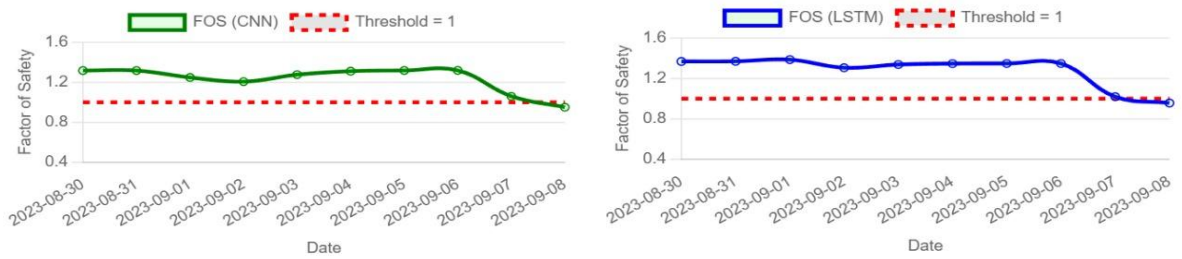



*Figure 6.7. The predicted FOS of the Yiu Hing landslide using the proposed models.*

Beyond conventional performance metrics such as AUC, the operational value of the proposed framework is demonstrated through its ability to provide timely and actionable warnings. Referring to Figure 6.1.e for the Yiu Hing Road landslide, Figure 6.8 presents a detailed view of the Tier 1 single-event predictions leading up to the failure. It is seen that no warning was issued prior to 7 September 2023, consistent with stable slope conditions. A warning was triggered on 7 September 2023, aligning closely with the FOS trajectories presented in Figure 6.7, which indicate a decline at the same time. This concordance between Tier 1 predictions and Tiers 2 and 3 builds trust and provides decision-makers with actionable, risk-informed information. The warning serves as a clear signal to consider proactive measures, such as road closures, slope maintenance, or public alerts, based on the assessed level of risk.

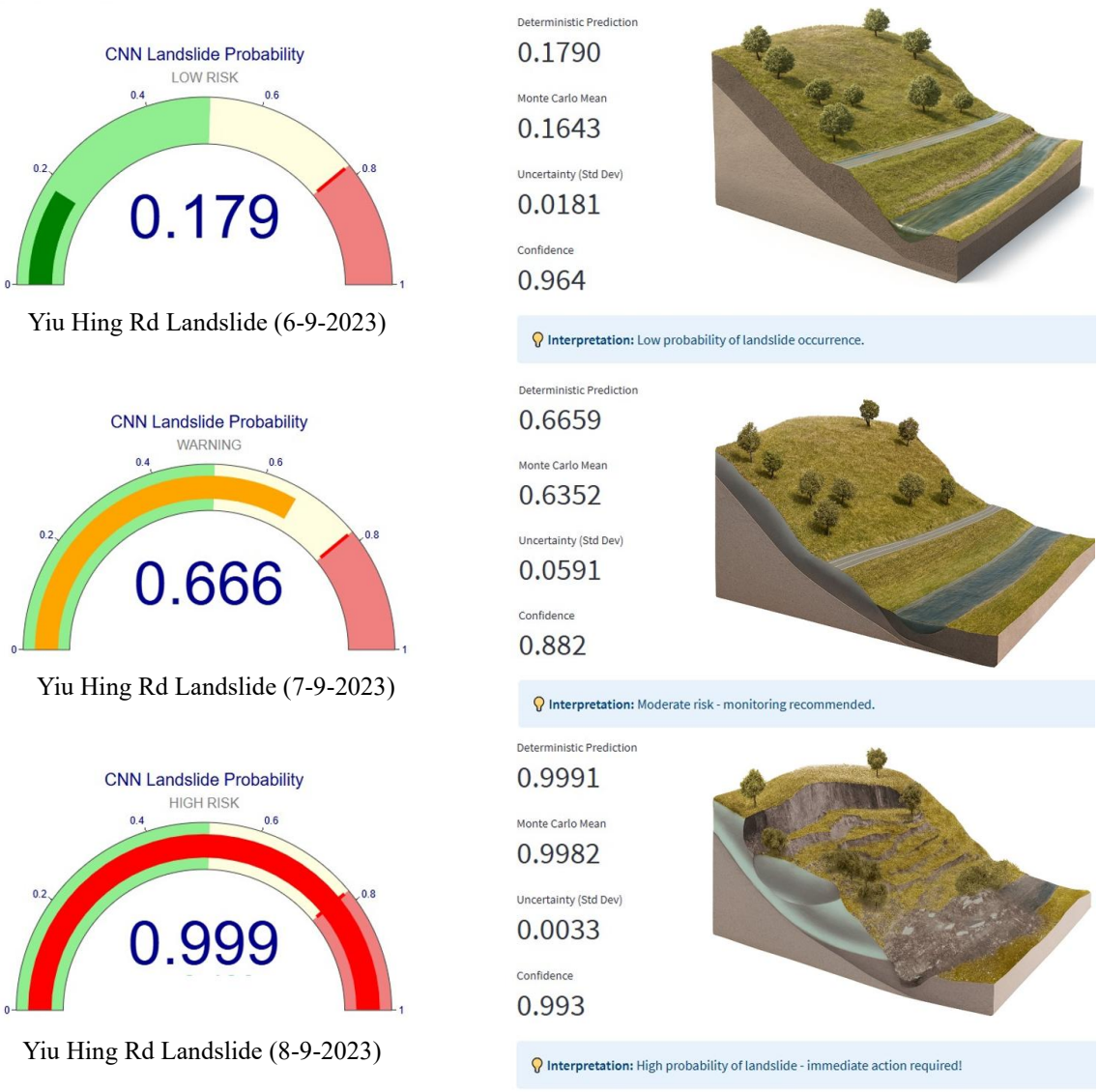



*Figure 6.8. Single-event predictions for the Yiu Hing Road landslide between 6–8 September 2023. Illustrations adapted from PBS LearningMedia*

The primary limitations stem from hydrological monitoring constraints and temporal uncertainties (Table 6.8). The assumption of binary saturation states (fully saturated during failures vs. dry conditions otherwise) represents a necessary simplification given the current absence of comprehensive regional pore-water pressure monitoring networks. Temporal uncertainties in exact failure timing and spatial disparities between landslide locations and rain gauges introduce additional complexities. While mitigated through extended antecedent durations and spatial distance features, these limitations highlight the need for enhanced monitoring infrastructure. Future research should prioritize the development of distributed

sensor networks for real-time pore-water pressure monitoring and high-resolution temporal characterization of failure mechanisms.

*Table 6.8. Current achievements, limitations, and future research directions.*

| Aspect | Current Achievement | Limitation | Future Direction |
|---|---|---|---|
| Hydrological Assumptions | Implemented saturation-dry binary states based on available data | Simplified saturation assumptions due to limited pore-water pressure monitoring | Integrate distributed pore-water pressure sensors and soil moisture monitoring networks |
| Temporal Precision | Incorporated antecedent rainfall patterns with 28-day sequences | Uncertainty in exact failure timing | Develop probabilistic timing models and implement high-resolution temporal monitoring. |
| Model Architecture | Developed and validated both LSTM and CNN architectures | Limited exploration of hybrid architectures and transfer learning | Investigate transformer-based models and cross-regional transfer learning approaches |

## 6.6. SUMMARY

This chapter describes the development and deployment of four web-based applications created to operationalize the predictive models presented in this study. The first three tools were implemented as Streamlit web applications; the fourth was developed as an HTML-based interface. The section enumerates limitations and planned improvements for each tool and specifies the input data required by every web interface. Screenshots and descriptions illustrate the user-friendly design and interaction workflow for each application.

# CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

## 7.1. INTRODUCTION

This research aims to develop an effective, scalable, operational, and cost-effective system to mitigate landslide risk in Hong Kong by integrating multiple technologies. Four core objectives guided the work: understand the factors and mechanisms of landslide initiation and critically review existing monitoring and prediction methods; design a spatiotemporal model for shallow, rainfall-induced landslides using historical records and data-driven techniques to capture triggering randomness; establish an integrated prediction framework that fuses subsurface hydrological and mechanical responses to improve early-warning accuracy; and streamline and automate the stability-analysis workflow, including parametrization and FOS computation. The methodology, datasets, model development, testing results, and automated tools created to support implementation are described in the preceding chapters. This concluding chapter distills the main findings, highlights the study's significance and contributions, acknowledges limitations, and presents recommendations for future work.

## 7.2. SUMMARY OF THE KEY FINDINGS

### *7.2.1. Mechanisms, Predicting Approaches, and Monitoring Systems*

The most effective strategy for shallow landslide prediction integrates real patterns of rainfall-driven triggers, internal slope responses, and multi-scale monitoring with hybrid physical–AI models tailored to data availability and site conditions. This is underscored as follows: Shallow landslides are primarily triggered by rainfall, where both intensity–duration and antecedent moisture conditions govern initiation, while internal responses involve changes in displacement, soil moisture, suction stress, groundwater level, and effective stress. Displacement frequently exhibits superposed trend and periodic components driven by hydrologic forcing, seasonality, and hydro-fluctuation effects. Empirical rainfall thresholds remain practical when physical inputs are scarce: I–D thresholds can provide reasonable

accuracy but suffer high false positives, whereas API thresholds better emulate initial moisture, improving accuracy and reducing false alarms when derived via regression or probabilistic analysis. Physically based causative thresholds directly link rainfall to internal state variables; displacement thresholds are most common due to data availability, while thresholds tied to suction, moisture, or groundwater require site-specific geotechnical and hydrologic inputs. Regression models struggle with sudden surges, single AI models improve accuracy but miss periodicity, and hybrid approaches that decompose displacement into trend (polynomial/simple AI) and periodic components (hybrid/deep models) perform best, though they remain constrained by training-domain generalization and simplified initial conditions.

Deterministic physical models deliver higher fidelity by explicitly representing initial and boundary conditions, infiltration, runoff/evaporation, soil properties, and both saturated and unsaturated states. Analytical frameworks can incorporate steady/transient flow, vegetation, weathering/degradation, and neighbor-grid effects, with recent evidence highlighting the accuracy of TAG-FLOW. Numerical models reproduce real rainfall patterns with time lags, complex 3D geometries, and bedrock–soil systems under hydro-fluctuation. These physical models can be combined with probabilistic analysis to create susceptibility maps and fused with AI to enhance early warning, but they require detailed and costly site characterization and are best suited to local scales where spatial inputs are available. For susceptibility mapping, qualitative (expert-based) methods are being supplanted by quantitative approaches, where linear bivariate/multivariate models are limited in capturing complex interactions, and hybrid AI models offer better handling of nonlinear, spatiotemporal relationships, albeit with strong dependence on input data quality and selection.

Monitoring systems must bridge surface indicators and subsurface mechanisms to enable reliable early warning. Surface techniques provide millimeter-level 3D positions and broad spatial coverage but face challenges with real-time acquisition, coarse temporal

resolution, and environmental interference, making them more suitable for susceptibility, risk, and vulnerability mapping than for alerts. Subsurface methods provide higher diagnostic value: extensometers offer high temporal precision for translational motion; inclinometers detect subsurface displacement but have limited vertical resolution, while optical-fiber solutions (FBG/BOTDA) improve spatial resolution and couple strain–temperature at lower cost but are wiring-limited; TDR pinpoints sliding surfaces at fine resolution though displacement quantification is difficult; AE is sensitive to micro-deformation and suits slow movements; tilt sensors provide accurate angular change and low-cost modular arrays but miss pure translational motion; strain-gauge systems (SDS, SG-DEP) capture micro-displacement with 360° directionality; and accelerometers detect rapid translational motion independent of triggers. Hydrologic and thermal monitoring strengthens causative insight: novel self-powered rainfall sensors expand range and resolution; soil moisture can be captured via FBG, UHF RFID, smart aggregates, and ERT for plot-to-regional scales; and temperature in shear bands and at the surface offers stability indicators for shallow strata. MEMS/IoT nodes integrated into WSNs reduce wiring and cost, suiting shallow landslides, though networks must address power, communication range, and data integrity via sleep thresholds, rechargeable power, optimized density, and buried magnetic induction links. Deployment strategies (random, matrix, vulnerable, hybrid) should prioritize vulnerable placement at critical points for maximal value. Emerging early warning signals from elastic waves and infrasound show promise but remain challenging to operationalize, with novel triboelectric and chemiluminescent sensors under development.

Time series modeling emphasizes that data frequency and temporal ordering critically shape predictions. Monthly series typically require decomposition into trend and seasonal components, while daily series may lack seasonality and demand tailored preprocessing; lag selection should reflect site hydraulics. Dynamic deep models (e.g., LSTM) outperform static

ANN, SVM, RF, and ARIMA by capturing nonlinear temporal dependencies, provided that careful data preparation (frequency choice, sampling ratio, decomposition, temporal correlations, and feature selection) is performed. Random search for hyperparameters, Adam optimization, and MSE/Huber losses are effective, with overfitting mitigated by targeted feature sets, monitored training, regularization, and judicious activation choices. Evaluation predominantly uses RMSE, MAE, and MAPE, but weighted metrics and interval/multi-step forecasts are recommended to reflect the criticality of extreme events and to broaden beyond single-step displacement prediction.

### *7.2.2. Landslide Prediction Model*

This research has successfully developed and validated an operational, multi-data-driven framework for predicting shallow rainfall-induced landslides, demonstrating that DL models, specifically LSTM, GRU, and CNN, can achieve high-performance benchmarks with AUCs exceeding 90%, outperforming traditional LR and prior studies. The superior performance is attributed to a meticulously crafted methodology centered on a time-dependent dataset design that explicitly captures rainfall-triggering randomness and essential spatiotemporal features. This foundational approach was bolstered by rigorous data preparation to correct inherent spatial-temporal imbalances, a comprehensive coarse-to-fine hyperparameter tuning strategy, and the novel introduction of a dual-threshold decision scheme. Validation against the 2005 Fei Ngo Shan case history confirms the model's robustness and reliability, establishing it as a practical, standalone decision-support tool that effectively integrates probabilistic forecasting with deterministic seepage stability for verification.

The investigation yielded critical insights into model behavior and operational practicality. Analysis revealed that rainfall-related predictors were the most influential, accounting for approximately 36% of the total feature importance, with short-term lags, two-

day antecedent rainfall, and rainfall pattern features being key contributors. Furthermore, the inclusion of spatiotemporal descriptors was shown to increase model precision by 7%, underscoring the necessity of moving beyond rainfall-only inputs to accurately regulate slope response. From an operational standpoint, the CNN emerged as the most efficient architecture, achieving comparable precision (~78%) to the more complex LSTM and GRU models but with substantially lower computational cost, tuning time, and predictive uncertainty, while also producing a hydrological response more consistent with physical models. The proposed dual-threshold method proved to be a significant enhancement over a conventional 0.5 cut-off, providing a more nuanced risk assessment, reducing false decisions, and delivering a narrower, more actionable warning zone for end-users. Although calibrated for Hong Kong's colluvium, the underlying framework is designed for transferability to other regions, contingent upon the integration of local data and appropriate model retuning.

### *7.2.3. Subsurface Sensor-Based Predictions*

#### a) VWC Prediction Operational Model (OWHK)

This research has established a robust operational framework for forecasting VWC in colluvial slopes, effectively addressing the complex challenge of quantifying infiltrated rainfall as a critical proxy for stability assessment. The core innovation lies in the integration of independence-oriented time series windowing with a multi-parallel LSTM network architecture and deterministic seepage modeling, a synergistic approach designed to capture rainfall-triggering randomness and overcome inherent spatiotemporal data limitations. The model's physical grounding is reflected in its feature design, where boundary conditions, represented by initial VWC, act as proxies for unmonitored environmental factors, while comprehensive rainfall descriptors drive the dynamic hydrological response. This methodology was rigorously validated against a substantial dataset comprising 15 in-situ sensors across three

sites, augmented by a numerically validated case and supported by field samples, which collectively corroborated the initial hypotheses and demonstrated the model's ability to learn distinct hydrologic regimes.

The operational outcome of this research is the OWHK tool, which delivers strong predictive skill with a mean absolute error of less than 0.6% and an $R^2$ exceeding 0.92, alongside quantified uncertainty through Monte Carlo dropout. A key contribution is its provision of one of the first systematic comparisons between an operational, regional model and site-specific analyses. The specialized design of the LSTM networks, each mapped to a specific segment of the soil-water retention curve and trained on data grouped by boundary conditions, proved effective in reducing bias from missing variables and enhancing model adaptability. However, the model's performance, while robust within the observed data domain, highlights a critical dependency on data comprehensiveness, as accuracy diminishes during rapid transitional soil states due to limited training samples for these events. This underscores the necessity for future work to expand datasets, particularly for these dry-to-saturated transitions, and to further enrich model inputs with data on vegetation and detailed soil properties to ensure sustained operational reliability and broader generalization.

### b) Hypertuning Strategy for Enhanced Model Performance

This research establishes a comprehensive hypertuning strategy that successfully couples DL with time series forecasting to analyze subsurface VWC dynamics, providing a robust and practical framework for operational landslide risk decision-making. The investigation demonstrates that while the selection of a sophisticated architecture like LSTM is important, the rigor of the tuning process and the quality of the feature selection are paramount. The strategy, which employed both statistical and knowledge-based techniques for feature engineering, revealed critical insights, including significant correlations between spatially variable responses and cumulative antecedent rainfall, and identified distinct optimal

antecedent periods of 10 and 20 days for shallow and deeper soil layers, respectively. Furthermore, the analysis established that allocating 50% to 70% of the data for training is a key determinant of superior model performance. A central finding corroborates the "No Free Lunch" theorem, indicating that with limited data, DL offers no distinct advantage over simpler machine learning models, thereby underscoring that meticulous tuning and high-quality data are more critical than algorithmic complexity alone.

The practical application and validation of this tuned framework yielded strong and transferable results, with models achieving high predictive accuracy (MAE 0.35%–1.25%) and reliability (PICP 0.86–0.95) across multiple landslide sites, including Pa Mei and Tung Chung. The propagation of these predicted VWC values into FOS calculations further confirmed the model's operational utility, producing robust metrics ($R^2$ = 0.89, PICP = 0.85) for stability assessment. Beyond direct prediction, the study proposes an innovative, collaborative approach to enhance physical monitoring systems, demonstrating that data from a network of sensors can be leveraged to forecast the status of any individual sensor, thereby mitigating pervasive challenges related to power consumption and data integrity. Ultimately, this work affirms that while data-driven mining with advanced ML algorithms is a powerful predictive tool, its efficacy and credibility are contingent upon its consistency with fundamental geological theory and its role as a complement to, not a replacement for, physical understanding.

### c) Distributed Multi-Sensor (Subsurface) Prediction Model

This research has established a robust DL framework that fundamentally advances the prediction of subsurface processes leading to rainfall-induced landslides by demonstrating that accurate forecasting necessitates modeling the complex, multivariate interaction between rainfall and the slope's antecedent hydrological state. The core innovation of the tailored LSTM-based approach lies in its successful capture of this critical spatiotemporal relationship, moving beyond traditional models that treat rainfall as a sole trigger by integrating historical

sequences from spatially distributed sensors. This enables the model to learn that failure is preconditioned not by a simple rainfall total, but by a specific hydrological input interacting with a slope state defined by its internal hydromechanical boundary conditions. The framework's effectiveness is empirically validated by its consistent and high predictive accuracy across all key subsurface variables: it accurately captured the build-up of suction, a direct stability indicator, with high precision (RMSE of 1.72 kPa, $R^2$ = 0.93); significantly outperformed a naïve model in predicting matric suction (RMSE of 5.40 kPa); and demonstrated a dramatic improvement in VWC prediction using multivariate data (RMSE of 3.26%) over rainfall-alone inputs (RMSE of 6.23%). Beyond these metrics, the model's strong physical consistency, evidenced by its accurate replication of transient wetting front dynamics, temporal lags at different depths, and distinct seasonal responses, confirmed through successful generalization across independent case studies, underscores its operational robustness.

### *7.2.4. Slope Stability Analysis and Process Automation*

#### a) Automated parametrization tool

This research introduces a transformative approach to soil parameterization by developing the APHs tool, leveraging the power of LSTM networks to achieve unprecedented accuracy and efficiency in estimating critical strength and stiffness parameters ($c'$, $\phi'$, $R_f$, $E_{oed}^{ref}$, $E_{50}^{ref}$, $E_{ur}^{ref}$, $m$). Key findings include: With a model accuracy of 99.4%, the APHs tool demonstrates precision in predicting parameters solely from triaxial test data, making the process more accessible and cost-effective. Validation cases affirm the tool's reliability, with predicted parameters aligning with laboratory results. Notably, for CDG-recompacted soil, the tool accurately retrieved parameters even when they fell outside the range of ratios traditionally

proposed in the literature, showcasing that this tool not only accelerates the HS model parameterization process but also simplifies it while maintaining high precision.

This outstanding performance stems from two key advancements. Firstly, the study employed advanced data generation techniques using a combination of the LHS and SIGMA/W simulations to address the limitations of insufficient experimental data. This approach created a robust and comprehensive parameter space tailored to Hong Kong's unique geological conditions, ensuring the tool's practical applicability. Secondly, the tool's innovative multi-model architecture enabled the assignment of distinct weights to each parameter, effectively mimicking the physical processes involved in parameter derivation and ensuring robustness across varying scales. By framing stress-strain relationships as a time series problem, the tool leveraged the sequential nature of triaxial test results to enable precise and consistent parameter estimation. This research establishes a paradigm shift in geotechnical engineering, paving the way for smart laboratory systems that integrate DL to optimize and automate traditionally manual processes. While current findings validate the tool's performance under defined conditions, future work will focus on expanding its parameter space to accommodate a broader spectrum of soil types and boundary conditions.

### b) Automated Slope Stability Analysis Model (A Hybrid DL–Physics Model)

This research successfully establishes a regional, operational landslide early-warning framework that effectively bridges the gap between data-driven learning and physical modeling, demonstrating that a hybrid approach can overcome the traditional trade-off between physical rigor and data availability. By training DL models to learn saturation dynamics from real, temporally random rainfall sequences and embedding these outputs within an infinite slope stability equation, the framework captures the essential time-dependent mechanisms of rainfall-induced failures while maintaining computational tractability across Hong Kong's extensive natural terrain. The system delivers reliable predictive skill with robust

generalization, as evidenced by CNN and LSTM models achieving balanced accuracy of approximately 0.82, precision between 69-74%, and AUC values ranging from 0.87 to 0.91 across all dataset splits, while maintaining close physical agreement with deterministic seepage-stability analyses.

The framework's validity is further confirmed through multiple verification pathways, including strong alignment with the documented 2005 Fei Nno Shan failure case (MAE in FOS 0.06-0.07), where it correctly identified the failure timing and reflected post-event drainage hysteresis, and successful external testing on the 2023 Yiu Hing Road landslide, demonstrating consistent alignment with professionally assessed stability conditions. Mechanistic interpretability analyses substantiate the model's physical consistency, confirming rainfall as the dominant triggering mechanism while identifying how topographic features modulate susceptibility, with CNN's spatial filters and LSTM's sequential memory providing complementary strengths. Ultimately, the research delivers an operational system that integrates 28-day antecedent rainfall sequences from 96 gauges to output real-time, SSI-informed FOS trajectories, supporting actionable thresholds for proactive risk management and demonstrating that coupling DL with physics-based stability can provide regionally scalable, time-aware warnings without requiring exhaustive site-specific geotechnical characterization.

### c) Factors Controlling Rainfall-Induced Landslides: A Parametric Study

This research provides critical new insights derived from physical deterministic analysis to enhance the accuracy of landslide early warning systems by rigorously analyzing the factors governing the triggering randomness of rainfall-induced landslides, with a specific focus on the typical slope conditions of Hong Kong. The analysis demonstrates that relying solely on cumulative rainfall, as is common in empirical statistical thresholds, can be fundamentally misleading, as it ignores the decisive role of temporal rainfall patterns. The study establishes that for an identical cumulative rainfall total, distinct patterns, specifically delayed versus

advanced rainfall, produce significantly different hydrological responses and failure timings. A delayed pattern allows adequate time for water seepage, while an advanced pattern intensifies pore-pressure buildup, thereby delaying the time to failure. This finding challenges the foundation of many statistical approaches and underscores the necessity of incorporating temporal rainfall characteristics into predictive models.

Furthermore, the investigation quantitatively clarifies the complex dynamics controlling a slope's recovery time, or the effective non-rainfall duration required for stabilization post-rainfall. It is demonstrated that this healing period is not a fixed value but is primarily controlled by the soil's hydraulic conductivity, with higher conductivity leading to faster drainage and shorter recovery. However, this process is significantly modulated by initial soil suction conditions, rainfall duration, and intensity; smaller initial suction, longer rainfall duration, and heavier rainfall all contribute to a prolonged recovery time. Consequently, this study provides a principled basis for defining the critical antecedent rainfall period, concluding that for Hong Kong's colluvium, a variable window of at least 15 days, extending beyond 32.5 days depending on specific topographic, hydrological, mechanical, and rainfall characteristics, must be considered to accurately account for the slope's antecedent state, thereby offering a physically-grounded method to reduce false positives in early-warning systems.

### d) Geometry effects on rainfall-induced landslides: A Parametric Study

This research establishes that subsurface flow geometry, specifically the inflow-to-outflow area ratio (R/r), exerts a first-order control on regional vulnerability to rainfall-induced landslides, providing a mechanistic explanation for observed spatial patterns of failure. The analysis reveals two distinct topographic regimes: convergent hillslopes, characterized by high R/r ratios, act to focus infiltrating water and lateral seepage into confined zones, which promotes rapid pore-pressure accumulation and establishes an antecedent-dominated regime where residual moisture critically accelerates failure onset even under modest rainfall. In

contrast, divergent hillslopes distribute hydrological inflow across larger outflow areas, attenuating groundwater development and creating an intensity-dominated regime where failure potential scales more directly with rainfall rate. This fundamental geometric control not only clarifies why landslides cluster in specific topographies under ostensibly similar conditions but also directly informs practical hazard assessment. By incorporating quantitative geometric descriptors like R/r into regional models, predictive skill is significantly improved, enabling the prioritization of monitoring and mitigation efforts in the most climate-sensitive terrain and ultimately supporting more resilient land-use planning and infrastructure design in the face of intensifying precipitation extremes.

### e) Numerical Slope Stability Modeling (Focused Site Specific Analysis)

This research provides a detailed numerical reconstruction of the 2005 Fei Ngo Shan landslide, delivering critical, case-specific insights into the prediction and monitoring of shallow failures in Hong Kong's colluvium through a comprehensive analysis combining non-deformable, small-deformation, and large-deformation approaches. The investigation yields three fundamental findings that clarify both the temporal and spatial characteristics of slope failure: first, it demonstrates that while the FOS remains largely invariant for a given cumulative rainfall total, the temporal resolution and pattern of rainfall data exert a primary control on the timing of the predicted slope response, underscoring that high-frequency rainfall inputs are essential for accurate temporal forecasting without materially altering the ultimate stability assessment. Second, the analysis enables precise local vulnerability mapping, identifying the BH3 vicinity as the critical monitoring location due to its conjunction of the thinnest colluvium layer, its position at the slope toe where both lateral inflow and direct infiltration converge, and its characteristic groundwater lag response, thereby providing a scientifically-grounded rationale for prioritizing instrumentation.

The third key finding elucidates the specific failure mechanism, confirming that shallow landslides in this context are toe-initiated, with large deformations first developing at the toe that precipitate a loss of support and subsequent retrogressive mobilization of the upslope shallow layer, while the soil along the rupture surface itself exhibits only minor sliding. Collectively, these results offer an integrated understanding of how rainfall temporal structure governs response timing, where monitoring resources should be concentrated to detect early instability, and the precise mechanical sequence by which shallow collapses propagate, thereby providing directly actionable guidance for optimizing operational monitoring layouts and refining numerical model configurations for more effective early warning systems.

### *7.2.5. Synthesis of Key Findings and Conclusions*

The following overarching conclusions, which synthesize the key contributions of this research, are presented in Table A16 in more detail as bullet points:

1. A tiered AI-physics framework is essential: A tiered structure integrating data-driven AI with physics-based modeling is essential for achieving both regional scalability and geotechnical accuracy in operational early warning systems.
2. DL captures spatiotemporal complexity: DL models, particularly LSTM, GRU, and 1D CNN, can capture the complex spatiotemporal dependencies governing landslide initiation when trained on appropriately designed datasets. The 1D CNN offers the best balance between predictive accuracy and computational efficiency for operational deployment.
3. Dual-threshold classification enhances operational utility: Moving beyond a conventional 0.5 probability cutoff, the dual-threshold framework provides nuanced risk assessment by creating an actionable "warning zone" that flags ambiguous cases for priority investigation, significantly reducing false decisions.
4. Rainfall is more than cumulative totals: Rainfall is not a single event, but a random process defined by intensity, duration, pattern, and dry spells. Models that treat rainfall as cumulative totals alone are fundamentally limited. The "effective non-rainfall duration" provides a crucial metric for quantifying slope recovery and should be incorporated into predictive frameworks.
5. Multi-parallel architecture enhances subsurface prediction: For subsurface forecasting, multi-parallel LSTM structures with specialized networks targeting distinct hydrological regimes prove more effective than single-model approaches, successfully capturing both temporal dynamics and spatial heterogeneity across monitoring networks.

6. VWC emerges as the most consistent predictive feature: VWC proves to be the most reliable predictive feature, capturing infiltration dynamics and mitigating rainfall spatiotemporal uncertainty.
7. Geometry and spatial features control infiltration and landslide vulnerability: This research demonstrates that hillslope geometry and spatial features exert a first-order control on landslide susceptibility by governing how much rainfall infiltrates and where water accumulates.
8. Automated parameterization achieves laboratory-grade accuracy: The APHs tool demonstrates that DL can automate soil parameterization with 99.4% accuracy using only standard triaxial test data, transforming what once required specialized tests and expert judgment into a rapid, accessible workflow.
9. The automated stability tool enables time-aware assessment: A hybrid pathway fusing data-driven Soil Saturation Index with physics-based FOS calculations enables successful regional-scale, time-aware stability assessment.

## 7.3. SIGNIFICANCE AND CONTRIBUTION OF THE RESEARCH

This research makes significant strides in the field of landslide forecasting by moving beyond traditional methods and creating an integrated, data-driven framework. The contributions are categorized into theoretical, practical, and broader impacts as follows:

### *7.3.1. Theoretical Contributions*

1. A Novel Multi-Framework for Landslide Prediction: This study establishes a new paradigm by synergistically combining data-driven DL models with physical deterministic analysis. It demonstrates that DL models (LSTM, GRU, CNN) can effectively capture the inherent triggering randomness of rainfall-induced landslides, a factor often oversimplified in purely physical or statistical approaches.
2. An Integrated, Automated Stability Analysis for Regional Early Warning: This study operationalizes a physics-informed, data-driven framework by learning saturation dynamics from temporally random rainfall sequences and embedding them within the infinite-slope stability model, demonstrating physical rigor with data limitations to provide scalable, operational early warning across Hong Kong's natural hillsides.
3. A Novel Operational Forecasting Framework for Regional Landslide Prediction: This study introduces a novel operational forecasting framework that prioritizes regional, cross-site predictions over conventional site-specific models by explicitly accounting for spatiotemporal limitations and rainfall-triggering randomness, thereby delivering scalable predictive performance across diverse terrain while maintaining computational efficiency.
4. Advanced Understanding of Spatiotemporal Controls: The research provides new insights into the critical factors controlling landslide triggers. It quantitatively demonstrates that rainfall patterns, not just cumulative rainfall, are a primary control on temporal prediction. Furthermore, it elucidates the role of hillslope geometry (specifically the inflow-to-outflow area ratio, R/r) as a first-order control on pore-pressure buildup, explaining spatial

clustering and defining two distinct failure mechanisms: the intensity effect (diverging slopes) and the antecedent effect (converging slopes).

5. Innovative Data Preparation and Model Development Techniques: The study introduces several novel methodologies, including: a) independence-oriented time series windowing for robust sensor data forecasting, which mitigates chronological bias; b) a dual-threshold approach for risk classification, proving that a conventional 0.5 cut-off is inadequate for capturing nuanced risk gradations; and c) a comprehensive coarse-to-fine model tuning strategy with weighted loss monitoring tailored for geotechnical data imbalances; d) novel multi-model DL frameworks that replace traditional ad hoc weighting with data-driven importance learning, leveraging multi-parallel architectures to achieve higher accuracy, improved robustness, and better generalization across sites and conditions.

### *7.3.2. Practical Contributions, Cost-Benefit Considerations, and Policy Integration*

1. Development of Operational, User-Friendly Tools: This research has translated complex models into practical decision-support systems. Key tools developed include: a) A spatiotemporal landslide forecasting tool for Hong Kong, achieving over 90% AUC, which serves as a standalone early-warning system; b) OWHK: An Operational VWC Forecasting Model, a user-friendly tool that enables efficient prediction of a key landslide precursor metric; c) APHs: An Automated Parameterization Tool, which uses LSTM to achieve 99.4% accuracy in estimating soil parameters from triaxial test data, drastically reducing the time and expertise required for geotechnical modeling; d) An automated stability tool provides time-resolved, regional forecasts with balanced accuracy (~0.82), strong discriminative performance (AUC 0.87–0.91), and close agreement with deterministic seepage–stability analysis (FOS MAE 0.06–0.07).
2. Smart Sensing for Energy-Efficient Monitoring: The rainfall thresholds identified, 177 mm (7-day), 270 mm (14-day), 341 mm (21-day), 424 mm (28-day) cumulative rainfall, 146 mm/day/month maximum intensity, and 7 rainy days per month, enable intelligent sensor scheduling. Sensors activate only when thresholds are approached or exceeded, with data frequency reduced during quiescent periods. This approach directly reduces power consumption, extends sensor lifespan, and lowers maintenance costs.
3. Data Loss Mitigation and Virtual Sensing: Subsurface monitoring systems represent a significant capital investment. The operational prediction framework mitigates data loss through: a) Continuous model-based estimates when sensor data is temporarily unavailable, preserving monitoring investments; and b) Virtual sensing capabilities where model-based estimates substitute for direct measurements, enabling strategic switching between physical sensors to conserve power while maintaining continuous coverage.

4. Guidelines for Effective Monitoring: The research provides actionable insights, including optimal antecedent periods (15–32.5 days for Hong Kong), vulnerable locations for sensor placement (slope toes with thin colluvium), and geometry-informed strategies recognizing that convergent (antecedent-dominated) and divergent (intensity-dominated) slopes require different monitoring approaches.
5. Rapid Assessment vs. Conventional Investigation: The operational tools deliver stability assessments in seconds, offering dramatic cost advantages. The 2023 Yiu Hing Road landslide was predicted in seconds, whereas subsequent investigation required thousands of dollars and months of analysis. These tools provide rapid, low-cost screening to prioritize sites for detailed investigation.
6. Democratizing Geotechnical Parameterization: Conventional soil parameterization requires specialized expertise and multiple laboratory tests. APHs deliver accurate parameters in seconds using only triaxial test data, making high-quality parameterization accessible to practitioners without specialized geotechnical expertise.
7. False Alarm Reduction: The framework demonstrated zero false alarms across 436 time steps (339 for 2023 Yiu Hing, 97 for 2005 Fei Ngo Shan), eliminating costs of unnecessary site visits, evacuations, and public distrust associated with empirical threshold methods.

*Table 7.1. Summary of cost-benefit and policy integration pathways.*

| Contribution | Cost-Benefit Implication | Policy/Integration Pathway |
|---|---|---|
| Smart sensing thresholds (rainfall thresholds) | Extended sensor lifespan and lower maintenance costs | Integrate threshold-based activation into GEO monitoring protocols. |
| Data loss mitigation & Virtual sensing | Protects monitoring investment; avoids redeployment costs | Implement as a backup during sensor outages in operational networks |
| Geometry-controlled vulnerability (R/r ratio) | Prioritizes high-risk areas, optimizes sensor placement | Inform land-use zoning; target convergent slopes for instrumentation |
| Rapid web-based assessment | Avoids unnecessary investigations; screens sites before detailed study | Deploy as a first-pass screening tool within the GEO workflow |
| Automated parameterization | Eliminates specialized tests; reduces laboratory costs; no expertise required | Adopt in geotechnical laboratories; update design codes to accept AI-derived parameters. |
| Zero false alarms (over 436 test cases) | Eliminates costs of unnecessary evacuations and site visits | Build public trust; reduce emergency response expenditures |

### *7.3.3. Broader Contributions*

1. Paradigm Shift towards Smart Geotechnics: This work paves the way for a new era of "smart" geotechnical engineering by demonstrating the successful integration of DL into core practices such as laboratory parameterization and field monitoring, promoting automation and data-driven decision-making.
2. Enhanced Climate Change Resilience: By improving the accuracy of landslide early warning systems and clarifying the controls on landslide vulnerability, this research provides urban planners and engineers with the tools and knowledge to design more resilient infrastructure and land-use strategies in the face of intensifying rainfall events due to climate change.
3. Adaptable Methodological Framework: While developed and validated using Hong Kong's unique geological conditions, the underlying methodological framework, from data preparation to model integration, is highly adaptable. It provides a blueprint that can be transferred to other landslide-prone regions worldwide with the integration of local datasets and model retuning.

## 7.4. RESEARCH LIMITATIONS

1. Data Constraints Hydrological Sensing: The performance of the data-driven models is constrained by the temporal span and resolution of the available dataset. Approximately two years of daily sensor data, while extensive, limits the models' exposure to rare but critical scenarios, such as rapid transitional soil states (e.g., sudden drying or wetting), leading to predictive uncertainty for these events. Furthermore, the reliance on binary saturation assumptions stems from sparse direct pore-water pressure monitoring.
2. Spatial Simplifications: The numerical analyses and tool generation are based on 2D axisymmetric or idealized slope geometries. Natural slopes are complex 3D characterized by significant heterogeneity in stratigraphy, topography, and material properties. These simplifications may not fully capture 3D effects like lateral flow convergence/divergence, which can critically influence pore-pressure evolution and failure mechanisms.
3. Model Architecture and Generalizability: The current DL models are specifically tuned for Hong Kong's colluvium soil and climatic conditions, corroborating the "No Free Lunch" theorem that no single model is universally superior. Their direct application to other regions is non-trivial and requires retraining with local data. There was also limited exploration of more recent or hybrid neural architectures.
4. Feature and Environmental Complexity: The current models primarily rely on rainfall and spatiotemporal features. While these are highly influential, they do not fully incorporate other environmental factors such as precise vegetation cover, evaporation rates, plant transpiration, and regional-scale soil mechanical properties, which could further refine prediction accuracy and resilience to future climatic challenges.
5. Operational Implementation and Computational Efficiency: The transition from a historically validated framework to a fully operational, real-time early warning system requires further development. While the CNN model was relatively efficient, other

developed models (e.g., LSTM, GRU) involved higher computational costs, which may pose challenges for real-time deployment in resource-constrained environments. The current system also lacks dynamic data handling protocols.

6. Data Resolution and Methodological Optimization: The current modeling framework is constrained by the limited incorporation of high-resolution temporal data, such as detailed vegetation indices and soil properties. This data gap restricts the ability to systematically determine the optimal class ratio for model training, which is a critical methodological parameter. Consequently, the influence of dynamic slope changes on model performance and the ideal balance between landslide and non-landslide instances remains an area requiring further investigation with a more comprehensive and refined dataset.

## 7.5. RECOMMENDATIONS FOR FUTURE WORK

Building upon the findings, limitations, and achievements of this research, future work is organized along two interconnected pathways: the enhancement of the core methodologies developed in this study and the extension of the framework to new domains and applications. The first strand focuses on deepening the technical robustness, accuracy, and physical grounding of the proposed models and tools, while the second aims to scale the research outcomes into operational systems and validate them across a wider range of scenarios and regions. Furthermore, the methodologies and tools developed in this thesis have demonstrated potential that extends beyond their initial application to landslide stability, giving rise to a third pathway dedicated to novel research avenues and technological spin-offs.

### *7.5.1. Enhancements for the Developed Research*

a) Advanced data acquisition and synthesis:

- Future work should prioritize deploying denser sensor networks, including distributed pore-water pressure, to move beyond binary saturation assumptions. This should be coupled with high-resolution temporal data on evaporation, soil temperature, and vegetation indices (e.g., NDVI) to capture land-surface energy balance and its impact on soil hydrology. Incorporating vegetation parameters such as root architecture, root area ratio, and seasonal dynamics would enable a more accurate representation of hydrological interception and mechanical reinforcement effects. As field data becomes available, hysteretic SWCC models should be incorporated to characterize wetting and drying paths separately. Additionally, higher-resolution rainfall data and sub-daily modeling would enable systematic evaluation of performance across 1-, 3-, 6-, and 12-hour horizons, further enhancing operational applicability.
- To address the scarcity of data for rare failure scenarios, a program of high-fidelity synthetic data generation should be implemented. This involves using numerically calibrated 3D simulations to model a wide range of conditions. Advanced sampling techniques, such as Sobol sequences, should be employed to maximize the diversity and coverage of the parameter space within these simulations, ensuring the synthetic dataset is both comprehensive and efficient.
- Future work should establish a systematic framework linking predictive uncertainty to operational decisions through reliability diagrams, cost–loss analysis, and probabilistic trigger levels. Reliability diagrams would assess calibration between predicted and observed failure frequencies across prediction horizons. Cost–loss analyses would optimize decision thresholds based on explicit ratios between false alarm and missed event costs, providing an economic basis for stakeholder-specific threshold selection.

b) Next-generation model architecture and training:

- The promising results from LSTM and CNN models warrant investigation into more sophisticated DL architectures. Transformer models should be explored for their potential to better capture long-term dependencies in temporal data with improved computational efficiency. Future work will also employ tree-based models and ablation studies to rigorously isolate and quantify the contribution of spatial features to prediction performance.
- A key research direction is the systematic investigation of the optimal class ratio for model training. This requires a more comprehensive dataset to robustly analyze how this critical methodological parameter affects model performance, particularly the balance between landslide and non-landslide instances and its interaction with dynamic slope conditions. Furthermore, future work will incorporate time-blocked and spatial-blocked validation strategies to further enhance operational credibility.

c) Enhancement of numerical and physical grounding:

- A critical step forward is the transition from 2D axisymmetric models to high-resolution 3D simulation frameworks that utilize realistic terrain meshes derived from LiDAR or other topographic surveys. This is essential to capture 3D effects like lateral flow convergence/divergence.
- The numerical backbone, particularly for the APHs tool and synthetic data generation, must be strengthened by incorporating advanced constitutive soil models. These models should account for critical behaviors such as anisotropy, cyclic loading effects, and partial saturation, moving beyond the capabilities of the standard HS model.

### *7.5.2. Extensions and Broader Applications*

a) Development of a geotechnical digital twin:

- The forecasting, VWC prediction, and parameterization tools should be integrated into a single platform with robust data normalization and ingestion pipelines capable of processing real-time streaming sensor data. This integrated system would enable continuous validation against field observations, creating a closed-loop framework where model predictions are constantly refined against actual slope behavior.
- To prevent performance drift over time, continual learning protocols must be established. This system would periodically update model weights based on new monitoring data, allowing the models to adapt to changing environmental conditions and improve their predictive accuracy throughout their operational lifespan.

b) Increased tool versatility and applicability:

- The methodological framework should be applied and validated across diverse geographic and climatic regions to test transferability and adapt to different landslide triggering mechanisms. This involves: (i) collecting local historical landslide data; (ii) recalibrating region-specific antecedent windows through sensitivity analysis; (iii) retraining deep learning models with hyperparameter tuning optimized for local characteristics; (iv) re-establishing decision thresholds using the dual-threshold method calibrated to regional risk tolerance; and (v) validating against local case studies.
- The framework should undergo structured validation with end-users (e.g., geotechnical agencies) to assess usability, integration into existing workflows, and decision-support effectiveness across varying geological conditions. These efforts would build confidence in the framework's adaptability, reliability, and practical utility for real-world operational deployment.

c) Climate resilience and proactive hazard management:

- The early-warning systems should be stress-tested against future climate scenario data. This involves using downscaled climate projections to evaluate system performance under changing rainfall patterns, including increased intensity and prolonged droughts, ensuring its resilience to future climatic challenges.
- To address uncertainty in exact failure timing, future work should focus on developing probabilistic timing models. This would enhance early warnings by providing confidence intervals for predicted failure events, moving from a binary alert to a risk-informed forecast.

### *7.5.3. Novel Research Directions and Spin-Off Technologies*

a) Development of a smart laboratory system for automated soil parametrization:

- Rationale: The APHs tool has proven highly effective in rapidly and accurately identifying soil parameters from pressuremeter tests. This capability forms the foundation for a transformative leap in geotechnical laboratory practices.
- Vision: The current process of calibrating advanced constitutive soil models (e.g., HS, Small-Strain Hardening) is a time-consuming and often subjective manual task. This project envisions a "Smart Lab" system that fully automates this process. By integrating the APHs tool directly with automated triaxial or pressuremeter testing apparatus, the system would close the loop between data acquisition and model calibration, delivering a fully characterized and validated soil model in near-real time.
- Implementation and Impact: The development would proceed through prototyping with a specific test apparatus (e.g., a pressuremeter), followed by expansion to triaxial systems. This technology would dramatically reduce the time and expertise required for advanced geotechnical analysis, making sophisticated constitutive modeling more accessible and reproducible.

### b) Development of a Smart Urban Flood Forecasting System:

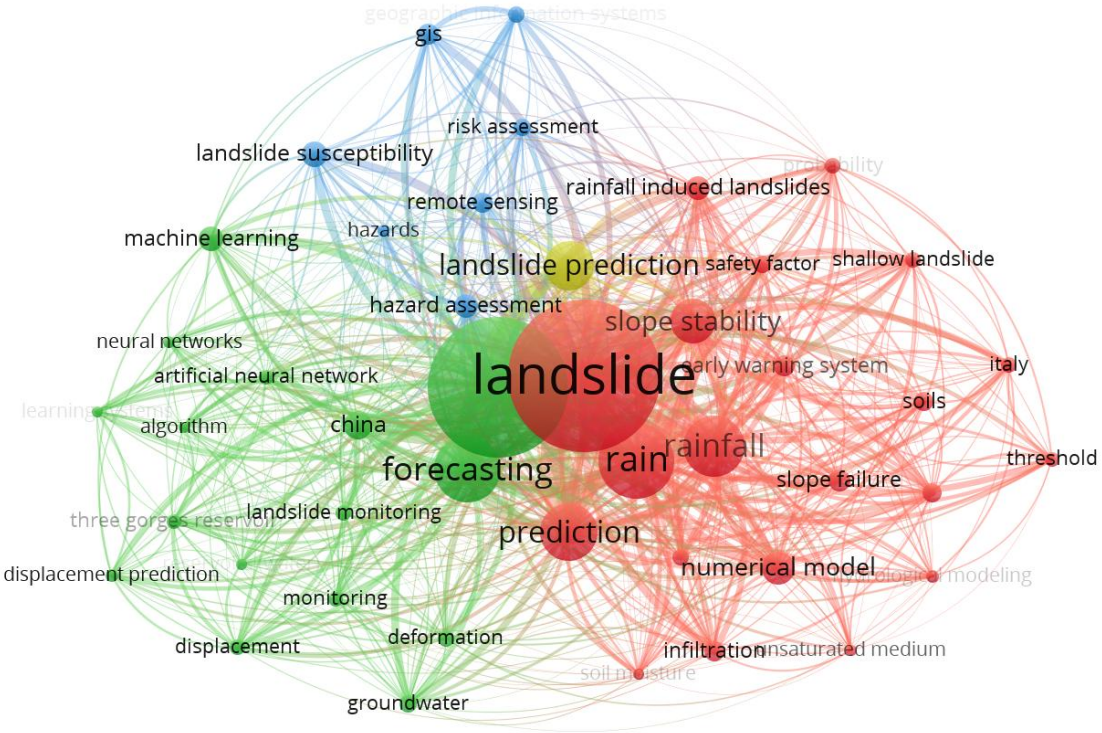
- Rationale: The OWHK model's core achievement is the operational forecasting of a hydrological response (VWC) to rainfall. This same architectural principle, proven effective for slopes, can be directly adapted to forecast a different but critically important hydrological variable in urban settings: flood water levels.
- Vision: Urban flood models often rely on complex 2D hydraulic simulations that are computationally intensive and not ideal for real-time, city-wide early warning. This project proposes leveraging the OWHK's data-driven, spatiotemporal forecasting framework to create a high-resolution urban flood nowcasting system.
- Implementation and Impact: A pilot study would focus on a frequently flooded urban catchment. Success would demonstrate a paradigm shift from slow, physics-intensive flood modeling to rapid, data-driven forecasting, providing city managers with a crucial tool for enhancing urban resilience to extreme weather events.

## 7.3. SUMMARY

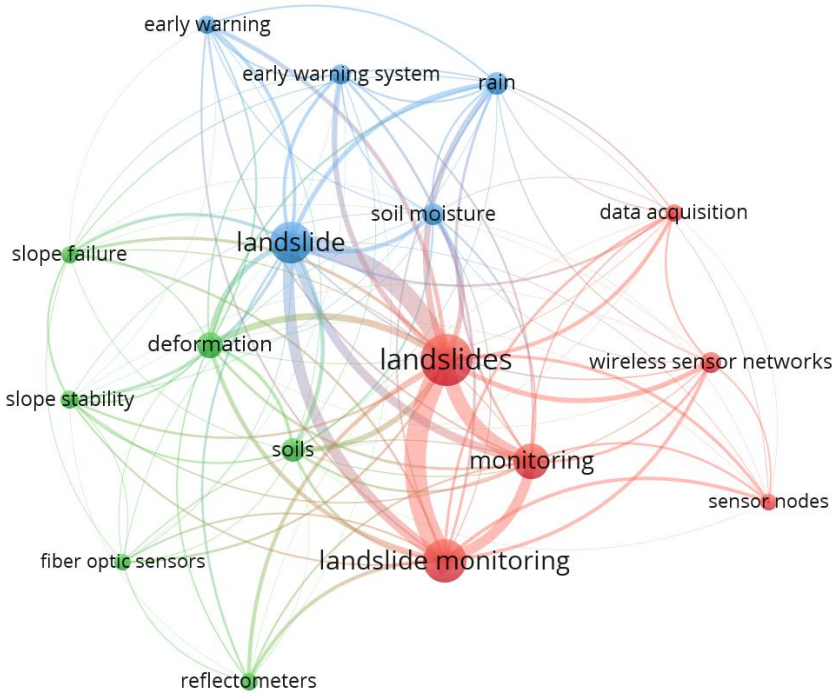
This chapter recaps the objectives of this research, followed by a summary of the key findings derived from the study. It then outlines the significant contributions and implications of the research. Additionally, the chapter acknowledges the limitations of the study. Based on these discussions, suggestions for potential future research directions are explained.

## APPENDIX A

VOSviewer software was used to identify frequently used keywords through co-occurrence analysis of "all keywords," with a minimum occurrence threshold of ten. For prediction models (Figure A1.a), 45 of 1785 keywords met the threshold, forming three clusters: blue (susceptibility mapping), red (physical models), and green (physically based thresholds). For monitoring techniques (Figure A1.b), 17 of 1050 keywords formed clusters for rainfall/soil moisture (blue), sensors/wireless networks (red), and subsurface displacement systems like optical fiber (green). Node size corresponds to keyword frequency, with "landslides" and "landslide monitoring" as dominant terms. This mapping elucidates primary research streams, susceptibility assessment, mechanistic modeling, threshold determination, and multi-scale monitoring, providing objective clustering that complements narrative reviews presented in Chapter 2.




(a) In landslide prediction.




(b) In landslide monitoring.

*Figure A1. Keyword mapping weighted by occurrence.*

A comprehensive evaluation framework is essential not only for selecting reliable features and optimizing model architecture, but also for rigorously validating model performance and dependability. This study employs multiple evaluation metrics to address different aspects of model performance, with particular attention to the challenges posed by imbalanced datasets. The Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the Curve (AUC) value serve as primary evaluation measures (Figure A2). The AUC quantifies the model's ability to distinguish between classes, with values ranging from 0.5 (random performance) to 1.0 (perfect discrimination). Additionally, confusion matrices (Table A5) enable detailed analysis of classification performance across all categories, facilitating identification of specific error patterns and model strengths. Given that accuracy can be misleading for imbalanced datasets, we implemented a comprehensive suite of metrics from scikit-learn to provide a nuanced performance assessment (Table A2).

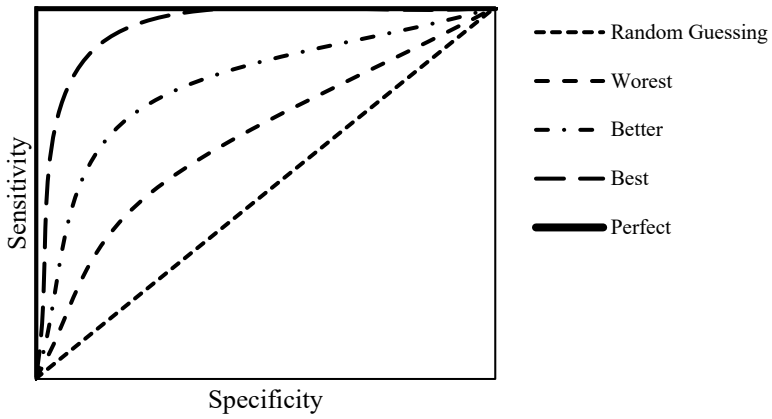



*Figure A2. The receiver operating characteristic (ROC).*

*Table A1. Error matrix.*

| **The total population P + N** | | **Predicted condition** | |
|---|---|---|---|
| | | Negative (PN) | Positive (PP) |
| Actual condition | Negative (N) | True negative (TN) | False positive (FP) |
| | Positive (P) | False negative (FN) | True positive (TP) |

FP and FN are the inaccurately recognized landslide or non-landslide occurrences; TP is the successfully identified landslide event; and TN is the correctly identified non-landslide event. P represents the total number of landslides, and N represents the total number of non-landslide events.

*Table A2. Evaluation metrics definitions.*

| Metric | Definitions |
|---|---|
| Precision | Measures the reliability of positive landslide predictions |
| Recall | Assesses the model's ability to identify actual landslide cases |
| F1-score | Provides a balanced measure between precision and recall |
| Balanced Accuracy | Compensates for dataset imbalance |
| Cohen's Kappa | Evaluates agreement beyond chance |
| Matthews Correlation Coefficient | Offers robust assessment for binary classification |

Geotechnically, it is seen that vegetation strengthens the cohesion parameter (Equation A1), while the friction angle is poorly affected. The factors of safety for unsaturated conditions and saturated conditions can be estimated by Equations A2 and A3, respectively (Salvatici et al., 2018)

$$C_{tot} = C' + fT_r(A_r / A) \quad \textit{(Equation A1)}$$

$$FS = \frac{\tan\phi}{\tan\beta} + \frac{C_{tot}}{\gamma_d H \sin\beta} + \left( \gamma_w u_w \tan\phi \left\{ \left[ 1 + \left( h_b^{-1} \mid u_w \mid \right)^{\lambda+1} \right]^{\frac{\lambda}{\lambda+1}} \right\}^{-1} / \gamma_d H \sin\beta \right) \quad \textit{(Equation A2)}$$

$$FS = \frac{\tan\phi}{\tan\beta} + \frac{C_{tot}}{(\gamma_d (H - u_w) + \gamma_{sat} u_w)\sin\beta} + \frac{\gamma_w u_w \tan\phi}{(\gamma_d (H - u_w) + \gamma_{sat} u_w)\tan\beta} \quad \textit{(Equation A3)}$$

Where $T_r$ is the root failure strength, $A_r/A$ is the root area ratio, $f$ is a coefficient dependent on the effective soil friction angle and the orientation of roots, $\phi$ is the friction angle, $\beta$ is the slope angle, $\gamma_d$ is the dry soil unit weight, $H$ is the depth, $\gamma_w$ is the water unit weight, $u_w$ is the pressure head, $h_b$ is the bubbling pressure, $\lambda$ is the pore size index distribution, and $\gamma_{sat}$ is the saturated soil unit weight.

*Table A3. Key advantages and disadvantages of DL and physically based models.*

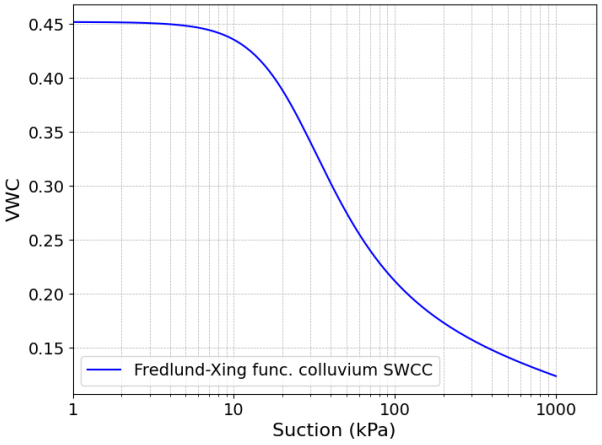
| Aspect | DL | Physically based models |
|---|---|---|
| Principle | Data-driven; learns spatiotemporal dynamics from historical observations. | Process-based; uses physical laws to simulate soil-water interactions. |
| Requirements | Requires large, high-quality historical datasets for training. | Relies on detailed physical parameters and site-specific calibration data. |
| Efficiency | Generally, faster once trained; well-suited for real-time forecasting and handling large data volumes. | Typically, computationally intensive due to detailed simulations. |
| Suitability | When sensor data is available, rapid forecasting is needed, especially under complex conditions. | When physical consistency is required or when extrapolating to new conditions. |

*Table A4. Statistical, shallow ML, and DL models*

| | |
|---|---|
| Naïve | Forecasts the next value to be equal to the last observed value. |
| Simple average | Predicts future values using the mean of all historical observations. |
| Single exponential average | Uses an exponentially weighted average of all past observations, with higher weights on more recent data. |
| Triple exponential average | Extends exponential smoothing to model data with trend and seasonality using three smoothing equations: level, trend, and seasonality. |
| SARIMA | Models time series using Autoregressive (AR), Moving Average (MA), and Differencing (I) components for both non-seasonal and seasonal patterns. |
| Rolling moving average | Forecasts are obtained by averaging the most recent observations within a fixed-size, sliding window. |
| Multivariate regression | Predicts future values of a time series using its own past and the past values of other related variables. |
| DNN | Uses multiple hidden layers to learn complex, hierarchical patterns and non-linear relationships in historical data. |
| Bi-LSTM | Processes sequential data in both forward and backward directions to capture context from past and future points in a sequence (Graves & Schmidhuber, 2005). |

*Table A5. Mechanical, geotechnical, and hydrological soil characteristics at Pa Mei*

| Material | | Colluvium |
|---|---|---|
| c (kPa) [1] | | 3 |
| $\phi$' (degree) [1] | | 34.3 |
| $\phi^b$ (degree) [2] | | 15 |
| Dry unit weight (kN/m$^3$) [1] | | 14 |
| Slope (degrees) [4] | | 30 |
| Fredlund – Xing Parameters[3] | a (kPa) | 20.3 |
| | n | 2.356 |
| | m | 0.562 |




[1](Wong, 2020); [2](Ng et al., 1998); [3](Fredlund, 1998); [4](Geotechnical Engineering Office, 2010a)

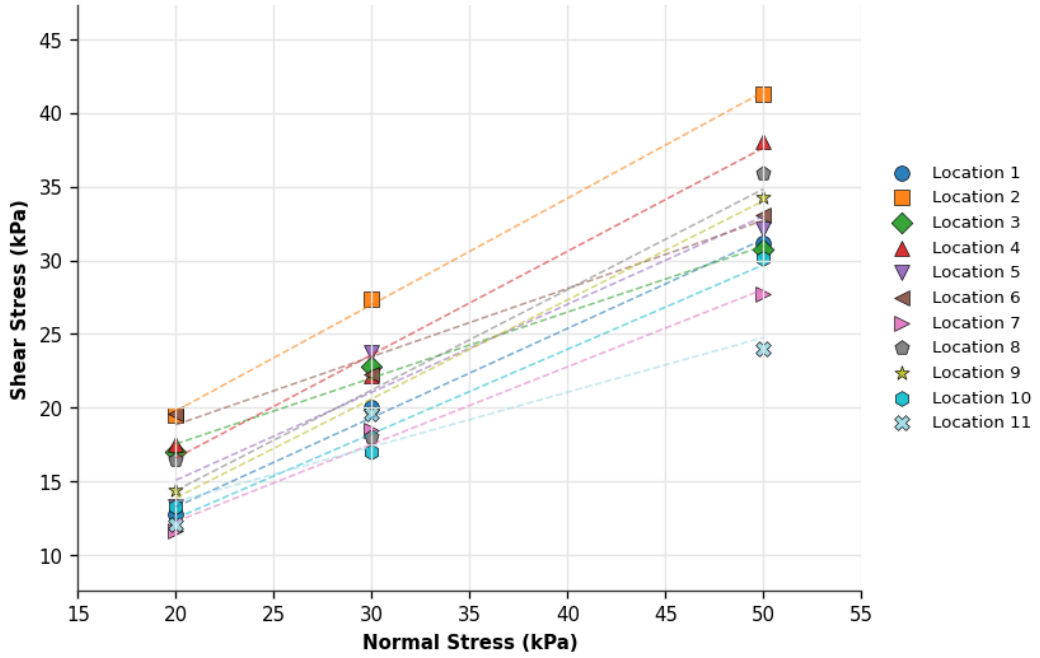



*Figure A3. Shear strength vs. normal stress relationship.*

*Table A6. HS model parameter range for model training and validation*

| Parameter | Minimum value | Maximum value |
|---|---|---|
| Over-consolidation ratio (OCR) | 1 | |
| $K_o^{NC}$ | Calculated by the solver $(1-\sin\phi)$ | |
| $\upsilon_{ur}$ | 0.15 | 0.35 |
| $E_{ur}^{ref}$ (kPa) | 10000 | 50000 |
| $E_{50}^{ref}$ (kPa) | 2000 | 0.5 $E_{ur}^{ref}$ |
| $E_{oed}^{ref}$ (kPa) | 0.5 * $E_{50}^{ref}$ | 0.8 * $E_{50}^{ref}$ |
| $m$ | 0.15 | 0.99 |
| $\sigma^{ref}$ (kPa) | 100 | |
| $c'$ (kPa) | 0 | 8 |
| $\phi'$ (°) | 29 | 36 |
| $\psi$ (°) | 0 | |
| $R_f$ | 0.5 | 1 |

*Table A7. Numerical features statistics (Historical landslide records)*

| Feature | Class | Count | Mean | Std | Min | 25% | 50% | 75% | Max |
|---|---|---|---|---|---|---|---|---|---|
| CR1W (mm) | Non-Landslide | 3733.0 | 78.8 | 114.4 | 0.0 | 4.0 | 31.0 | 107.5 | 1067.0 |
| | Landslide | 1307.0 | 337.2 | 236.7 | 0.0 | 140.5 | 327.5 | 514.0 | 1045.0 |
| CR2W (mm) | Non-Landslide | 3733.0 | 148.6 | 176.8 | 0.0 | 24.0 | 84.5 | 212.5 | 1269.5 |
| | Landslide | 1307.0 | 457.8 | 275.6 | 0.0 | 249.0 | 446.5 | 678.8 | 1255.5 |
| CR3W (mm) | Non-Landslide | 3733.0 | 206.5 | 215.6 | 0.0 | 44.0 | 141.0 | 302.5 | 1476.0 |
| | Landslide | 1307.0 | 551.5 | 307.8 | 0.0 | 336.3 | 538.0 | 791.0 | 1535.0 |
| CR4W (mm) | Non-Landslide | 3733.0 | 267.4 | 244.6 | 0.0 | 82.5 | 215.5 | 367.0 | 1714.0 |
| | Landslide | 1307.0 | 629.4 | 327.2 | 0.0 | 403.5 | 635.0 | 838.5 | 1731.0 |
| MRI (mm) | Non-Landslide | 3733.0 | 91.4 | 80.7 | 0.0 | 34.0 | 68.5 | 132.0 | 769.0 |
| | Landslide | 1307.0 | 235.5 | 136.6 | 0.0 | 127.0 | 224.0 | 326.0 | 769.0 |
| RD (d) | Non-Landslide | 3733.0 | 5.3 | 3.8 | 0.0 | 2.0 | 4.0 | 8.0 | 20.0 |
| | Landslide | 1307.0 | 8.6 | 4.1 | 0.0 | 6.0 | 9.0 | 11.0 | 19.0 |
| Slope (deg) | Non-Landslide | 3733.0 | 34.4 | 7.3 | 9.9 | 29.6 | 34.1 | 38.8 | 66.3 |
| | Landslide | 1307.0 | 34.1 | 7.0 | 9.9 | 29.7 | 33.9 | 38.3 | 66.3 |
| Elevation (m) | Non-Landslide | 3733.0 | 118.3 | 113.2 | 2.6 | 32.0 | 85.1 | 158.5 | 604.6 |
| | Landslide | 1307.0 | 110.6 | 105.1 | 2.6 | 31.2 | 80.1 | 148.8 | 557.7 |
| Zone distance (m) | Non-Landslide | 3733.0 | 1427.0 | 1192.2 | 24.0 | 597.1 | 1050.5 | 1893.4 | 7849.8 |
| | Landslide | 1307.0 | 1566.1 | 1269.9 | 24.0 | 653.6 | 1179.6 | 2134.6 | 7809.9 |

*Table A8. Categorical features statistics (Historical landslide records)*

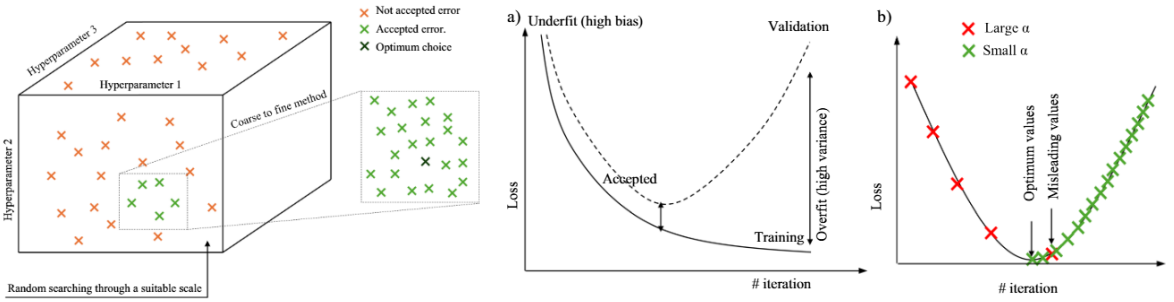
| Feature | | Non-Landslide | Landslide | Total | Landslide rate (%) |
|---|---|---|---|---|---|
| Material | Soil | 2833 | 1043 | 3876 | 26.91 |
| | Soil and Rock | 824 | 253 | 1077 | 23.49 |
| | Rock | 76 | 11 | 87 | 12.64 |
| Nearby structures | Roads | 943 | 284 | 1227 | 23.15 |
| | Paths | 964 | 344 | 1308 | 26.3 |
| | Drainages | 191 | 85 | 276 | 30.8 |
| | Residential Areas | 653 | 235 | 888 | 26.46 |
| | Construction & Industrial Areas | 52 | 20 | 72 | 27.78 |
| | Recreational Area | 57 | 18 | 75 | 24 |
| | Public Facility | 78 | 21 | 99 | 21.21 |
| | Open Spaces | 552 | 207 | 759 | 27.27 |
| | Hillside | 243 | 93 | 336 | 27.68 |
| Zone (spatial dependencies) | H01 | 51 | 9 | 60 | 15 |
| | H02 | 22 | 5 | 27 | 18.52 |
| | H03 | 54 | 18 | 72 | 25 |
| | H04 | 23 | 10 | 33 | 30.3 |
| | H05 | 22 | 5 | 27 | 18.52 |
| | H06 | 44 | 13 | 57 | 22.81 |
| | H07 | 55 | 14 | 69 | 20.29 |
| | H08 | 97 | 32 | 129 | 24.81 |
| | H09 | 42 | 6 | 48 | 12.5 |
| | H10 | 44 | 4 | 48 | 8.33 |
| | H11 | 9 | 3 | 12 | 25 |
| | H12 | 8 | 1 | 9 | 11.11 |
| | H13 | 17 | 4 | 21 | 19.05 |
| | H14 | 14 | 4 | 18 | 22.22 |
| | H15 | 53 | 13 | 66 | 19.7 |
| | H16 | 129 | 30 | 159 | 18.87 |
| | H17 | 97 | 29 | 126 | 23.02 |
| | H18 | 8 | 1 | 9 | 11.11 |
| | H19 | 28 | 5 | 33 | 15.15 |
| | H20 | 19 | 5 | 24 | 20.83 |
| | H21 | 16 | 5 | 21 | 23.81 |
| | H22 | 16 | 5 | 21 | 23.81 |
| | H23 | 7 | 2 | 9 | 22.22 |
| | H24 | 15 | 6 | 21 | 28.57 |
| | H26 | 32 | 7 | 39 | 17.95 |
| | H27 | 19 | 5 | 24 | 20.83 |
| | H28 | 9 | 3 | 12 | 25 |

*Table A8. Cont.*

| Feature | | Non-Landslide | Landslide | Total | Landslide rate (%) |
|---|---|---|---|---|---|
| Zone (spatial dependencies) | H29 | 13 | 5 | 18 | 27.78 |
| | H30 | 47 | 10 | 57 | 17.54 |
| | K01 | 21 | 9 | 30 | 30 |
| | K02 | 37 | 11 | 48 | 22.92 |
| | K03 | 4 | 2 | 6 | 33.33 |
| | K04 | 26 | 4 | 30 | 13.33 |
| | K05 | 32 | 13 | 45 | 28.89 |
| | K06 | 51 | 24 | 75 | 32 |
| | K07 | 102 | 27 | 129 | 20.93 |
| | K08 | 41 | 13 | 54 | 24.07 |
| | K09 | 16 | 5 | 21 | 23.81 |
| | K10 | 15 | 3 | 18 | 16.67 |
| | N01 | 49 | 20 | 69 | 28.99 |
| | N02 | 106 | 41 | 147 | 27.89 |
| | N03 | 41 | 19 | 60 | 31.67 |
| | N04 | 58 | 26 | 84 | 30.95 |
| | N05 | 52 | 23 | 75 | 30.67 |
| | N06 | 42 | 15 | 57 | 26.32 |
| | N07 | 20 | 7 | 27 | 25.93 |
| | N08 | 46 | 14 | 60 | 23.33 |
| | N09 | 33 | 15 | 48 | 31.25 |
| | N10 | 43 | 17 | 60 | 28.33 |
| | N11 | 8 | 4 | 12 | 33.33 |
| | N12 | 21 | 9 | 30 | 30 |
| | N13 | 30 | 15 | 45 | 33.33 |
| | N14 | 9 | 3 | 12 | 25 |
| | N15 | 28 | 11 | 39 | 28.21 |
| | N16 | 41 | 10 | 51 | 19.61 |
| | N17 | 37 | 17 | 54 | 31.48 |
| | N18 | 110 | 43 | 153 | 28.1 |
| | N19 | 130 | 53 | 183 | 28.96 |
| | N20 | 31 | 8 | 39 | 20.51 |
| | N21 | 27 | 12 | 39 | 30.77 |
| | N22 | 28 | 14 | 42 | 33.33 |
| | N23 | 27 | 12 | 39 | 30.77 |
| | N24 | 2 | 1 | 3 | 33.33 |
| | N25 | 63 | 24 | 87 | 27.59 |
| | N26 | 54 | 18 | 72 | 25 |
| | N27 | 79 | 20 | 99 | 20.2 |
| | N28 | 16 | 5 | 21 | 23.81 |
| | N29 | 37 | 17 | 54 | 31.48 |

*Table A8. Cont.*

| Feature | | Non-Landslide | Landslide | Total | Landslide rate (%) |
|---|---|---|---|---|---|
| Zone (spatial dependencies) | N30 | 39 | 12 | 51 | 23.53 |
| | N31 | 50 | 19 | 69 | 27.54 |
| | N32 | 39 | 18 | 57 | 31.58 |
| | N33 | 7 | 2 | 9 | 22.22 |
| | N34 | 38 | 19 | 57 | 33.33 |
| | N35 | 122 | 52 | 174 | 29.89 |
| | N36 | 51 | 21 | 72 | 29.17 |
| | N37 | 20 | 7 | 27 | 25.93 |
| | N38 | 43 | 17 | 60 | 28.33 |
| | N39 | 12 | 6 | 18 | 33.33 |
| | N40 | 42 | 18 | 60 | 30 |
| | N41 | 89 | 34 | 123 | 27.64 |
| | N42 | 54 | 21 | 75 | 28 |
| | N43 | 48 | 15 | 63 | 23.81 |
| | N44 | 20 | 10 | 30 | 33.33 |
| | N45 | 46 | 20 | 66 | 30.3 |
| | N46 | 140 | 55 | 195 | 28.21 |
| | N47 | 11 | 1 | 12 | 8.33 |
| | N48 | 24 | 9 | 33 | 27.27 |
| | N49 | 10 | 5 | 15 | 33.33 |
| | N50 | 40 | 8 | 48 | 16.67 |
| | N51 | 2 | 1 | 3 | 33.33 |
| | N52 | 27 | 9 | 36 | 25 |
| | N53 | 110 | 52 | 162 | 32.1 |
| | N54 | 12 | 6 | 18 | 33.33 |
| | N55 | 14 | 7 | 21 | 33.33 |
| Curvature | Diverging | 562 | 164 | 726 | 22.59 |
| | Planner | 572 | 169 | 741 | 22.81 |
| | Converging | 2599 | 974 | 3573 | 27.26 |
| Initial month (temporal dependencies) | January | 125 | 10 | 135 | 7.41 |
| | February | 277 | 11 | 288 | 3.82 |
| | March | 653 | 67 | 720 | 9.31 |
| | April | 753 | 149 | 902 | 16.52 |
| | May | 605 | 395 | 1000 | 39.5 |
| | June | 562 | 199 | 761 | 26.15 |
| | July | 276 | 250 | 526 | 47.53 |
| | August | 189 | 103 | 292 | 35.27 |
| | September | 109 | 44 | 153 | 28.76 |
| | October | 66 | 52 | 118 | 44.07 |
| | November | 56 | 10 | 66 | 15.15 |
| | December | 62 | 17 | 79 | 21.52 |




Coarse-tuning approach

Fine-tuning approach: a) loss versus iterations for cross-validation (CV) and training sets; b) loss versus iterations for different learning rates (modified from KhangPham, 2023))

*Figure A4. Coarse to fine-tuning approach*

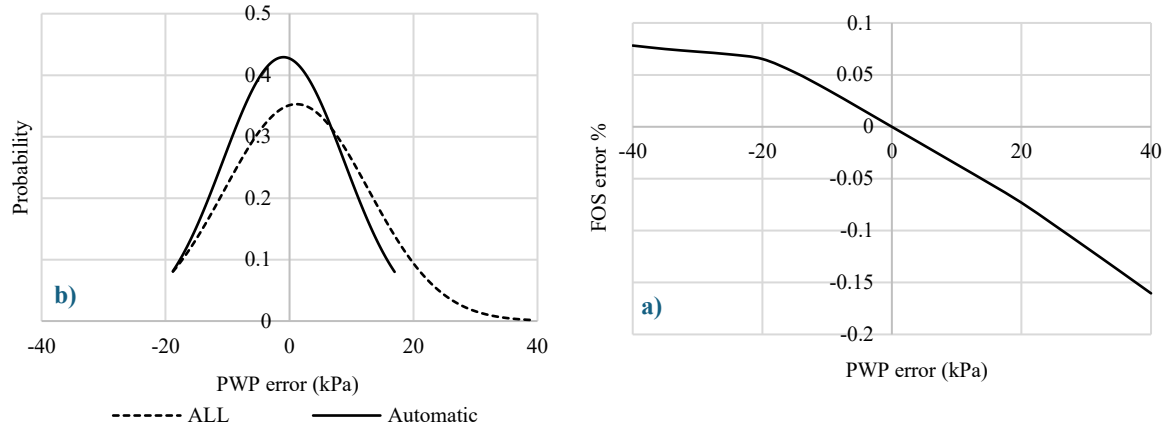



*Figure A5. a) Probability density function for the PWP error; b) The FOS error corresponding to the PWP.*

| Training (70%) | | Valid. (15%) | |
|---|---|---|---|
| 2237 | 377 | 488 | 72 |
| 249 | 667 | 58 | 138 |

| Testing (15%) | |
|---|---|
| 486 | 75 |
| 65 | 131 |

(a) CR1W

| Training (70%) | | Valid. (15%) | |
|---|---|---|---|
| 2187 | 427 | 449 | 111 |
| 240 | 676 | 53 | 143 |

| Testing (15%) | |
|---|---|
| 466 | 95 |
| 56 | 140 |

(b) CR2W

| Training (70%) | | Valid. (15%) | |
|---|---|---|---|
| 2086 | 528 | 424 | 136 |
| 237 | 679 | 46 | 150 |

| Testing (15%) | |
|---|---|
| 467 | 94 |
| 49 | 147 |

(c) CR3W

| Training (70%) | | Valid. (15%) | |
|---|---|---|---|
| 2118 | 496 | 440 | 120 |
| 247 | 669 | 47 | 149 |

| Testing (15%) | |
|---|---|
| 448 | 113 |
| 51 | 145 |

(d) CR4W

| Training (70%) | | Valid. (15%) | |
|---|---|---|---|
| 2079 | 535 | 429 | 131 |
| 273 | 643 | 53 | 143 |

| Testing (15%) | |
|---|---|
| 435 | 126 |
| 46 | 150 |

(e) MRI

| Training (70%) | | Valid. (15%) | |
|---|---|---|---|
| 1710 | 904 | 366 | 194 |
| 294 | 622 | 62 | 134 |

| Testing (15%) | |
|---|---|
| 354 | 207 |
| 58 | 138 |

(f) RD

*Figure A6. Detailed performance of LR models through confusion matrix analysis.*

*Table A9. LSTM model hyperparameters and results for earth pressure predictions at Tung Chung.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization - Dropout | LSTM architecture | | | | | Sampling ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | EP1A (Earth Pressure) | Rainfall + WTDR1C + WTDR2C | 0.95 | 0.97 | 0.52 | -0.01 | 0.97 | 0.04 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.4) | 128 | 3 | LeakyReLU (0.4) | 1 | linear | 72%:27% |
| Naïve | | - | 1.41 | 1.19 | 0.44 | -0.49 | 1.19 | 0.25 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | EP1B (Earth Pressure) | Rainfall + WTDR1C + WTDR2C | 1.33 | 1.15 | 0.75 | -0.15 | 1.12 | 0.00 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.9) | 128 | 3 | LeakyReLU (0.4) | 1 | linear | 72%:27% |
| Naïve | | - | 1.60 | 1.27 | 0.51 | -0.09 | 1.27 | 0.34 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | EP2A (Earth Pressure) | Rainfall | 5.70 | 2.39 | 1.68 | 0.64 | 2.35 | 0.81 | 20 | 32 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.8) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1C + WTDR2C | 4.84 | 2.20 | 1.38 | 0.69 | 2.17 | 0.84 | 20 | 4 | | | | (alpha = 0.2) | 64 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1C + WTDR2C | 4.64 | 2.15 | 1.35 | 0.70 | 2.12 | 0.85 | 50 | 4 | | | | (alpha = 0.2) | 64 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Naïve | | - | 5.95 | 2.44 | 1.15 | 0.62 | 2.44 | 0.81 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A10. LSTM model hyperparameters and results for earth pressure predictions at Tsing Shan.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization - Dropout | LSTM architecture | | | | | Sampling ratio |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | EP1P (earth pressure) | WTDR3A + WTDR3B + WTDR3C (no rainfall was used as it contains some missing data) | 7.46 | 2.73 | 1.60 | 0.63 | 2.67 | 0.81 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.3) | 256 | 3 | LeakyReLU (0.9) | 1 | linear | 70%:30% |
| | | WTDR3B + WTDR3C (no rainfall was used as it contains some missing data) | 5.77 | 2.40 | 1.33 | 0.71 | 2.40 | 0.87 | 50 | 16 | | | | (alpha = 0.3) | 256 | 3 | LeakyReLU (0.9) | 1 | linear | 70%:30% |
| | | WTDR3A + WTDR3B + WTDR3C + EP2P (no rainfall was used as it contains some missing data) | 4.45 | 2.11 | 0.94 | 0.78 | 2.11 | 0.89 | 50 | 16 | | | | (alpha = 0.3) | 256 | 3 | LeakyReLU (0.9) | 1 | linear | 70%:30% |
| Naïve | | - | 3.98 | 1.99 | 0.62 | 0.80 | 1.99 | 0.90 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | EP2P (earth pressure) | WTDR3A + WTDR3B + WTDR3C (no rainfall was used as it contains some missing data) | 4.27 | 2.07 | 1.66 | 0.89 | 1.99 | 0.96 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.3) | 256 | 3 | LeakyReLU (0.9) | 1 | linear | 70%:30% |
| | | WTDR3A + WTDR3B + WTDR3C + EP1P (no rainfall was used as it contains some missing data) | 2.97 | 1.72 | 1.26 | 0.93 | 1.72 | 0.96 | | | | | | | | | | | | |
| Naïve | | - | 1.07 | 1.03 | 0.49 | 0.97 | 1.03 | 0.99 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A11. LSTM model hyperparameters and results for suction (TC1) at Tung Chung.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization - Dropout | LSTM architecture | | | | | Sampling |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | TC1A (Suction) | Rainfall + WTDR1A + WTDR2A | 130.47 | 11.42 | 8.42 | 0.87 | 11.38 | 0.93 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.4) | 32&128&128 | 3 | LeakyReLU (0.2) | 1 | linear | 72%:27% |
| Naïve | | - | 205.19 | 14.32 | 3.99 | 0.80 | 14.32 | 0.90 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | TC1B (Suction) | Rainfall + WTDR1A + WTDR2A | 31.49 | 5.61 | 3.63 | 0.90 | 5.58 | 0.95 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.4) | 128 | 3 | LeakyReLU (0.1) | 1 | linear | 72%:27% |
| Naïve | | - | 37.14 | 6.09 | 1.85 | 0.88 | 6.09 | 0.94 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | TC1C (Suction) | Rainfall + WTDR1A + WTDR2A | 14.54 | 3.81 | 2.25 | 0.87 | 3.81 | 0.94 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.6) | 64 | 3 | LeakyReLU (0.5) | 1 | linear | 72%:27% |
| Naïve | | - | 15.25 | 3.91 | 1.17 | 0.87 | 3.90 | 0.93 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A12. LSTM model hyperparameters and results for suction (TC1) at Tung Chung – Extended dataset.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization - Dropout | LSTM architecture | | | | | Sampling |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | TC1A (Suction) | Rainfall + WTDR1A + WTDR2A | 58.20 | 7.63 | 5.03 | 0.95 | 7.53 | 0.98 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.4) | 256 | 3 | LeakyReLU (0.8) | 1 | linear | 72%:27% |
| Naïve | | - | 105.29 | 10.26 | 2.92 | 0.91 | 10.26 | 0.96 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | TC1B (Suction) | Rainfall + WTDR1A + WTDR2A | 29.13 | 5.40 | 3.32 | 0.94 | 5.12 | 0.98 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.8) | 256 | 3 | LeakyReLU (0.5) | 1 | linear | 72%:27% |
| Naïve | | - | 47.05 | 6.86 | 1.99 | 0.90 | 6.86 | 0.95 | - | - | - | - | - | - | - | - | - | - | - | - |
| Fredlund-Xing | | - | 1112.63 | 33.36 | 25.62 | -1.34 | 23.93 | 0.53 | | | | | | | | | | | | |
| LSTM | TC1C (Suction) | Rainfall + WTDR1A + WTDR2A | 28.41 | 5.33 | 2.88 | 0.88 | 4.93 | 0.94 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.7) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Naïve | | - | 13.26 | 3.64 | 1.34 | 0.94 | 3.64 | 0.97 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A13. LSTM model hyperparameters and results for suction (JFT1 and JFT2) at Tung Chung.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization - Dropout | LSTM architecture | | | | | Sampling |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | JFT1A (Suction) | Rainfall + WTDR1A + WTDR2A | 28.63 | 5.35 | 3.73 | 0.77 | 5.28 | 0.88 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.1) | 32 | 3 | LeakyReLU (0.9) | 1 | linear | 72%:27% |
| Naïve | | - | 32.88 | 5.73 | 1.39 | 0.74 | 5.73 | 0.87 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT2A (Suction) | Rainfall + WTDR1A + WTDR2A | 18.92 | 4.35 | 3.25 | 0.87 | 4.35 | 0.94 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.3) | 32 | 3 | LeakyReLU (0.9) | 1 | linear | 72%:27% |
| Naïve | | - | 24.11 | 4.91 | 1.68 | 0.83 | 4.91 | 0.91 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT1B (Suction) | Rainfall + WTDR1B + WTDR2B | 96.66 | 9.83 | 6.12 | 0.10 | 9.62 | 0.42 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.9) | 256 | 3 | LeakyReLU (0.5) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1B + WTDR2B + TC1C - Corrected | 70.90 | 8.42 | 4.68 | 0.83 | 8.35 | 0.92 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Fredlund-Xing | | - | 628.33 | 25.07 | 16.19 | -4.85 | 22.34 | 0.65 | | | | | | | | | | | | |
| Naïve | | - | 45.05 | 6.71 | 2.66 | 0.58 | 6.71 | 0.80 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT2B (Suction) | Rainfall + WTDR1B + WTDR2B | 168.59 | 12.98 | 8.74 | 0.37 | 12.94 | 0.62 | 20 | 64 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 32 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1B + WTDR2B + TC1C - Corrected | 15.52 | 3.94 | 2.41 | 0.95 | 3.66 | 0.98 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Fredlund-Xing | | - | 345.01 | 18.57 | 12.87 | -0.28 | 13.40 | 0.76 | | | | | | | | | | | | |
| Naïve | | - | 9.49 | 3.08 | 1.18 | 0.96 | 3.06 | 0.98 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT1C (Suction) | Rainfall + WTDR1C + WTDR2C | 119.73 | 10.94 | 6.70 | 0.52 | 9.73 | 0.82 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.3) | 256 | 3 | LeakyReLU (0.3) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1C + TC1C - Corrected | 33.45 | 5.78 | 3.81 | 0.87 | 5.56 | 0.94 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Fredlund-Xing | | | 154.54 | 12.43 | 7.71 | 0.38 | 11.17 | 0.84 | | | | | | | | | | | | |
| Naïve | | - | 23.74 | 4.87 | 1.99 | 0.91 | 4.86 | 0.95 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT2C (Suction) | Rainfall + WTDR1C + WTDR2C | 211.19 | 14.53 | 9.52 | 0.31 | 14.47 | 0.57 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.5) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1C + WTDR2B + TC1C - Corrected | 10.11 | 3.18 | 2.35 | 0.98 | 3.15 | 0.99 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.9) | 256 | 3 | LeakyReLU (0.5) | 1 | linear | 72%:27% |
| Fredlund-Xing | | | 773.56 | 27.81 | 16.06 | -1.54 | 24.24 | 0.63 | | | | | | | | | | | | |
| Naïve | | - | 41.79 | 6.46 | 1.94 | 0.86 | 6.46 | 0.93 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A14. LSTM model hyperparameters and results for suction (JFT1) at Tsing Shan.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization-Dropout | LSTM architecture | | | | | Sampling |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | JFT1A (Suction) | Rainfall + WTDR1A + WTDR2A + WTDR3A - Corrected | 1.74 | 1.32 | 1.02 | 0.88 | 1.29 | 0.96 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Fredlund-Xing | | WTDR1A | 10.68 | 3.27 | 3.00 | 0.29 | 1.35 | 0.95 | | | | | | | | | | | | |
| Naïve | | - | 4.88 | 2.21 | 0.68 | 0.68 | 2.21 | 0.84 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT1B (Suction) | Rainfall + WTDR1B + WTDR2B + WTDR3B - Corrected | 17.83 | 4.22 | 3.00 | 0.89 | 4.22 | 0.96 | 20 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Fredlund-Xing | | WTDR1B | 40.78 | 6.39 | 3.56 | 0.75 | 5.76 | 0.95 | | | | | | | | | | | | |
| Naïve | | - | 22.68 | 4.76 | 2.25 | 0.86 | 4.76 | 0.93 | - | - | - | - | - | - | - | - | - | - | - | - |
| LSTM | JFT1C (Suction) | Rainfall + WTDR1C + WTDR2C + WTDR3C - Corrected | 5.66 | 2.38 | 1.68 | 0.41 | 2.37 | 0.64 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.3) | 256 | 3 | LeakyReLU (0.3) | 1 | linear | 72%:27% |
| Fredlund-Xing | | WTDR1C | 140.42 | 11.85 | 9.07 | -13.72 | 7.83 | 0.46 | | | | | | | | | | | | |
| Naïve | | - | 2.88 | 1.70 | 0.80 | 0.70 | 1.70 | 0.85 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A15. LSTM model hyperparameters and results for VWC (WTDR2C) at Tung Chung.*

| Model | Target | Features | Metrics | | | | | | Input Parameters | | Loss function | Optimizer | Normalization | Regularization - Dropout | LSTM architecture | | | | | Sampling |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | MSE | RMSE | MAE | R2 | SDR | PRC | Window size | Batch size | | | | | Hidden neurons | Hidden layers | Hidden activation | Output neurons | Output activation | |
| LSTM | WTDR2C (VWC) | Rainfall + WTDR1C | 10.63 | 3.26 | 1.43 | 0.89 | 3.13 | 0.95 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.1) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| | | Rainfall + WTDR1C + JFT2B | 10.71 | 3.27 | 1.38 | 0.89 | 3.20 | 0.94 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.1) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| | | Rainfall + JFT2B + TC1C | 14.49 | 3.81 | 3.04 | 0.85 | 3.69 | 0.93 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.9) | 32 | 3 | LeakyReLU (0.9) | 1 | linear | 72%:27% |
| | | Rainfall + EP2A + EP2B | 13.25 | 3.64 | 2.43 | 0.86 | 3.63 | 0.93 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.1) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| | | Rainfall +BH7 | 11.14 | 3.34 | 1.97 | 0.88 | 3.29 | 0.94 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.5) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| | | Rainfall | 38.87 | 6.23 | 4.16 | 0.59 | 6.20 | 0.78 | 50 | 16 | Mean squared error | Adam | MinMaxScaler | (alpha = 0.1) | 256 | 3 | LeakyReLU (0.6) | 1 | linear | 72%:27% |
| Naïve | | - | 10.11 | 3.18 | 1.07 | 0.89 | 3.18 | 0.95 | - | - | - | - | - | - | - | - | - | - | - | - |

*Table A16. Synthesis of key findings and conclusions.*

| a) Mechanisms, Predicting Approaches, and Monitoring Systems for Shallow Landslides | |
|---|---|
| On failure mechanisms: | • Shallow landslides are primarily triggered by rainfall, where both intensity–duration and antecedent moisture conditions govern initiation.<br>• Displacement typically exhibits superposed trend and periodic components driven by hydrologic forcing, seasonality, and hydro-fluctuation effects. |
| On prediction approaches: | • Empirical rainfall thresholds (I-D) remain practical when physical inputs are scarce but suffer from high false-positive rates.<br>• API thresholds better emulate initial moisture conditions, improving accuracy when derived via regression or probabilistic analysis.<br>• Physically based causative thresholds directly link rainfall to internal state variables, though they require site-specific geotechnical inputs.<br>• Deterministic physical models deliver higher fidelity by explicitly representing initial/boundary conditions, infiltration, runoff, evaporation, and soil properties, but require detailed and costly site characterization. |
| On monitoring systems: | • Surface techniques provide millimeter-level positioning and broad spatial coverage but face challenges with real-time acquisition and environmental interference—suitable for susceptibility mapping but not for real-time alerts.<br>• Subsurface methods offer higher diagnostic value: extensometers, inclinometers, optical-fiber solutions (FBG/BOTDA), TDR, AE, tilt sensors, and accelerometers each provide complementary capabilities for detecting internal deformation.<br>• Hydrologic and thermal monitoring (soil moisture, suction, temperature) strengthens causative understanding and early-warning capability.<br>• MEMS/IoT nodes integrated into WSNs reduce wiring and cost but must address power, communication, and data integrity through optimized deployment strategies. |
| On time series modeling: | • Data frequency and temporal ordering critically shape predictions: monthly series require decomposition into trend and seasonal components, while daily series demand tailored preprocessing.<br>• DL models (e.g., LSTM) outperform shallow ML (i.e., ANN, SVM, RF) by capturing nonlinear temporal dependencies.<br>• Evaluation should move beyond RMSE/MAE to include weighted metrics and interval/multi-step forecasts that reflect the criticality of extreme events. |
| b) Landslide Prediction Model | |
| On model performance: | • DL models (LSTM, GRU, 1D CNN) achieved AUC values exceeding 90%, significantly outperforming traditional LR and prior studies.<br>• The superior performance is attributed to: (i) time-dependent dataset design capturing rainfall-triggering randomness, (ii) rigorous data preparation correcting spatiotemporal imbalances, (iii) comprehensive coarse-to-fine hyperparameter tuning, and (iv) novel dual-threshold decision scheme. |
| On feature importance: | • Rainfall-related predictors accounted for approximately 36% of total feature importance, with short-term lags, two-day antecedent rainfall, and rainfall pattern features being key contributors.<br>• Inclusion of spatiotemporal descriptors increased model precision by 7%, demonstrating the necessity of moving beyond rainfall-only inputs. |
| On operational efficiency: | • The 1D CNN emerged as the most efficient architecture, achieving comparable precision (~78%) to LSTM and GRU but with substantially lower computational cost, tuning time, and predictive uncertainty.<br>• The CNN also produced a hydrological response more consistent with physical models, capturing post-event drainage hysteresis. |
| On decision framework: | • The dual-threshold method proved superior to the conventional 0.5 cut-off, providing more nuanced risk assessment, reducing false decisions, and delivering a narrower, more actionable warning zone. |
| On transferability: | • Although calibrated for Hong Kong's colluvium, the underlying framework is designed for transferability to other regions, contingent upon integration of local data and appropriate model retuning. |

*Table A16. Cont.*

| | |
|---|---|
| c) VWC Prediction Operational Model (OWHK) | |
| On model innovation: | • The OWHK tool integrates independence-oriented time series windowing with a multi-parallel LSTM network architecture and deterministic seepage modeling.<br>• Boundary conditions (initial VWC) act as proxies for unmonitored environmental factors, while comprehensive rainfall descriptors drive dynamic hydrological response. |
| On predictive performance: | • The model delivers mean absolute error < 0.6% and $R^2 > 0.92$ across multiple sites, with quantified uncertainty via Monte Carlo dropout.<br>• Specialized LSTM networks mapped to specific segments of the soil-water retention curve proved effective in reducing bias from missing variables. |
| On limitations: | • Accuracy diminishes during rapid transitional soil states (dry-to-saturated) due to limited training samples, highlighting the need for expanded datasets capturing these events. |
| d) Hypertuning Strategy for Enhanced Model Performance | |
| On tuning and feature selection: | • Meticulous tuning and high-quality data are more critical than algorithmic complexity alone, corroborating the "No Free Lunch" theorem.<br>• Optimal antecedent periods of 10 and 20 days were identified for shallow and deeper soil layers, respectively.<br>• Allocating 50%–70% of data for training is a key determinant of superior model performance. |
| On model transferability: | • The tuned framework achieved high predictive accuracy (MAE 0.35%–1.25%) and reliability (PICP 0.86–0.95) across multiple landslide sites (Pa Mei, Tung Chung).<br>• Propagation of predicted VWC into FOS calculations yielded robust metrics ($R^2 = 0.89$, PICP = 0.85) for stability assessment. |
| On monitoring innovation: | • The study demonstrates that data from a network of sensors can forecast the status of any individual sensor, mitigating challenges related to power consumption and data integrity |
| e) Distributed Multi-Sensor (Subsurface) Prediction Model | |
| On the Modeling approach: | • The LSTM-based framework successfully captures the critical spatiotemporal relationship between rainfall and the slope's antecedent hydrological state by integrating historical sequences from spatially distributed sensors. |
| On predictive accuracy: | • Matric suction prediction: Significantly outperformed naïve model (RMSE of 5.40 kPa)<br>• VWC prediction: Dramatic improvement using multivariate data (RMSE of 3.26%) over rainfall-alone inputs (RMSE of 6.23%) |
| On physical consistency: | • The model accurately replicates transient wetting front dynamics, temporal lags at different depths, and distinct seasonal responses, confirmed through successful generalization across independent case studies. |
| f) Automated Parameterization Tool (APHs) | |
| On performance: | • Model accuracy of 99.4% in estimating critical strength and stiffness parameters solely from triaxial test data.<br>• For CDG-recompacted soil, the tool accurately retrieved parameters even when they fell outside literature-proposed ratios, demonstrating robust generalization. |
| On methodological innovations: | • Advanced data generation using Latin Hypercube Sampling (LHS) and SIGMA/W simulations created a robust parameter space tailored to Hong Kong's geological conditions.<br>• Multi-model architecture enabled assignment of distinct weights to each parameter, mimicking physical processes in parameter derivation.<br>• Framing stress-strain relationships as a time series problem leveraged the sequential nature of triaxial test results for precise estimation. |
| On practical impact: | • The tool represents a paradigm shift toward smart laboratory systems, automating traditionally manual parameterization processes. |

*Table A16. Cont.*

| | |
|---|---|
| g) Automated Slope Stability Analysis Model (Hybrid DL–Physics Model) | |
| On framework innovation: | • The hybrid approach trains DL models to learn saturation dynamics from real rainfall sequences and embeds these outputs within an infinite slope stability equation.<br>• This captures time-dependent failure mechanisms while maintaining computational tractability across Hong Kong's extensive natural terrain. |
| On predictive performance: | • Balanced accuracy ≈ 0.82, precision between 69–74%, and AUC values ranging from 0.87 to 0.91 across all dataset splits.<br>• Strong physical agreement with deterministic seepage-stability analyses (MAE in FOS 0.06–0.07 for the 2005 Fei Ngo Shan case). |
| On validation: | • Successfully tested on the 2023 Yiu Hing Road landslide, demonstrating consistent alignment with professionally assessed stability conditions.<br>• The system integrates 28-day antecedent rainfall sequences from 96 gauges to output real-time, SSI-informed FOS trajectories. |
| h) Factors Controlling Rainfall-Induced Landslides: A Parametric Study | |
| On rainfall pattern effects: | • Relying solely on cumulative rainfall can be fundamentally misleading, as it ignores the decisive role of temporal rainfall patterns.<br>• For identical cumulative rainfall totals, delayed versus advanced patterns produce significantly different hydrological responses and failure timings:<br>- The delayed pattern allows adequate time for water seepage<br>- Advanced pattern intensifies pore-pressure buildup, delaying time to failure |
| On Slope recovery (healing) dynamics: | • Recovery time is not fixed but controlled primarily by soil hydraulic conductivity—higher conductivity leads to faster drainage and shorter recovery.<br>• Recovery is significantly modulated by initial soil suction conditions, rainfall duration, and intensity:<br>- Smaller initial suction, longer duration, and heavier rainfall → prolonged recovery time |
| On the antecedent period definition: | • For Hong Kong's colluvium, a variable window of at least 15 days, extending beyond 32.5 days, must be considered to accurately account for the slope's antecedent state. |
| i) Geometry Effects on Rainfall-Induced Landslides: A Parametric Study | |
| On topographic control: | • Subsurface flow geometry, specifically the inflow-to-outflow area ratio (R/r), exerts a first-order control on regional vulnerability to rainfall-induced landslides. |
| On failure regimes: | • Convergent hillslopes (high R/r): Focus infiltrating water into confined zones → rapid pore-pressure accumulation → antecedent-dominated regime where residual moisture accelerates failure onset even under modest rainfall.<br>• Divergent hillslopes (low R/r): Distribute inflow across larger areas → attenuated groundwater development → intensity-dominated regime where failure potential scales directly with rainfall rate |
| On practical application: | • Incorporating quantitative geometric descriptors (R/r) into regional models significantly improves predictive skill, enabling prioritization of monitoring and mitigation in the most climate-sensitive terrain. |
| j) Numerical Slope Stability Modeling (Focused Site-Specific Analysis) | |
| On temporal prediction: | • While the FOS remains largely invariant for a given cumulative rainfall total, the temporal resolution and pattern of rainfall data exert a primary control on the timing of predicted slope response.<br>• High-frequency rainfall inputs are essential for accurate temporal forecasting without materially altering ultimate stability assessment. |
| On vulnerable zone identification: | • The BH3 vicinity at Fei Ngo Shan was identified as the critical monitoring location due to: (i) the thinnest colluvium layer, (ii) position at slope toe where lateral inflow and direct infiltration converge, and (iii) characteristic groundwater lag response. |
| On the failure mechanism: | • Shallow landslides in this context are toe-initiated: large deformations first develop at the toe → loss of support → retrogressive mobilization of the upslope shallow layer, while soil along the rupture surface exhibits only minor sliding. |

## REFERENCES


Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jozefowicz, R., Jia, Y., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Schuster, M., Monga, R., Moore, S., Murray, D., Olah, C., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., & Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Abbaszadeh Shahri, A., Shan, C., & Larsson, S. (2022). A Novel Approach to Uncertainty Quantification in Groundwater Table Modeling by Automated Predictive Deep Learning. Natural Resources Research, 31(3), 1351-1373. https://doi.org/10.1007/s11053-022-10051-w

Abbaszadeh Shahri, A., Shan, C., Larsson, S., & Johansson, F. (2024). Normalizing Large Scale Sensor-Based MWD Data: An Automated Method toward A Unified Database. Sensors, 24(4), 1209. https://doi.org/10.3390/s24041209

Abbaszadeh Shahri, A., Shan, C., Larsson, S., & Johansson, F. (2024). Normalizing Large Scale Sensor-Based MWD Data: An Automated Method toward A Unified Database. Sensors, 24(4), 1209. https://www.mdpi.com/1424-8220/24/4/1209

Abbaszadeh Shahri, A., Shan, C., Zäll, E., & Larsson, S. (2021). Spatial distribution modeling of subsurface bedrock using a developed automated intelligence deep learning procedure: A case study in Sweden. Journal of Rock Mechanics and Geotechnical Engineering, 13(6), 1300-1310. https://doi.org/10.1016/j.jrmge.2021.07.006

Abolmasov, B., Milenković, S., Marjanović, M., Đurić, U., & Jelisavac, B. (2015). A geotechnical model of the Umka landslide with reference to landslides in weathered Neogene marls in Serbia. Landslides, 12(4), 689-702. https://doi.org/10.1007/s10346-014-0499-4

Abraham, M. T., Satyam, N., Bulzinetti, M. A., Pradhan, B., Pham, B. T., & Segoni, S. (2020b). Using field-based monitoring to enhance the performance of rainfall thresholds for landslide warning. Water, 12(12), 1-21, 3453. https://doi.org/10.3390/w12123453

Abraham, M. T., Satyam, N., Pradhan, B., & Alamri, A. M. (2020a). IoT-Based Geotechnical Monitoring of Unstable Slopes for Landslide Early Warning in the Darjeeling Himalayas. Sensors, 20(9), 2611. https://www.mdpi.com/1424-8220/20/9/2611

Agafonkin, V. (2023). Leaflet: an open-source JavaScript library for interactive maps. https://leafletjs.com

Ahmed, F. S., Bryson, L. S., & Crawford, M. M. (2021). Prediction of seasonal variation of in-situ hydrologic behavior using an analytical transient infiltration model. Engineering Geology, 294, 106383. https://doi.org/10.1016/j.enggeo.2021.106383

Albanese, D., Filosi, M., Visintainer, R., Riccadonna, S., Jurman, G., & Furlanello, C. (2013). minerva and minepy: a C engine for the MINE suite and its R, Python and MATLAB wrappers. Bioinformatics, 29(3), 407-408. https://doi.org/10.1093/bioinformatics/bts707

An, H., Viet, T. T., Lee, G., Kim, Y., Kim, M., Noh, S., & Noh, J. (2016). Development of time-variant landslide-prediction software considering three-dimensional subsurface unsaturated flow. Environmental Modelling & Software, 85, 172-183. https://doi.org/10.1016/j.envsoft.2016.08.009

Angeli, M.-G., Pasuto, A., & Silvano, S. (2000). A critical review of landslide monitoring experiences. Engineering Geology, 55(3), 133-147. https://doi.org/10.1016/S0013-7952(99)00122-2

Anura3D MPM Research Community, 2023. Anura3D Version 2023 Source Code, www.anura3d.com

Arribas-Bel, D., & Contextily Contributors. (2020). Contextily. © Copyright 2020, Dani Arribas-Bel & Contextily Contributors. Retrieved from https://contextily.readthedocs.io/en/latest/

Askarinejad, A., Akca, D., & Springman, S. M. (2018). Precursors of instability in a natural slope due to rainfall: a full-scale experiment. Landslides, 15(9), 1745-1759. https://doi.org/10.1007/s10346-018-0994-0

Au, S.W.C. (1998). Rain-induced slope instability in Hong Kong. Engineering Geology, 51(1),1–36. https://doi.org/10.1016/S0013-7952(98)00038-6.

Auflic, M. J., Herrera, G., Mateos, R. M., Poyiadji, E., Quental, L., Severine, B., Peternel, T., Podolszki, L., Calcaterra, S., Kociu, A., Warmuz, B., Jelenek, J., Hadjicharalambous, K., Becher, G. P., Dashwood, C., Ondrus, P., Minkevicius, V., Todorovic, S., Moller, J. J., & Marturia, J. (2023). Landslide monitoring techniques in the Geological Surveys of Europe. Landslides, 20(5), 951-965. https://doi.org/10.1007/s10346-022-02007-1

Azmoon, B., Biniyaz, A., Liu, Z., & Sun, Y. (2021). Image-data-driven slope stability analysis for preventing landslides using deep learning. IEEE Access, 9, 150623-150636. https://doi.org/10.1109/ACCESS.2021.3123501.

Bai, W., Li, R., Pan, J., Li, R., Wang, L., & Yang, Z. (2023). Measured Rainfall Infiltration and the Infiltration Interface Effect on Double-Layer Loess Slope. Water, 15(14), 2505. https://doi.org/10.3390/w15142505

Baker, R., Leshchinsky, D., (2001). Spatial Distribution of Safety Factors. Journal of Geotechnical and Geoenvironmental Engineering, 127(2), 135-145. https://doi.org/10.1061/(ASCE)1090-0241(2001)127:2(135)

Barron, J. T. (2019). A general and adaptive robust loss function. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4331-4339). https://doi.org/10.48550/arXiv.1701.03077

Baum, R.L., Savage, W.Z., and Godt, J.W. (2008), TRIGRS—A Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0: U.S. Geological Survey Open-File Report, 2008-1159, 75 p. (Assessed September 23, 2024 at https://pubs.usgs.gov/of/2008/1159/)

Bednarczyk, Z. (2018). Identification of flysch landslide triggers using conventional and 'nearly real-time' monitoring methods – An example from the Carpathian Mountains, Poland. Engineering Geology, 244, 41-56. https://doi.org/10.1016/j.enggeo.2018.07.012

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of K-fold cross-validation. Journal of Machine Learning Research, 5, 1089–1105.

Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. Hydrological Sciences Bulletin, 24(1), 43-69. https://doi.org/10.1080/02626667909491834

Bezak, N., Jemec Auflič, M., & Mikoš, M. (2019). Application of hydrological modelling for temporal prediction of rainfall-induced shallow landslides. Landslides, 16(7), 1273-1283. https://doi.org/10.1007/s10346-019-01169-9

Bicocchi, G., Tofani, V., D'Ambrosio, M., Tacconi-Stefanelli, C., Vannocci, P., Casagli, N., Lavorini, G., Trevisani, M., & Catani, F. (2019). Geotechnical and hydrological characterization of hillslope deposits for regional landslide prediction modeling. Bulletin of Engineering Geology and the Environment, 78(7), 4875-4891. https://doi.org/10.1007/s10064-018-01449-z

Birkinshaw, S. J., & Ewen, J. (2000). Nitrogen transformation component for SHETRAN catchment nitrate transport modelling. Journal of Hydrology, 230(1), 1-17. https://doi.org/10.1016/S0022-1694(00)00174-8

Bishop, A. W. (1955). The use of the Slip Circle in the Stability Analysis of Slopes. Geotechnique, 5(1), 7-17. https://doi.org/10.1680/geot.1955.5.1.7

Bland, J. M., & Altman, D. G. (1996). Statistics notes: Measurement error. BMJ, 312(7047), 1654. https://doi.org/10.1136/bmj.313.7059.744

Bontemps, N., Lacroix, P., Larose, E., Jara, J., & Taipe, E. (2020). Rain and small earthquakes maintain a slow-moving landslide in a persistent critical state. Nature Communication, 11(1), 780. https://doi.org/10.1038/s41467-020-14445-3

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994) Time Series Analysis; Forecasting and Control. 3rd Edition, Prentice Hall, Englewood Cliff, New Jersey.

Brand, E. W. (1984). Landslides in Southeast Asia; a state-of-the-art report. Proceedings of the 4th International Symposium on Landslides, Toronto, 1, 17-59

Breglio, G., Bernini, R., Berruti, G. M., Bruno, F. A., Buontempo, S., Campopiano, S., Catalano, E., Consales, M., Coscetta, A., Cutolo, A., Cutolo, M. A., Di Palma, P., Esposito, F., Fienga, F., Giordano, M., Iele, A., Iadicicco, A., Irace, A., Janneh, M., . . . Cusano, A. (2023). Innovative Photonic Sensors for Safety and Security, Part III: Environment, Agriculture and Soil Monitoring. Sensors, 23(6), 3187. https://www.mdpi.com/1424-8220/23/6/3187

Brunetti, M. T., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F. (2010). Rainfall thresholds for the possible occurrence of landslides in Italy. Natural Hazards and Earth System Sciences, 10(3), 447-458.https://doi.org/10.5194/nhess-10-447-2010

Cacciari, I., & Ranfagni, A. (2024). Hands-On Fundamentals of 1D Convolutional Neural Networks—A Tutorial for Beginner Users. Applied Sciences, 14(18), 8500. https://www.mdpi.com/2076-3417/14/18/8500

Caine, N. (1980). The rainfall intensity-duration control of shallow landslides and debris flows. Geografiska Annaler Series A, Physical Geography, 62(1-2), 23-27. https://doi.org/10.1080/04353676.1980.11879996

Calvello, M., Cascini, L., & Grimaldi, G. M. (2009). Displacement scenarios of a rainfall-controlled slow moving active slide in stiff clays. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 3(3), 116-125. https://doi.org/10.1080/17499510902793850

Cao, Y., Yin, K., Alexander, D. E., & Zhou, C. (2016). Using an extreme learning machine to predict the displacement of step-like landslides in relation to controlling factors. Landslides, 13(4), 725-736. https://doi.org/10.1007/s10346-015-0596-z

Cao, Y., Yin, K., Zhou, C., & Ahmed, B. (2020). Establishment of landslide groundwater level prediction model based on GA-SVM and influencing factor analysis. Sensors, 20(3), 845. https://doi.org/10.3390/s20030845.

Carsel, R. F., Parrish, R. S., (1988). Developing joint probability distributions of soil water retention characteristics. Water Resources Research, 24(5), 755-769. https://doi.org/10.1029/WR024i005p00755

Ceccato, F., Yerro, A., Girardi, V., & Simonini, P. (2021). Two-phase dynamic MPM formulation for unsaturated soil. Computers and Geotechnics, 129, 103876. https://doi.org/https://doi.org/10.1016/j.compgeo.2020.103876

Ceccato, Francesca & Girardi, Veronica & Yerro, Alba & Simonini, Paolo. (2019). Evaluation Of Dynamic Explicit MPM Formulations For Unsaturated Soils. Fundamentals and Applications, PARTICLES, 167064. Retrieved from http://hdl.handle.net/2117/186457

Chae, B.-G., Park, H.-J., Catani, F., Simoni, A., & Berti, M. (2017). Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. Geosciences Journal, 21(6), 1033-1070. https://doi.org/10.1007/s12303-017-0034-4

Chan, C. C. L., Chiu, D. S. M., Lo, F. L. C., Kwan, J. S. H., Lee, S. W., & Leung, A. C. O. (2021). Back analyses of two deep excavations in Hong Kong using the Mohr-Coulomb model with linear elasticity and the Hardening Soil model. ISSMGE International Journal of Geoengineering Case Histories, 7(1), 137–163 .

Chen, G., Zhang, G., Lu, S., & Wang, X. (2018b). An attempt to quantify the lag time of hydrodynamic action based on the long-term monitoring of a typical landslide, Three Gorges, China. Mathematical Problems in Engineering, 2018, 5958436. https://doi.org/10.1155/2018/5958436.

Chen, X., Liu, W., Mao, X., & Yang, Z. (2020). Distributed high-dimensional regression under a quantile loss function. Journal of Machine Learning Research, 21(182), 1-43. https://doi.org/10.48550/arXiv.1906.05741

Chen, X., Zhang, L., Chen, L., Li, X., & Liu, D. (2019). Slope stability analysis based on the Coupled Eulerian-Lagrangian finite element method. Bulletin of Engineering Geology and the Environment, 78(6), 4451-4463. https://doi.org/10.1007/s10064-018-1413-4

Chen, Y., Irfan, M., Uchimura, T., & Zhang, K. (2018a). Feasibility of using elastic wave velocity monitoring for early warning of rainfall-induced slope failure. Sensors, 18(4), 997. https://doi.org/10.3390/s18040997

Chen, Z.-Y., & Morgenstern, N. R. (1983). Extensions to the generalized method of slices for stability analysis. Canadian Geotechnical Journal, 20(1), 104-119. https://doi.org/10.1139/t83-010

Cheng, Y. M., Lansivaara, T., & Wei, W. B. (2007). Two-dimensional slope stability analysis by limit equilibrium and strength reduction methods. Computers and Geotechnics, 34(3), 137-150. https://doi.org/10.1016/j.compgeo.2006.10.011

Cheng, y., Choi, Y.-k., Yeung, A., Tham, L., Au, A., Wei, W., & Chen, J. (2009). New soil nail material - Pilot study of grouted GFRP pipe nails in Korea and Hong Kong. Journal of Materials in Civil Engineering, 21 (3). https://doi.org/10.1061/(ASCE)0899-1561(2009)21:3 (93)

Chhorn, C., Yune, C. Y., Kim, G., & Lee, S. W. (2016). Study for assessment of rainfall duration inducing landslides. Proceedings of the Institution of Civil Engineers: Municipal Engineer, 169(2), 100-108. https://doi.org/10.1680/jmuen.14.00028

Chiang, J. L., Kuo, C. M., & Fazeldehkordi, L. (2022). Deep learning is used to formulate the landslide rainfall threshold of the potential large-scale landslide. Water, 14(20), 3320. https://doi.org/10.3390/w14203320

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. arXiv preprint arXiv:1409.1259 .

Cho, S. E. (2016). Stability analysis of unsaturated soil slopes considering water-air flow caused by rainfall infiltration. Engineering geology, 211, 184-197. https://doi.org/10.1016/j.enggeo.2016.07.008

Chu, M., Patton, A., Roering, J., Siebert, C., Selker, J., Walter, C., & Udell, C. (2021). SitkaNet: A low-cost, distributed sensor network for landslide monitoring and study. HardwareX, 9, e00191. https://doi.org/10.1016/j.ohx.2021.e00191

Chuan, W., Wen-Qiao, W., Guo-Jun, W., & Xiao-Ming, W. (2013). Multiple parameter monitoring system for landslide. International Journal on Smart Sensing and Intelligent Systems, 6(3), 1200-1229. https://doi.org/10.21307/ijssis-2017-587

Chung, C. C., Tran, V. N., & Azhar, M. (2022). Guidelines from direct shear modeling in centrifuge for TDR landslide monitoring. Engineering Geology, 310, 106870. https://doi.org/10.1016/j.enggeo.2022.106870

Chung, C.,, Lin, C., (2019).A comprehensive framework of TDR landslide monitoring and early warning substantiated by field examples, Engineering Geology, 262,105330.https://doi.org/10.1016/j.enggeo.2019.105330

Cohen, E. (1998). An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. Measurement Science and Technology, 9(6), 022. https://doi.org/10.1088/0957-0233/9/6/022

Collins, B. D., & Znidarcic, D. (2004). Stability Analyses of Rainfall Induced Landslides. Journal of Geotechnical and Geoenvironmental Engineering, 130(4), 362-372. https://doi.org/doi:10.1061/(ASCE)1090-0241(2004)130:4 (362)

Contextily. (2023). Contextily Documentation. Retrieved from https://github.com/geopandas/contextily

Cuomo, S., & Di Perna, A. 2020. Coupled hydro-mechanical modelling of a 1995 Hong Kong landslide. E3S Web Conference, 195, 01028. https://doi.org/10.1051/e3sconf/202019501028

Cuomo, S., Di Perna, A., & Martinelli, M. (2021). Modelling the spatio-temporal evolution of a rainfall-induced retrogressive landslide in an unsaturated slope. Engineering Geology, 294, 106371. https://doi.org/10.1016/j.enggeo.2021.106371

Dai, C., Li, W., Wang, D., Lu, H., Xu, Q., & Jian, J. (2021). Active landslide detection based on Sentinel-1 data and InSAR technology in Zhouqu county, Gansu province, Northwest China. Journal of Earth Science, 32, 1092-1103. https://doi.org/10.1007/s12583-020-1380-0

Dai, Y., Dai, W., Yu, W., & Bai, D. (2022). Determination of Landslide Displacement Warning Thresholds by Applying DBA-LSTM and Numerical Simulation Algorithms. Applied Sciences, 12(13), 6690. https://www.mdpi.com/2076-3417/12/13/6690

Das, P., Patwa, D., G, V., & Bharat, T. V. (2022). Influencing factors on the simulation of rainfall-induced landslide prediction based on case study. Bulletin of Engineering Geology and the Environment, 81(5), 194. https://doi.org/10.1007/s10064-022-02682-3

Dassanayake, S. M., Mousa, A., Fowmes, G. J., Susilawati, S., & Zamara, K. (2023). Forecasting the moisture dynamics of a landfill capping system comprising different geosynthetics: A NARX neural network approach. Geotextiles and Geomembranes, 51(1), 282-292. https://doi.org/10.1016/j.geotexmem.2022.08.005

Davar, S., Nobahar, M., Khan, M. S., & Amini, F. (2022). The Development of PSO-ANN and BOA-ANN Models for Predicting Matric Suction in Expansive Clay Soil. Mathematics, 10(16), 2825. https://www.mdpi.com/2227-7390/10/16/2825

Davis, S. G. (1952). The Geology of Hong Kong. The Government Printer, Hong Kong.

De Backer, M., Ghouch, A. E., & Van Keilegom, I. (2019). An adapted loss function for censored quantile regression. Journal of the American Statistical Association, 114(527), 1126-1137. https://doi.org/10.1080/01621459.2018.1469996

De Graff, J. V. (2011). Perspectives for systematic landslide monitoring. Environmental and Engineering Geoscience, 17(1), 67-76. https://doi.org/10.2113/gseegeosci.17.1.67

De Luca, D. L., & Versace, P. (2017). A comprehensive framework for empirical modeling of landslides induced by rainfall: the Generalized FLaIR Model (GFM). Landslides, 14(3), 1009-1030. https://doi.org/10.1007/s10346-016-0768-5

De Souza, F. T., & Ebecken, N. F. F. (2012). A Data Based Model to Predict Landslide Induced by Rainfall in Rio de Janeiro City. Geotechnical and Geological Engineering, 30(1), 85-94. https://doi.org/10.1007/s10706-011-9451-8

Deng, L., Yuan, H., Chen, J., Fu, M., Chen, Y., Li, K., Yu, M., & Chen, T. (2021). Experimental investigation on integrated subsurface monitoring of soil slope using acoustic emission and mechanical measurement. Applied Sciences, 11(16), 7173. https://doi.org/10.3390/app11167173

Dou, J., Xing, K., Wang, L., Guo, H., Wang, D., Peng, Y., Xiang, X., Ciren, D., Zhang, S., Zhang, L., & Peng, B. (2025). Air-Space-Ground Synergistic Observations for Rapid Post-Seismic Disaster Assessment of 2025 Ms6.8 Xigazê Earthquake, Xizang. Journal of Earth Science. https://doi.org/10.1007/s12583-025-0160-2

Draper, N. R., & Smith, H. (1998). Applied Regression Analysis (3rd ed.). John Wiley & Sons. https://doi.org/10.1002/9781118625590

Du, W., Sheng, Q., Fu, X., Chen, J., Wei, P., Zhou, Y. (2023). Post-failure analysis of landslide blocking river using the two-phase double-point material point method: a case of western Hubei, China. Bulletin of Engineering Geology and the Environment , 82(3), 98. https://doi.org/10.1007/s10064-023-03122-6

Duncan, J. M. (1996). State of the Art: Limit Equilibrium and Finite-Element Analysis of Slopes. Journal of Geotechnical Engineering, 122(7), 577-596. https://doi.org/doi:10.1061/(ASCE)0733-9410(1996)122:7 (577)

Egrioglu, E., Bas, E., & Karahasan, O. (2023). Winsorized dendritic neuron model artificial neural network and a robust training algorithm with Tukey's biweight loss function based on particle swarm optimization. Granular Computing, 8(3), 491-501. https://doi.org/10.1007/s41066-022-00345-y

Emmanuel Emeka, A., Ugochukwu Nnatuanya, O., Samad Opeyemi, A., Chioma Margaret, A., & Michael Oduh, O. (2021). Geotechnical and Geological Analysis of Amuzukwu Landslide. American Journal of Environmental Protection, 10(4), 84-89. https://doi.org/10.11648/j.ajep.20211004.12

ETS-Testconsult Ltd. (2023). The direct shear test (small shear box apparatus). 8/F Block B, Veristong Industrial Center, Au Pui Wan Street, Fo Tan, Hong Kong.

Eyo, E., Musa, T. A., Omar, K. M., Idris, K. M., Bayrak, T., Onuigbo, I., & Opaluwa, Y. D. (2014). Application of Low-Cost GPS Tools and Techniques for Landslide Monitoring: A Review. Jurnal Teknologi, 71(4). https://doi.org/10.11113/jt.v71.3828

Feng, Y., Yan, F., Wu, L., Lu, G., & Liu, T. (2023). Numerical Analyses of Slope Stability Considering Grading and Seepage Prevention. Water, 15(9), 1745. https://doi.org/10.3390/w15091745

Filipović, N., Brdar, S., Mimić, G., Marko, O., & Crnojević, V. (2022). Regional soil moisture prediction system based on Long Short-Term Memory network. Biosystems engineering, 213, 30-38. https://doi.org/10.1016/j.biosystemseng.2021.11.019

Fisher, A., Rudin, C., & Dominici, F. (2019). All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. Journal of machine learning research: JMLR, 20, 177. https://doi.org/10.48550/arXiv.1801.01489

Folino, G., Guarascio, M., & Chiaravalloti, F. (2023). Learning ensembles of deep neural networks for extreme rainfall event detection. Neural Computing and Applications, 35(14), 10347-10360. https://doi.org/10.1007/s00521-023-08238-0

Formetta, G., & Capparelli, G. (2019). Quantifying the three-dimensional effects of anisotropic soil horizons on hillslope hydrology and stability. Journal of Hydrology, 570, 329-342. https://doi.org/https://doi.org/10.1016/j.jhydrol.2018.12.064

Fredlund, D. G. (1998). Bringing Unsaturated Soil Mechanics into Engineering Practice (pp. 1-51). Keynote Address presented at the Second International Conference on Unsaturated Soil, Beijing, China, August 27-30, 1998. University of Saskatchewan, Saskatoon, Canada.

Fredlund, D. G., & Krahn, J. (1977). Comparison of slope stability methods of analysis. Canadian Geotechnical Journal, 14(3), 429-439. https://doi.org/10.1139/t77-045

Fredlund, D. G., & Rahardjo, H. (1993). Soil mechanics for unsaturated soils. John Wiley & Sons. https://doi.org/10.1002/9780470172759

Fredlund, D. G., Morgenstern, N. R., & Widger, R. A. (1978). The shear strength of unsaturated soils. Canadian Geotechnical Journal, 15(3), 313-321. https://doi.org/10.1139/t78-029

Fredlund, D. G., Xing, A. (1994). Equations for the soil-water characteristic curve. Canadian Geotechnical Journal, 31(4), 521-532. https://doi.org/10.1139/t94-061

Fredlund, D., Xing, A., Fredlund, M., Barbour, S. (1996). The relationship of the unsaturated soil shear strength to the soil-water characteristic curve. Canadian Geotechnical Journal, 33, 440-448. https://doi.org/10.1139/t96-065

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. Natural Hazards and Earth System Sciences, 18(8), 2161-2181. https://doi.org/10.5194/nhess-18-2161-2018.

Fugro (Hong Kong) Limited. (2025). Detailed study of the 8 September 2023 landslides on the natural hillside above Yiu Hing Road, Yiu Tung Estate, Shau Kei Wan (GEO Report No. 377). Geotechnical Engineering Office, Civil Engineering and Development Department, The Government of the Hong Kong Special Administrative Region.

Fukuhara, M., Uchimura, T., Tang, J., & Tao, S. (2022). Verify the Prediction of Slope Failure by Soil Creep With Laboratory Experiments. GEOMATE Journal, 23(98), 164-170. https://geomatejournal.com/geomate/article/view/3674

Gal, Y., & Ghahramani, Z. (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. Proceedings of The 33rd International Conference on Machine Learning, PMLR 48:1050-1059. https://doi.org/10.48550/arXiv.1506.02142

Gan, J. K., & Fredlund, D. G. (1996). Direct shear and triaxial testing of a Hong Kong soil under saturated and unsaturated conditions. Geotechnical Engineering Office, Civil Engineering Department. Retrived from https://www.cedd.gov.hk/eng/publications/geo/geo-reports/geo_rpt046/index.html

Gao, L., Zhang, L. M., & Cheung, R. W. M. (2018). Relationships between natural terrain landslide magnitudes and triggering rainfall based on a large landslide inventory in Hong Kong. Landslides, 15(4), 727-740. https://doi.org/10.1007/s10346-017-0904-x

Gao, L., Zhang, L. M., Chen, H. X., Fei, K., & Hong, Y. (2021). Topography and geology effects on travel distances of natural terrain landslides: Evidence from a large multi-temporal landslide inventory in Hong Kong. Engineering Geology, 292, 106266. https://doi.org/10.1016/j.enggeo.2021.106266

Gao, W., Dai, S., & Chen, X. (2020). Landslide prediction based on a combination intelligent method using the GM and ENN: two cases of landslides in the Three Gorges Reservoir, China. Landslides, 17(1), 111-126. https://doi.org/10.1007/s10346-019-01273-w

Gariano, S. L., Sarkar, R., Dikshit, A., Dorji, K., Brunetti, M. T., Peruccacci, S., & Melillo, M. (2019). Automatic calculation of rainfall thresholds for landslide occurrence in Chukha Dzongkhag, Bhutan. Bulletin of Engineering Geology and the Environment, 78(6), 4325-4332. https://doi.org/10.1007/s10064-018-1415-2

Ge, Q., Sun, H., Liu, Z., & Wang, X. (2023). A data-driven intelligent model for landslide displacement prediction. Geological Journal, 58(6), 2211-2230. https://doi.org/10.1002/gj.4675

GeoPandas. (2023). GeoPandas User Guide. Retrieved from https://geopandas.org

GEOSLOPE International Ltd. (2020). Stress-strain modeling with GeoStudio 2021 [Technical Report]. Calgary, Alberta, Canada.

GEOSLOPE International Ltd. (2021). Parameterizing the Hardening Soil Model [Technical Report]. Calgary, Alberta, Canada. Retrieved from https://www.seequent.com/sigma-w-material-model-series-hardening-soil-model/

Geo-Studio (2023a). User's Guide for SEEP/W, version 2023. Geo-slope International Ltd, Canada. Retrieved from https://www.geoslope.support/kb/article/10-geostudio-reference-manuals/

Geo-Studio (2023b). User's Guide for SLOPE/W, version 2023. Geo-slope International Ltd, Canada. Retrieved from https://www.geoslope.support/kb/article/10-geostudio-reference-manuals/

Geo-Studio (2023c). User's Guide for SIGMA/W, version 2023. Geo-slope International Ltd, Canada. Retrieved from https://www.geoslope.support/kb/article/10-geostudio-reference-manuals/

Geotechnical Control Office (GCO) (1982). Mid-levels study : Report on Geology, Hydrology and Soil Properities. Geotechnical Control Office, Hong Kong, 1-265.

Geotechnical Control Office (GCO) (1984). Geotechnical Manual for Slopes. (Second Edition). Geotechnical Control Office, Hong Kong, 1-295.

Geotechnical Control Office (GCO) (2000). Guide to Retaining Wall Design (Geoguide 1). Geotechnical Control Office, Civil Engineering Services Department, Hong Kong Government.

Geotechnical Engineering Office (1993). Guide to Retaining Wall Design (Geoguide 1). Geotechnical Engineering Office, Hong Kong Government, Hong Kong.

Geotechnical Engineering Office (2010a). Landslide Assessment and Monitoring Works at Four Selected Sites: Detailed Analysis of Data Obtained and the Slope Behavior - Pa Mei Study Area (Report No. CE4/2005(GE)).

Geotechnical Engineering Office (2010b). Landslide Assessment and Monitoring Works at Four Selected Sites: Detailed Analysis of Data Obtained and the Slope Behavior – Tung Chung Study Area (Report No. CE4/2005(GE)).

Geotechnical Engineering Office (2010c). Landslide Assessment and Monitoring Works at Four Selected Sites: Detailed Analysis of Data Obtained and the Slope Behavior – Tsing Shan Foothills Study Area (Report No. CE4/2005(GE)).

Geotechnical Engineering Office (2016). Natural Terrain Landslide Hazards in Hong Kong. Civil Engineering and Development Department, The Government of the Hong Kong Special Administrative Region, Hong Kong, PP.(33)

Gerscovich, D. M. S., Vargas Jr, E. A., & de Campos, T. M. P. (2006). On the evaluation of unsaturated flow in a natural slope in Rio de Janeiro, Brazil. Engineering Geology, 88(1-2), 23-40. https://doi.org/10.1016/j.enggeo.2006.07.008

Ghaderi, A., Abbaszadeh Shahri, A., & Larsson, S. (2022). A visualized hybrid intelligent model to delineate Swedish fine-grained soil layers using clay sensitivity. CATENA, 214, 106289. https://doi.org/https://doi.org/10.1016/j.catena.2022.106289

Gian Quoc, A., Nguyen Dinh, C., Tran Duc, N., Tran Duc, T., & Kumbesan, S. (2018). Wireless technology for monitoring site-specific landslide in Vietnam. International Journal of Electrical and Computer Engineering, 8, https://doi.org/4448–4455.10.11591/ijece.v8i6.pp4448-4455

Gillies, S., van der Wel, C., Van den Bossche, J., Taves, M. W., Arnott, J., Ward, B. C., & others. (2025). Shapely (Version 2.1.1). https://doi.org/10.5281/zenodo.5597138

Giri, P., Ng, K., & Phillips, W. (2019). Wireless Sensor Network System for Landslide Monitoring and Warning. IEEE Transactions on Instrumentation and Measurement, 68(4), 1210-1220, 8444699. https://doi.org/10.1109/TIM.2018.2861999

Giuseppe, F., Simoni, S., Godt, J. W., Lu, N., & Rigon, R. (2016). Geomorphological control on variably saturated hillslope hydrology and slope instability. Water Resources Research, 52(6), 4590-4607. https://doi.org/10.1002/2015WR017626

Google Research. (2024). Google Colaboratory. Retrieved from https://colab.research.google.com

Granata, F., Di Nunno, F., Najafzadeh, M., & Demir, I. (2023). A Stacked Machine Learning Algorithm for Multi-Step Ahead Prediction of Soil Moisture. Hydrology, 10(1), 1. https://www.mdpi.com/2306-5338/10/1/1

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Networks, 18(5-6), 602-610. https://doi.org/10.1016/j.neunet.2005.06.042

Gupta, D., Hazarika, B. B., & Berlin, M. (2020). Robust regularized extreme learning machine with asymmetric Huber loss function. Neural Computing and Applications, 32(16), 12971-12998. https://doi.org/10.1007/s00521-020-04741-w

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 31(1), 181-216. https://doi.org/10.1016/S0169-555X(99)00078-1

Guzzetti, F., Melillo, M., & Mondini, A. C. (2025). Landslide predictions through combined rainfall threshold models. Landslides, 22(1), 137-147. https://doi.org/10.1007/s10346-024-02340-7

Han, H., Shi, B., & Zhang, L. (2021). Prediction of landslide sharp increase displacement by SVM with considering hysteresis of groundwater change. Engineering Geology, 280, 105876. https://doi.org/10.1016/j.enggeo.2020.105876

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362. https://doi.org/10.1038/s41586-020-2649-2

He, J., Wang, S., Liu, H., Nguyen, V., & Han, W. (2021). The critical curve for shallow saturated zone in soil slope under rainfall and its prediction for landslide characteristics. Bulletin of Engineering Geology and the Environment, 80(3), 1927-1945. https://doi.org/10.1007/s10064-020-02016-1

He, X., Hong, Y., Vergara, H., Zhang, K., Kirstetter, P.-E., Gourley, J. J., Zhang, Y., Qiao, G., & Liu, C. (2016). Development of a coupled hydrological-geotechnical framework for rainfall-induced landslides prediction. Journal of Hydrology, 543, 395-405. https://doi.org/10.1016/j.jhydrol.2016.10.016

Head, T., MechCoder, Louppe, G., Shcherbatyi, I., fcharras, Ambrogio, C., & htsharkey. (2018). Scikit-Optimize: Efficient and reproducible hyperparameter optimization in Python.

Heber Green, W., & Ampt, G. A. (1911). Studies on Soil Phyics. The Journal of Agricultural Science, 4(1), 1-24. https://doi.org/10.1017/S0021859600001441

Hemalatha, T., Ramesh, M. V., & Rangan, V. P. (2019). Effective and accelerated forewarning of landslides using wireless sensor networks and machine learning. IEEE Sensors Journal, 19(21), 9964-9975, 8760567. https://doi.org/10.1109/JSEN.2019.2928358

Hencher, S. R., & Lee, S. G. (2010). Landslide mechanisms in Hong Kong. Geological Society, London, Engineering Geology Special Publications, 23(1), 77-103. https://doi.org/10.1144/EGSP23.6

Hindriks, R., Adhikari, M. H., Murayama, Y., Ganzetti, M., Mantini, D., Logothetis, N. K., & Deco, G. (2016). Can sliding-window correlations reveal dynamic functional connectivity in resting-state fMRI?. Neuroimage, 127, 242-256. https://doi.org/10.1016/j.neuroimage.2015.11.055

Ho, D., & Fredlund, D. (1982). A Multistage Triaxial Test for Unsaturated Soils. Geotechnical Testing Journal, 5, 18-25 .

Ho, H. Y., & Roberts, K. J. (2016). Guidelines for natural terrain hazard studies (GEO Report No. 138, 2nd ed.). Geotechnical Engineering Office, Civil Engineering and Development Department, The Government of the Hong Kong Special Administrative Region.

Ho, J.-Y., & Lee, K. T. (2017). Performance evaluation of a physically based model for shallow landslide prediction. Landslides, 14(3), 961-980. https://doi.org/10.1007/s10346-016-0762-y
Ho, J.-Y., Lee, K. T., Chang, T.-C., Wang, Z.-Y., & Liao, Y.-H. (2012). Influences of spatial distribution of soil thickness on shallow landslide prediction. Engineering Geology, 124, 38-46. https://doi.org/10.1016/j.enggeo.2011.09.013
Hobbs, P. R. N., Jones, L. D., Kirkham, M. P., Pennington, C. V. L., Morgan, D. J. R., & Dashwood, C. (2019). Coastal landslide monitoring at Aldbrough, East Riding of Yorkshire, UK. Quarterly Journal of Engineering Geology and Hydrogeology, 53(1), 101-116. https://doi.org/10.1144/qjegh2018-210
Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735
Holt, C.C. (1957) Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. ONR Memorandum, Vol. 52, Carnegie Institute of Technology, Pittsburgh. Available from the Engineering Library, University of Texas, Austin.
Hong Kong Observatory (HKO) (2015). Thunderstorm, rainstorm, flood and landslip warning services in Hong Kong. Hong Kong Observatory. https://www.hko.gov.hk
Hong Kong Observatory (HKO) (2019). Climate of Hong Kong, 2019. Available at: https://www.hko.gov.hk/cis/climahk_e.htm
Hong, M., Kim, J., & Jeong, S. (2018). Rainfall intensity-duration thresholds for landslide prediction in South Korea by considering the effects of antecedent rainfall. Landslides, 15(3), 523-534. https://doi.org/10.1007/s10346-017-0892-x
Hong, Y., Adler, R., & Huffman, G. (2006). Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. Geophysical Research Letters, 33(22). https://doi.org/https://doi.org/10.1029/2006GL028010
Hosmer, D.W., & Lemeshow, S. (2000). Applied logistic regression, 2nd edn. Wiley, New York
Hu, Y., Zhou, J., Li, J., Ma, J., Hu, Y., Lu, F., He, X., Wen, J., & Cheng, T. (2022). Tipping-bucket self-powered rain gauge based on triboelectric nanogenerators for rainfall measurement. Nano Energy, 98, 107234. https://doi.org/10.1016/j.nanoen.2022.107234
Huang, D., He, J., Song, Y., Guo, Z., Huang, X., & Guo, Y. (2022b). Displacement Prediction of the Muyubao Landslide Based on a GPS Time-Series Analysis and Temporal Convolutional Network Model. Remote Sensing, 14(11), 2656. https://www.mdpi.com/2072-4292/14/11/2656
Huang, G. (2021). Missing data filling method based on linear interpolation and lightgbm. Journal of Physics: Conference Series, 1754 (1), 012187. https://doi.org/10.1088/1742-6596/1754/1/012187
Huang, G., Du, S., & Wang, D. (2023). GNSS techniques for real-time monitoring of landslides: a review. Satellite Navigation, 4(1), 5. https://doi.org/10.1186/s43020-023-00095-5
Huang, J., Ju, N. P., Liao, Y. J., & Liu, D. D. (2015). Determination of rainfall thresholds for shallow landslides by a probabilistic and empirical method. Natural Hazards and Earth System Sciences, 15(12), 2715-2723.https://doi.org/10.5194/nhess-15-2715-2015, 2015.
Huang, J., Wu, X., Ling, S., Li, X., Wu, Y., Peng, L., & He, Z. (2022a). A bibliometric and content analysis of research trends on GIS-based landslide susceptibility from 2001 to 2020. Environmental Science and Pollution Research, 29(58), 86954-86993. https://doi.org/10.1007/s11356-022-23732-z
Huang, W., Loveridge, F., Milledge, D., & Ji, J. (2025). Boundary effect correction of the infinite slope model for saturated and unsaturated soils. Journal of Geotechnical and Geoenvironmental Engineering, 151(3), Article 02824002. https://doi.org/10.1061/JGGEFK.GTENG-13137
Huang, Y., & He, Z. (2023). Rainfall-oriented resilient design for slope system: Resilience-enhancing strategies. Soils and Foundations, 63(2), 101297. https://doi.org/https://doi.org/10.1016/j.sandf.2023.101297
Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. Landslides, 11(2), 167-194. https://doi.org/10.1007/s10346-013-0436-y
Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), 90-95. https://doi.org/10.1109/MCSE.2007.55
Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts.
Jeong, S., Ko, J., & Kim, J. (2020). The effectiveness of a wireless sensor network system for landslide monitoring. IEEE Access, 8, 8073-8086, 8936371. https://doi.org/10.1109/ACCESS.2019.2960570
Jiang, S., Ma, J., Liu, Z., & Guo, H. (2022). Scientometric analysis of artificial intelligence (AI) for geohazard research. Sensors, 22(20), 7814. https://doi.org/10.3390/s22207814

Jin, Y.-F., Yin, Z.-Y., Shen, S.-L., & Hicher, P.-Y. (2016). Selection of sand models and identification of parameters using an enhanced genetic algorithm. International Journal for Numerical and Analytical Methods in Geomechanics, 40(8), 1219–1240.

Johari, A., & Hooshmand Nejad, A. (2018). An Approach to Estimate Wetting Path of Soil–Water Retention Curve from Drying Path. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 42(1), 85-89. https://doi.org/10.1007/s40996-017-0074-z

Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Garcia Badaracco, A., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, & Bilogur, A., ... Leblanc, F. (2020). *geopandas/geopandas: v0.8.1* (Version 0.8.1). Zenodo. https://doi.org/10.5281/zenodo.3946761

Khalil, S. A., & Shaffie, A. M. (2013). A comparative study of total, direct and diffuse solar irradiance by using different models on horizontal and inclined surfaces for Cairo, Egypt. Renewable and Sustainable Energy Reviews, 27, 853-863. https://doi.org/https://doi.org/10.1016/j.rser.2013.06.038

Khan, M. I., & Wang, S. (2021). Slope Stability Analysis to Correlate Shear Strength with Slope Angle and Shear Stress by Considering Saturated and Unsaturated Seismic Conditions. Applied Sciences, 11(10), 4568. https://www.mdpi.com/2076-3417/11/10/4568

KhangPham (2023).Overfitting,Generalization&theBias-VarianceTradeoff. Retrievedfrom https:// medium. com/@ khang. pham. exxact/ overfittinggeneraliza tion- the- bias- varia nce- trade off- 5800f 8c2200

Khoa, V. V., & Takayama, S. (2018). Wireless sensor network in landslide monitoring system with remote data management. Measurement: Journal of the International Measurement Confederation, 118, 214-229. https://doi.org/10.1016/j.measurement.2018.01.002

Khorasani, M., Abdou, M., Fern´andez, J.H. (2022) Web Application Development with Streamlit: Develop and Deploy Secure and Scalable Web Applications to the Cloud Using a Pure Python Framework. https://doi.org/10.1007/978-1-4842- 8111-6

Kim, H.-J., Mawuntu, K. B. A., Park, T.-W., Kim, H.-S., Park, J.-Y., & Jeong, Y.-S. (2023). Spatial Autocorrelation Incorporated Machine Learning Model for Geotechnical Subsurface Modeling. Applied Sciences, 13(7), 4497. https://doi.org/10.3390/app13074497

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. Mechanical Systems and Signal Processing, 151, 107398. https://doi.org/10.1016/j.ymssp.2020.107398

Kirkby, M. (1975). Hydrograph modeling strategies. Process in physical and human geography, 69-90. Retrieved from https://xueshu.baidu.com/usercenter/paper/show?paperid=ea42e2100aa6add49430a8df1b3efba9

Kong, Q., Chen, H., Mo, Y.-l., & Song, G. (2017). Real-Time Monitoring of Water Content in Sandy Soil Using Shear Mode Piezoceramic Transducers and Active Sensing—A Feasibility Study. Sensors, 17(10), 2395. https://www.mdpi.com/1424-8220/17/10/2395

Krahn, J. (2003). The 2001 RM Hardy Lecture: The limits of limit equilibrium analyses. Canadian Geotechnical Journal, 40(3), 643-660. https://doi.org/10.1139/t03-024

Krkač, M., Bernat Gazibara, S., Arbanas, Ž., Sečanj, M., & Mihalić Arbanas, S. (2020). A comparative study of random forests and multiple linear regression in the prediction of landslide velocity. Landslides, 17(11), 2515-2531. https://doi.org/10.1007/s10346-020-01476-6

Krkač, M., Špoljarić, D., Bernat, S., & Arbanas, S. M. (2017). Method for prediction of landslide movements based on random forests. Landslides, 14(3), 947-960. https://doi.org/10.1007/s10346-016-0761-z

Kuang, K. S. C. (2018). Wireless chemiluminescence-based sensor for soil deformation detection. Sensors and Actuators, A: Physical, 269, 70-78. https://doi.org/10.1016/j.sna.2017.11.017

Kumar, D., Roshni, T., Singh, A., Jha, M. K., & Samui, P. (2020a). Predicting groundwater depth fluctuations using deep learning, extreme learning machine and Gaussian process: a comparative study. Earth Science Informatics, 13(4), 1237-1250. https://doi.org/10.1007/s12145-020-00508-y

Kumar, S., Duttagupta, S., Rangan, V. P., & Ramesh, M. V. (2020b). Reliable network connectivity in wireless sensor networks for remote monitoring of landslides. Wireless Networks, 26(3), 2137-2152. https://doi.org/10.1007/s11276-019-02059-7

Lam, A. K. M. (2018). An engineering solution for a hillside project in Hong Kong. Geotechnical Research, 5(3), 170–181. https://doi.org/10.1680/jgere.18.00008

Lam, L., Fredlund, D. G., & Barbour, S. L. (1987). Transient seepage model for saturated–unsaturated soil systems: a geotechnical engineering approach. Canadian Geotechnical Journal, 24(4), 565-580. https://doi.org/10.1139/t87-071

Lapenna, V., & Perrone, A. (2022). Time-Lapse Electrical Resistivity Tomography (TL-ERT) for Landslide Monitoring: Recent Advances and Future Directions. Applied Sciences, 12(3), 1425. https://www.mdpi.com/2076-3417/12/3/1425

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324. https://doi.org/10.1109/5.726791

Lee, K. T., & Ho, J.-Y. (2009). Prediction of landslide occurrence based on slope-instability analysis and hydrological model simulation. Journal of Hydrology, 375(3), 489-497. https://doi.org/10.1016/j.jhydrol.2009.06.053

Lee, M. L., Ng, K. Y., Huang, Y. F., & Li, W. C. (2014). Rainfall-induced landslides in Hulu Kelang area, Malaysia. Natural Hazards, 70(1), 353-375. https://doi.org/10.1007/s11069-013-0814-8

Lee, S. W., Pickles, A. R., Henderson, T. O., & Li, E. S. F. (2008). Three-dimensional modelling of deep excavation in decomposed granite: Influence of small strain stiffness and presence of individual piles. In Proceedings of the HKIE Geotechnical Division Annual Seminar 2008 (pp. 171–180). Hong Kong

Lee, W.-L., Martinelli, M., & Shieh, C.-L. (2021). An Investigation of Rainfall-Induced Landslides From the Pre-Failure Stage to the Post-Failure Stage Using the Material Point Method [Original Research]. Frontiers in Earth Science, 9. https://doi.org/10.3389/feart.2021.764393

Leone, M., Principe, S., Consales, M., Parente, R., Laudati, A., Caliro, S., Cutolo, A., & Cusano, A. (2017). Fiber optic thermo-hygrometers for soil moisture monitoring. Sensors (Switzerland), 17(6), 1451. https://doi.org/10.3390/s17061451

Li, F., & Li, Y. (2023). Robust echo state network with Cauchy loss function and hybrid regularization for noisy time series prediction. Applied Soft Computing, 146, 110640. https://doi.org/10.1016/j.asoc.2023.110640

Li, W., Fang, H., Qin, G., Tan, X., Huang, Z., Zeng, F., Du, H., & Li, S. (2020). Concentration estimation of dissolved oxygen in Pearl River Basin using input variable selection and machine learning techniques. Science of The Total Environment, 731, 139099. https://doi.org/10.1016/j.scitotenv.2020.139099

Li, Z., Cheng, P., & Zheng, J. (2021). Prediction of time to slope failure based on a new model. Bulletin of Engineering Geology and the Environment, 80(7), 5279-5291. https://doi.org/10.1007/s10064-021-02234-1

Lian, C., Chen, C. L. P., Zeng, Z., Yao, W., & Tang, H. (2016). Prediction Intervals for Landslide Displacement Based on Switched Neural Networks. IEEE Transactions on Reliability, 65(3), 1483-1495. https://doi.org/10.1109/TR.2016.2570540

Liang, W.-L., & Uchida, T. (2022). Performance and topographic preferences of dynamic and steady models for shallow landslide prediction in a small catchment. Landslides, 19(1), 51-66. https://doi.org/10.1007/s10346-021-01771-w

Liao, Z., Hong, Y., Wang, J., Fukuoka, H., Sassa, K., Karnawati, D., & Fathani, F. (2010). Prototyping an experimental early warning system for rainfall-induced landslides in Indonesia using satellite remote sensing and geospatial datasets. Landslides, 7(3), 317-324. https://doi.org/10.1007/s10346-010-0219-7

Lin, Z., He, Q., Xiao, Y., Zhu, T., Yang, J., Sun, C., Zhou, Z., Zhang, H., Shen, Z., Yang, J., & Wang, Z. L. (2018). Flexible Timbo-Like Triboelectric Nanogenerator as Self-Powered Force and Bend Sensor for Wireless and Distributed Landslide Monitoring. Advanced Materials Technologies, 3(11), 1800144. https://doi.org/10.1002/admt.201800144

Liu, X., Wang, Y., & Li, D.-Q. (2020). Numerical simulation of the 1995 rainfall-induced Fei Tsui Road landslide in Hong Kong: new insights from hydro-mechanically coupled material point method. Landslides, 17(12), 2755-2775. https://doi.org/10.1007/s10346-020-01442-2

Liu, X., Wang, Y., Koo, R. C. H., & Kwan, J. S. H. (2022). Development of a slope digital twin for predicting temporal variation of rainfall-induced slope instability using past slope performance records and monitoring data. Engineering Geology, 308, 106825. https://doi.org/10.1016/j.enggeo.2022.106825

Liu, Y., Liu, D., Qin, Z., Liu, F., & Liu, L. (2016). Rainfall data feature extraction and its verification in displacement prediction of Baishuihe landslide in China. Bulletin of Engineering Geology and the Environment, 75(3), 897-907. https://doi.org/10.1007/s10064-015-0847-1

Long, J., Li, C., Liu, Y., Feng, P., & Zuo, Q. (2022). A multi-feature fusion transfer learning method for displacement prediction of rainfall reservoir-induced landslide with step-like deformation characteristics. Engineering Geology, 297, 106494. https://doi.org/10.1016/j.enggeo.2021.106494

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 .

Lu, K.-K., Yin, J.-H., & Lo, S.-C. (2018). Modeling small-strain behavior of Hong Kong CDG and its application to finite-element study of basement-raft footing. International Journal of Geomechanics, 18(9), Article 04018104

Lu, N., & Godt, J. W. (2013). Hillslope Hydrology and Stability. Cambridge University Press. https://doi.org/10.1017/CBO9781139108164

Lumb, P. B. (1975). Slope failures in Hong Kong. Quarterly Journal of Engineering Geology and Hydrogeology., 8, 31-65. https://doi.org/10.1144/GSL.QJEG.1975.008.01.02

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30. arXiv:1705.07874v1

Ma, J., Niu, X., Tang, H., Wang, Y., Wen, T., & Zhang, J. (2020). Displacement prediction of a complex landslide in the Three Gorges Reservoir Area (China) using a hybrid computational intelligence approach. Complexity, 2020, 1-15. https://doi.org/10.1155/2020/2624547

Ma, J., Tang, H., Hu, X., Bobet, A., Yong, R., & Ez Eldin, M. A. M. (2017). Model testing of the spatial–temporal evolution of a landslide failure. Bulletin of Engineering Geology and the Environment, 76(1), 323-339. https://doi.org/10.1007/s10064-016-0884-4

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier nonlinearities improve neural network acoustic models. In Proc. Icml, 30 (1), 3. https://api.semanticscholar.org/CorpusID:16489696

Marino, P., Roman Quintero, D. C., Santonastaso, G. F., & Greco, R. (2023). Prototype of an IoT-Based Low-Cost Sensor Network for the Hydrological Monitoring of Landslide-Prone Areas. Sensors, 23(4), 2299. https://doi.org/10.3390/s23042299

Marrapu, B. M., Kukunuri, A., & Jakka, R. S. (2021). Improvement in Prediction of Slope Stability & Relative Importance Factors Using ANN. Geotechnical and Geological Engineering, 39(8), 5879-5894. https://doi.org/10.1007/s10706-021-01872-2

Matyas, E. L., & Radhakrishna, H. S. (1968). Volume Change Characteristics of Partially Saturated Soils. Geotechnique, 18(4), 432-448. https://doi.org/10.1680/geot.1968.18.4.432

Maunsell Geotechnical Services Limited (2007). Detailed Study of the 21 August 2005 debris flow on the natural hillside near Fei Ngo Shan Service Reservoir. Geotechnical Engineering Office, Hong Kong, 233, 1-127.

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics, 21(2), 239–245

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., Roccati, A., & Guzzetti, F. (2018). A tool for the automatic calculation of rainfall thresholds for landslide occurrence. Environmental Modelling & Software, 105, 230-243. https://doi.org/10.1016/j.envsoft.2018.03.024

Meng, S., Shi, Z., Peng, M., Li, G., Zheng, H., Liu, L., & Zhang, L. (2024). Landslide displacement prediction with step-like curve based on convolutional neural network coupled with bi-directional gated recurrent unit optimized by attention mechanism. Engineering Applications of Artificial Intelligence, 133, 108078. https://doi.org/10.1016/j.engappai.2024.108078

Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., & Abderrahmane, B. (2020). Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. Earth-Science Reviews, 207, 103225. https://doi.org/10.1016/j.earscirev.2020.103225

Minardo, A., Zeni, L., Coscetta, A., Catalano, E., Zeni, G., Damiano, E., De Cristofaro, M., & Olivares, L. (2021). Distributed optical fiber sensor applications in geotechnical monitoring. Sensors, 21(22), 7514. https://doi.org/10.3390/s21227514

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med, 6(7), e1000097. https://doi.org/10.1371/journal.pmed.1000097

Molnar, C. (2020). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Retrieved from https://christophm.github.io/interpretable-ml-book/

Mondini, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. Nature Communications, 14(1), 2466. https://doi.org/10.1038/s41467-023-38135-y

Morbidelli, R., Saltalippi, C., Flammini, A., & Govindaraju, R. S. (2018). Role of slope on infiltration: A review. Journal of Hydrology, 557, 878-886. https://doi.org/10.1016/j.jhydrol.2018.01.019

Morgenstern, N. R., & Price, V. E. (1965). The Analysis of the Stability of General Slip Surfaces. Geotechnique, 15(1), 79-93. https://doi.org/10.1680/geot.1965.15.1.79

Motakabber, S. M. A., & Ibrahimy, M. I. (2015). An approach of differential capacitor sensor for landslide monitoring. International Journal of GEOMATE, 9(2), 1534-1537. https://doi.org/10.21660/2015.18.4363
Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research, 12(3), 513-522. https://doi.org/10.1029/WR012i003p00513
Mucchi, L., Jayousi, S., Martinelli, A., Caputo, S., Intrieri, E., Gigli, G., Gracchi, T., Mugnai, F., Favalli, M., Fornaciai, A., & Nannipieri, L. (2018). A flexible wireless sensor network based on ultra-wide band technology for ground instability monitoring. Sensors, 18(9), 2948. https://doi.org/10.3390/s18092948
Murray, C., Du Bois, N., Hollywood, L., & Coyle, D. (2023). State-of-The-Art Deep Learning Models are Superior for Time Series Forecasting and are Applied Optimally with Iterative Prediction Methods. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4361707
Natalia, L., & Yang, J. (2024). New insights into the catastrophic slope failures at Sau Mau Ping: the role of antecedent rainfall pattern. Landslides, 21(7), 1501-1514. https://doi.org/10.1007/s10346-024-02247-3
Nava, L., Carraro, E., Reyes-Carmona, C., Puliero, S., Bhuyan, K., Rosi, A., ... & Catani, F. (2023). Landslide displacement forecasting using deep learning and monitoring data across selected sites. Landslides, 20(10), 2111-2129. https://doi.org/10.1007/s10346-023-02104-9
Ng, C. W. W., & Shi, Q. (1998). A numerical investigation of the stability of unsaturated soil slopes subjected to transient seepage. Computers and Geotechnics, 22(1), 1-28. https://doi.org/10.1016/S0266-352X(97)00036-0
Ng, C. W., Wang, B., & Tung, Y. K. (2001). Three-dimensional numerical investigations of groundwater responses in an unsaturated slope subjected to various rainfall patterns. Canadian Geotechnical Journal, 38(5), 1049-1062. https://doi.org/10.1139/t01-057
Nguyen, T. S., Likitlersuang, S., & Jotisankasa, A. (2019). Influence of the spatial variability of the root cohesion on a slope-scale stability model: a case study of residual soil slope in Thailand. Bulletin of Engineering Geology and the Environment, 78(5), 3337-3351. https://doi.org/10.1007/s10064-018-1380-9
Oguz, E. A., Depina, I., & Thakur, V. (2022). Effects of soil heterogeneity on susceptibility of shallow landslides. Landslides, 19(1), 67-83, DOI: https://doi.org/10.1007/s10346-021-01738-x.
Oliveira, E. D. P., Acevedo, A. M. G., Moreira, V. S., Faro, V. P., & Kormann, A. C. M. (2022). The Key Parameters Involved in a Rainfall-Triggered Landslide. Water, 14(21), 3561. https://doi.org/10.3390/w14213561
O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., & others. (2019). KerasTuner. GitHub. https://github.com/keras-team/keras-tuner
OriginLab Corporation. (2024). OriginPro (Version 2025b) [Computer software]. https://www.originlab.com/index.aspx?go=Products/Origin/2025b&pid=5476
Pack, R. T., Tarboton, D. G., & Goodwin, C. N. (1998). The SINMAP approach to terrain stability mapping. In 8th Congress of the International Association of Engineering Geology (Vol. 21, pp. 1157-1165). Vancouver, Canada.
Pagano, L., Picarelli, L., Rianna, G., & Urciuoli, G. (2010). A simple numerical procedure for timely prediction of precipitation-induced landslides in unsaturated pyroclastic soils. Landslides, 7(3), 273-289. https://doi.org/10.1007/s10346-010-0216-x
Pearson, K. (1895). Notes on regression and inheritance in the case of two parents. Proceedings of the Royal Society of London, 58, 240–242.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res., 12(null), 2825–2830 .
Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. Journal of Hydrology, 279(1), 275-289. https://doi.org/10.1016/S0022-1694(03)00225-7
Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., & Guzzetti, F. (2017). Rainfall thresholds for possible landslide occurrence in Italy. Geomorphology, 290, 39-57. https://doi.org/10.1016/j.geomorph.2017.03.031
Petrucci, O. (2022). Landslide Fatality Occurrence: A Systematic Review of Research Published between January 2010 and March 2022. Sustainability, 14(15), 9346. https://www.mdpi.com/2071-1050/14/15/9346

Phoon, K. K., & Zhang, W. (2023). Future of machine learning in geotechnics. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 17(1), 7-22. https://doi.org/10.1080/17499518.2022.2087884

Pichorim, S. F., Gomes, N. J., & Batchelor, J. C. (2018). Two solutions of soil moisture sensing with rfid for landslide monitoring. Sensors (Switzerland), 18(2), 452. https://doi.org/10.3390/s18020452

Postill, H., Helm, P. R., Dixon, N., Glendinning, S., Smethurst, J. A., Rouainia, M., Briggs, K. M., El-Hamalawi, A., & Blake, A. P. (2021). Forecasting the long-term deterioration of a cut slope in high-plasticity clay using a numerical model. Engineering Geology, 280, 105912. https://doi.org/10.1016/j.enggeo.2020.105912

Prabha, R., Ramesh, M. V., Rangan, V. P., Ushakumari, P. V., & Hemalatha, T. (2017). Energy Efficient Data Acquisition Techniques Using Context Aware Sensing for Landslide Monitoring Systems. IEEE Sensors Journal, 17(18), 6006-6018, 7987684. https://doi.org/10.1109/JSEN.2017.2730225

Premchitt, J., Brand, E. W. & Phillipson, H.B. (1986). Landslides caused by rapid groundwater changes. Geological Society, London, Engineering Geology Special Publications, 87-94.

Premchitt, J., Lam, T.S.K., Shen, J.M., & Lam, H.F. (1992). Rainstorm runoff on slopes. Report 12, Geotechnical Engineering Office, Civil Engineering Department of the Hong Kong Special Administrative Region, Hong Kong .

PROJ contributors. (2023). PROJ coordinate transformation software library. Open Source Geospatial Foundation. https://proj.org

Purnama, A. Y., Latif, D. O., Kurniawan, A. W., & Adriyati, M. (2022). Implementation of building information modeling on slope stability and mitigation analysis in Aceh, Indonesia. Journal of Applied Engineering Science, 20(1), 293-299. https://doi.org/10.5937/jaes0-29558

Qi, X., Wei, Y., Mei, X., Chellali, R., & Yang, S. (2024). Comparative analysis of the linear regions in ReLU and LeakyReLU networks. In B. Luo, L. Cheng, Z.-G. Wu, H. Li, & C. Li (Eds.), Neural Information Processing: ICONIP 2023 (Communications in Computer and Information Science, Vol. 1962, pp. 528–539). Springer

Qiao, S., Feng, C., Yu, P., Tan, J., Uchimura, T., Wang, L., Tang, J., Shen, Q., & Xie, J. (2020). Investigation on surface tilting in the failure process of shallow landslides. Sensors (Switzerland), 20(9), 2662. https://doi.org/10.3390/s20092662

Qiu, A., Wang, Q., Chen, Y., Tao, K., Peng, X., He, W., Gao, L., Geli, O. e., & Zhang, F. (2024). Landslide Susceptibility Assessment in Hong Kong with Consideration of Spatio-Temporal Consistency. Applied Sciences, 14(22), 10654. https://www.mdpi.com/2076-3417/14/22/10654

Qiu, D., Wang, L., Luo, D., Huang, H., Ye, Q., & Zhang, Y. (2019). Landslide monitoring analysis of single-frequency BDS/GPS combined positioning with constraints on deformation characteristics. Survey Review, 51(367), 364-372. https://doi.org/10.1080/00396265.2018.1467075

Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106. https://doi.org/10.1007/BF00116251

Quoc-Anh, G., Dinh-Chinh, N., Duc-Nghia, T., Duc-Tan, T., Nguyen, K. T., & Sandrasegaran, K. (2018). Wireless technology for monitoring site-specific landslide in Vietnam. International Journal of Electrical and Computer Engineering, 8(6), 4448-4455. https://doi.org/10.11591/ijece.v8i6.pp4448-4455

Ragnoli, M., Leoni, A., Barile, G., Ferri, G., & Stornelli, V. (2022). LoRa-Based Wireless Sensors Network for Rockfall and Landslide Monitoring: A Case Study in Pantelleria Island with Portable LoRaWAN Access. Journal of Low Power Electronics and Applications, 12(3), 47. https://doi.org/10.3390/jlpea12030047

Rahardjo, H., Lim, T. T., Chang, M. F., & Fredlund, D. G. (1995). Shear-strength characteristics of a residual soil. Canadian Geotechnical Journal, 32(1), 60-77. https://doi.org/10.1139/t95-005

Rahardjo, H., Ong, T. H., Rezaur, R. B., & Leong, E. C. (2007). Factors Controlling Instability of Homogeneous Soil Slopes under Rainfall. Journal of Geotechnical and Geoenvironmental Engineering, 133(12), 1532-1543. https://doi.org/10.1061/(ASCE)1090-0241(2007)133:12(1532)

Rahimi, A., Rahardjo, H., & Leong, E.-C. (2011). Effect of antecedent rainfall patterns on rainfall-induced slope failure. Journal of Geotechnical and Geoenvironmental Engineering, 137(5), 483-491. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000451

Ramesh, M. V., & Rangan, V. P. (2014). Data Reduction and Energy Sustenance in Multisensor Networks for Landslide Monitoring. IEEE Sensors Journal, 14(5), 1555-1563. https://doi.org/10.1109/JSEN.2013.2296611

Ramesh, M. V., & Vasudevan, N. (2012). The deployment of deep-earth sensor probes for landslide detection. Landslides, 9(4), 457-474. https://doi.org/10.1007/s10346-011-0300-x

Reback, J., McKinney, W., jbrockmendel, Van den Bossche, J., Augspurger, T., Cloud, P., Hawkins, S., Gfyoung, Sinhrks, Hoefler, P., Klein, A., Roeschke, M., Tratner, J., She, C., Ayd, W., Naveh, S., Peterson, P., Garcia, M., Schendel, J., Hayden, A., … Shadrach, R. (2020). pandas: Powerful Python data analysis toolkit. Retrieved from https://pandas.pydata.org/

Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., ... & Sabeti, P. C. (2011). Detecting novel associations in large data sets. Science, 334(6062), 1518–1524. https://doi.org/10.1126/science.1205438

Richards, L. A. (1931). Capillary Conduction of Liquids through Porous Mediums. Physics, 1(5), 318-333. https://doi.org/10.1063/1.1745010

Rohatgi, A. (2022). WebPlotDigitizer (Version 4.6) [Computer software]. Retrieved from https://automeris.io/WebPlotDigitizer

Rosi, A., Berti, M., Bicocchi, N., Castelli, G., Corsini, A., Mamei, M., & Zambonelli, F. (2011). Landslide monitoring with sensor networks: Experiences and lessons learnt from a real-world deployment. International Journal of Sensor Networks, 10(3), 111-122. https://doi.org/10.1504/IJSNET.2011.042195

Rossi, G., Catani, F., Leoni, L., Segoni, S., & Tofani, V. (2013). HIRESSS: a physically based slope stability simulator for HPC applications. Natural Hazards and Earth System Sciences, 13(1), 151-166. https://doi.org/10.5194/nhess-13-151-2013

Rumelhart, D., Hinton, G. & Williams, R (1986). Learning representations by back-propagating errors. Nature 323, 533–536. https://doi.org/10.1038/323533a0

Ruzza, G., Guerriero, L., Revellino, P., & Guadagno, F. M. (2020). A multi-module fixed inclinometer for continuous monitoring of landslides: Design, development, and laboratory testing. Sensors (Switzerland), 20(11), 1-19, 3318. https://doi.org/10.3390/s20113318

Saadatkhah, N., Kassim, A., & Lee, L. M. (2015). Hulu Kelang, Malaysia regional mapping of rainfall-induced landslides using TRIGRS model. Arabian Journal of Geosciences, 8(5), 3183-3194. https://doi.org/10.1007/s12517-014-1410-2

Saha, R., Baranval, N. K., Das, I. C., Kumaranchat, V. K., & Reddy, K. S. (2022). Application of Machine Learning and Geospatial Techniques for Groundwater Potential Mapping. Journal of the Indian Society of Remote Sensing, 50(10), 1995-2010. https://doi.org/10.1007/s12524-022-01582-z

Salvatici, T., Tofani, V., Rossi, G., D'Ambrosio, M., Tacconi Stefanelli, C., Benedetta Masi, E., Rosi, A., Pazzi, V., Vannocci, P., Petrolo, M., Catani, F., & Ratto, S. (2018). Application of a physically based model to forecast shallow landslides at a regional scale. Natural Hazards and Earth System Sciences, 18(7), 1919-1935. https://doi.org/10.5194/nhess-18-1919-2018.

Sarma, S. K. (1987). A note on the stability analysis of slopes. Geotechnique, 37(1), 107-111. https://doi.org/10.1680/geot.1987.37.1.107

Sasahara, K. (2017). Prediction of the shear deformation of a sandy model slope generated by rainfall based on the monitoring of the shear strain and the pore pressure in the slope. Engineering Geology, 224, 75-86. https://doi.org/10.1016/j.enggeo.2017.05.003

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. Proceedings of the 9th Python in Science Conference, 57–61 .

Segoni, S., Piciullo, L., & Gariano, S. L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides, 15(8), 1483-1501. https://doi.org/10.1007/s10346-018-0966-4

Segui, C., & Veveakis, M. (2021). Continuous assessment of landslides by measuring their basal temperature. Landslides, 18(12), 3953-3961. https://doi.org/10.1007/s10346-021-01762-x

Seguí, C., & Veveakis, M. (2022). Forecasting and mitigating landslide collapse by fusing physics-based and data-driven approaches. Geomechanics for Energy and the Environment, 32, 100412. https://doi.org/10.1016/j.gete.2022.100412

Setiono, A., Qomaruddin Afandi, M.I., Adinanta, H., Rofianingrum, M.Y., Suryadi Mulyanto, I., Bayuwati, D., Anwar, A., & Widiyatmoko, B. (2023). Wire Extensometer Based on Optical Encoder for Translational Landslide Measurement. International Journal on Advanced Science, Engineering and Information Technology, 13, 17–23.https://doi.org/10.18517/ijaseit.13.1.15636

Shamshi, M. A. (2004). Technologies Convergence in Recent Instrumentation for Natural Disaster Monitoring and Mitigation. IETE Technical Review, 21(4), 277-290. https://doi.org/10.1080/02564602.2004.11417155

Shano, L., Raghuvanshi, T. K., & Meten, M. (2020). Landslide susceptibility evaluation and hazard zonation techniques – a review. Geoenvironmental Disasters, 7(1), 18. https://doi.org/10.1186/s40677-020-00152-0

Shapely. (2023). Shapely Documentation. Retrieved from https://shapely.readthedocs.io

Sheikh, M. R., Nakata, Y., Shitano, M., & Kaneko, M. (2021). Rainfall-induced unstable slope monitoring and early warning through tilt sensors . Soils and Foundations, 61(4), 1033-1053. https://doi.org/10.1016/j.sandf.2021.05.010

Shentu, N., Yang, J., Li, Q., Qiu, G., & Wang, F. (2023). Research on the Landslide Prediction Based on the Dual Mutual-Inductance Deep Displacement 3D Measuring Sensor. Applied Sciences, 13(1), 213. https://www.mdpi.com/2076-3417/13/1/213

Shi, C., & Wang, Y. (2022). Machine learning of three-dimensional subsurface geological model for a reclamation site in Hong Kong. Bulletin of Engineering Geology and the Environment, 81(12), 504. https://doi.org/10.1007/s10064-022-03009-y

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. Water Science and Technology, 82(12), 2635-2670. https://doi.org/10.2166/wst.2020.369

Sitarenios, P., Casini, F., Askarinejad, A., & Springman, S. (2021). Hydro-mechanical analysis of a surficial landslide triggered by artificial rainfall: the Ruedlingen field experiment. Geotechnique, 71(2), 96-109. https://doi.org/10.1680/jgeot.18.P.188

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, Lake Tahoe, Nevada .

Soga, K., Alonso, E., Yerro, A., Kumar, K., & Bandara, S. (2016). Trends in large-deformation analysis of landslide mass movements with particular emphasis on the material point method. Geotechnique, 66(3), 248-273. https://doi.org/10.1680/jgeot.15.LM.005

Song, D., Choi, C., Ng, C. W. W., & Zhou, G. (2017). Geophysical flows impacting a flexible barrier: effects of solid-fluid interaction. Landslides, 15, 99-110. https://doi.org/10.1007/s10346-017-0856-1

Song, Q., Wu, Y., Xin, X., Yang, L., Yang, M., Chen, H., ... & Li, J. (2019). Real-time tunnel crack analysis system via deep learning. Ieee Access, 7, 64186-64197. https://doi.org/10.1109/ACCESS.2019.2916330

Song, Z., Li, X., Lizárraga, J. J., Zhao, L., & Buscarnera, G. (2020). Spatially distributed landslide triggering analyses accounting for coupled infiltration and volume change. Landslides, 17(12), 2811-2824. https://doi.org/10.1007/s10346-020-01451-1

Spencer, E. (1967). A Method of analysis of the Stability of Embankments Assuming Parallel Inter-Slice Forces. Geotechnique, 17(1), 11-26. https://doi.org/10.1680/geot.1967.17.1.11

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958. https://doi.org/10.5555/2627435.2670313

Srivastava, P., Shukla, A., & Bansal, A. (2021). A comprehensive review on soil classification using deep learning and computer vision techniques. Multimedia Tools and Applications, 80(10), 14887-14914. https://doi.org/10.1007/s11042-021-10544-5

Suk, J. W., Jeong, H. S., Jung, M. S., Kang, H. S., Kim, H. J., & Choi, S. G. (2022). Prediction of Shallow Failure on a Slope Using Volumetric Water Content Gradient Characteristics. Applied Sciences, 12(11), 5308. https://doi.org/10.3390/app12115308

Sulsky, D., Zhou, S.-J., Schreyer, H. L. (1995). Application of a particle-in-cell method to solid mechanics. Computer Physics Communications, 87(1-2), 236-252. https://doi.org/10.1016/0010-4655(94)00170-7

Sun, J., Yu, T., & Dong, P. (2022). Evaluation of 3D slope stability based on the minimum potential energy principle. Computers and Geotechnics, 146, 104717. https://doi.org/10.1016/j.compgeo.2022.104717

Tang, Y., Wu, W., Yin, K., Wang, S., & Lei, G. (2019). A hydro-mechanical coupled analysis of rainfall induced landslide using a hypoplastic constitutive model. Computers and Geotechnics, 112, 284-292. https://doi.org/10.1016/j.compgeo.2019.04.024

Tao, Z., Wang, Y., Zhu, C., Xu, H., Li, G., & He, M. (2019). Mechanical evolution of constant resistance and large deformation anchor cables and their application in landslide monitoring. Bulletin of Engineering Geology and the Environment, 78(7), 4787-4803. https://doi.org/10.1007/s10064-018-01446-2

Tengtrairat, N., Woo, W. L., Parathai, P., Aryupong, C., Jitsangiam, P., & Rinchumphu, D. (2021). Automated Landslide-Risk Prediction Using Web GIS and Machine Learning Models. Sensors, 21(13), 4620. https://www.mdpi.com/1424-8220/21/13/4620

TensorFlow. TensorFlow Core v2.7.0: tf.Module. Retrieved [6 22, 2024], from https://www.tensorflow.org/api_docs/python/tf/all_symbols

Thang, N. V., Wakai, A., Sato, G., Viet, T. T., & Kitamura, N. (2022). Simple Method for Shallow Landslide Prediction Based on Wide-Area Terrain Analysis Incorporated with Surface and Subsurface Flows. Natural Hazards Review, 23(4), 04022028. https://doi.org/doi:10.1061/(ASCE)NH.1527-6996.0000578

The Hong Kong Polytechnic University HPC, (2025): https://studenthpc.polyu.edu.hk.
Togneri, R., dos Santos, D. F., Camponogara, G., Nagano, H., Custodio, G., Prati, R., ... & Kamienski, C. (2022). Soil moisture forecast for smart irrigation: The primetime for machine learning. Expert Systems with Applications, 207, 117653. https://doi.org/10.1016/j.eswa.2022.117653
Tohari, A., Nishigaki, M., & Komatsu, M. (2007). Laboratory rainfall-induced slope failure with moisture content measurement. Journal of Geotechnical and Geoenvironmental Engineering, 133(5), 575-587. https://doi.org/10.1061/(ASCE)1090-0241(2007)133:5(575)
Tsuchida, T., Athapaththu, A. M. R. G., Kawabata, S., Kano, S., Hanaoka, T., & Yuri, A. (2014). Individual landslide hazard assessment of natural valleys and slopes based on geotechnical investigation and analysis. Soils and Foundations, 54(4), 806-819. https://doi.org/10.1016/j.sandf.2014.06.012
Uchida, T., Mori, N., Tamura, K., Terada, H., Takiguchi, S., & Kamee, K. (2009). The role of data preparation on shallow landslide prediction. Journal of the Japan Society of Erosion Control Engineering, 62(1), 23-31. http://www.jsece.or.jp/
Uwihirwe, J., Hrachowitz, M., & Bogaard, T. A. (2020). Landslide precipitation thresholds in Rwanda. Landslides, 17(10), 2469-2481. https://doi.org/10.1007/s10346-020-01457-9
Valentino, R., Meisina, C., Montrasio, L., Losi, G. L., & Zizioli, D. (2014). Predictive Power Evaluation of a Physically Based Model for Shallow Landslides in the Area of Oltrepò Pavese, Northern Italy. Geotechnical and Geological Engineering, 32(4), 783-805. https://doi.org/10.1007/s10706-014-9758-3
van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics, 84(2), 523-538. https://doi.org/10.1007/s11192-009-0146-3
Van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal, 44(5), 892-898. https://doi.org/10.2136/sssaj1980.03615995004400050002x
Vanapalli, S. K., Fredlund, D. G., Pufahl, D. E., & Clifton, A. W. (1996). Model for the prediction of shear strength with respect to soil suction. Canadian Geotechnical Journal, 33(3), 379-392. https://doi.org/10.1139/t96-060
Varangaonkar, P., & Rode, S. V. (2023). Lightweight deep learning model for automatic landslide prediction and localization. Multimedia Tools and Applications, 82(21), 33245-33266. https://doi.org/10.1007/s11042-023-15049-x
Vargas, M. (1971). Effect of rainfall and ground water levels. Proceedings of the 4th Pan American Conference on Soil Mechanics and Foundation Engineering, San Juan, Puerto Rico, 138-141.
Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261-272.
Wang, C., & Zhao, Y. (2023). Time Series Prediction Model of Landslide Displacement Using Mean-Based Low-Rank Autoregressive Tensor Completion. Applied Sciences, 13(8), 5214. https://www.mdpi.com/2076-3417/13/8/5214
Wang, J., Hong, Y., Li, L., Gourley, J. J., Khan, S. I., Yilmaz, K. K., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., Limaye, A. S., Korme, T., & Okello, L. (2011). The coupled routing and excess storage (CREST) distributed hydrological model. Hydrological Sciences Journal, 56(1), 84-98. https://doi.org/10.1080/02626667.2010.543087
Wang, S., Zhang, K., van Beek, L. P. H., Tian, X., & Bogaard, T. A. (2020a). Physically-based landslide prediction over a large region: Scaling low-resolution hydrological model results for high-resolution slope stability assessment. Environmental Modelling & Software, 124, 104607. https://doi.org/10.1016/j.envsoft.2019.104607
Wang, Y., Feng, L., Li, S., Ren, F., & Du, Q. (2020). A hybrid model considering spatial heterogeneity for landslide susceptibility mapping in Zhejiang Province, China. Catena, 188, 104425.
Wang, Y., Liu, Z., Wang, D., Li, Y., & Yan, J. (2017). Anomaly Detection and Visual Perception for Landslide Monitoring Based on a Heterogeneous Sensor Network. IEEE Sensors Journal, 17(13), 4248-4257, 7929300. https://doi.org/10.1109/JSEN.2017.2704584
Wang, Z. (2022). Comparative study of Latin hypercube sampling and Monte Carlo method in structural reliability analysis. Highl. Sci. Eng. Technol., 28, 61–69
Waskom, M. L. (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021, https://doi.org/10.21105/joss.03021
Wasowski, J., & Bovenga, F. (2014). Investigating landslides and unstable slopes with satellite Multi Temporal Interferometry: Current issues and future perspectives. Engineering Geology, 174, 103-138. https://doi.org/10.1016/j.enggeo.2014.03.003

Wicki, A., & Hauck, C. (2022). Monitoring critically saturated conditions for shallow landslide occurrence using electrical resistivity tomography. Vadose Zone Journal, 21(4), e20204. https://doi.org/10.1002/vzj2.20204
Wicki, A., Jansson, P. E., Lehmann, P., Hauck, C., & Stähli, M. (2021). Simulated or measured soil moisture: which one is adding more value to regional landslide early warning? Hydrol. Earth Syst. Sci., 25(8), 4585-4610. https://doi.org/10.5194/hess-25-4585-2021
Willmott, C., & Matsuura, K. (2005). Advantages of the Mean Absolute Error (MAE) over the Root Mean Square Error (RMSE) in Assessing Average Model Performance. Climate Research, 30, 79. https://www.jstor.org/stable/24869236
Winters, P. R. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. Management Science, 6(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324
Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, England, United Kingdom. https://doi.org/10.1145/2601248.2601268
Wolpert, D. H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. Neural Computation, 8(7), 1341-1390. https://doi.org/10.1162/neco.1996.8.7.1341
Wolpert, D. H. (1996). The lack of a priori distinctions between learning algorithms. Neural computation, 8(7), 1341-1390. https://doi.org/10.1162/neco.1996.8.7.1341
Wong, H. N. & Ho, K. K. S. (1996). Systematic Investigation of Landslides Caused By a Severe Rainstorm in Hong Kong. HKIE Trans., 3(3), 17-27 .
Wong, H. N., & Ho, K. K. S. (1996). Systematic Investigation of Landslides Caused By a Severe Rainstorm in Hong Kong. HKIE Transactions, 3(3), 17-27. https://doi.org/10.1080/1023697X.1996.10667704
Wong, H. Y. (2020). Shear strength properties of Hong Kong soils for slope stability. HKIE Transactions Hong Kong Institution of Engineers, 27(1), 48-54 .
World Health Organization (WHO) (2023). Landslides. https://www.who.int/health-topics/landslides#tab=tab_1
Wu, L., Huang, R., & Li, X. (2020). Hydro-mechanical analysis of rainfall-induced landslides. Springer Singapore, 1-235, DOI: https://doi.org/10.1007/978-981-15-0761-8 .
Wu, S., Hu, X., Zheng, W., He, C., Zhang, G., Zhang, H., & Wang, X. (2021). Effects of reservoir water level fluctuations and rainfall on a landslide by two-way ANOVA and K-means clustering. Bulletin of Engineering Geology and the Environment, 80(7), 5405-5421. https://doi.org/10.1007/s10064-021-02273-8
Wuni, I. Y., & Shen, G. Q. (2020). Critical success factors for modular integrated construction projects: a review. Building Research & Information, 48(7), 763-784. https://doi.org/10.1080/09613218.2019.1669009
Xi, N., Zang, M., Lin, R., Sun, Y., & Mei, G. (2023). Spatiotemporal prediction of landslide displacement using deep learning approaches based on monitored time-series displacement data: a case in the Huangliangshu landslide. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, 17, 98 - 113. https://doi.org/10.1080/17499518.2023.2172186
Xiao, T., Zhang, L. M., Cheung, R. W. M., & Lacasse, S. (2023). Predicting spatio-temporal man-made slope failures induced by rainfall in Hong Kong using machine learning techniques. Géotechnique, 73(9), 749-765. https://doi.org/10.1680/jgeot.21.00160
Xiaochun, L., Xue, C., Bobo, X., Bin, T., Xiaolong, T., & Zhigang, T. (2022). Bi-LSTM-GPR algorithms based on a high-density electrical method for inversing the moisture content of landslide. Bulletin of Engineering Geology and the Environment, 81(11), 491. https://doi.org/10.1007/s10064-022-02982-8
Xie, J., Huang, J., Zeng, C., Huang, S., & Burton, G. J. (2022). A generic framework for geotechnical subsurface modeling with machine learning. Journal of Rock Mechanics and Geotechnical Engineering, 14(5), 1366-1379. https://doi.org/10.1016/j.jrmge.2022.08.001
Xing, Y., Yue, J., & Chen, C. (2020). Interval Estimation of Landslide Displacement Prediction Based on Time Series Decomposition and Long Short-Term Memory Network. IEEE Access, 8, 3187-3196. https://doi.org/10.1109/ACCESS.2019.2961295
Xu, W., Xu, H., Chen, J., Kang, Y., Pu, Y., Ye, Y., & Tong, J. (2022). Combining Numerical Simulation and Deep Learning for Landslide Displacement Prediction: An Attempt to Expand the Deep Learning Dataset. Sustainability, 14(11). https://doi.org/10.3390/su14116908
Xu, Y., Chen, Q., Tian, D., Zhang, Y., Li, B., & Tang, H. (2020). Selection of high transfer stability and optimal power-efficiency tradeoff with respect to distance region for underground wireless power transfer systems. Journal of Power Electronics, 20(6), 1662-1671. https://doi.org/10.1007/s43236-020-00156-x

Yanbin, M., Hongrui, L., Lin, W., Wengang, Z., Zhengwei, Z., Haiqing, Y., Luqi, W., & Xingzhong, Y. (2022). Machine learning algorithms and techniques for landslide susceptibility investigation: A literature review. Journal of Civil and Environmental Engineering., 44, 53-67. https://doi.org/10.11835/j.issn.2096-6717.2021.102.

Yang, C.-M., Yu, W.-L., Dong, J.-J., Kuo, C.-Y., Shimamoto, T., Lee, C.-T., & Togo, T., Miyamoto, Y. (2014). Initiation, movement, and run-out of the giant Tsaoling landslide — What can we learn from a simple rigid block model and a velocity–displacement dependent friction law? Engineering Geology, 182, 158-181. https://doi.org/10.1016/j.enggeo.2014.08.008

Yang, X., Diao, X., Zhou, T., Hu, F., & Wang, S. (2023). Study on the Stability of Accumulated Layer Landslide under the Coupling Action of Earthquake and Rainfall. KSCE Journal of Civil Engineering, 27(1), 98-108. https://doi.org/10.1007/s12205-022-0110-9

Yang, Z., Shao, W., Qiao, J., Huang, D., Tian, H., Lei, X., & Uchimura, T. (2017). A Multi-Source Early Warning System of MEMS Based Wireless Monitoring for Rainfall-Induced Landslides. Applied Sciences, 7(12), 1234. https://www.mdpi.com/2076-3417/7/12/1234

Yerro, A., Girardi, V., Martinelli, M., & Ceccato, F. (2022). Modelling unsaturated soils with the Material Point Method. A discussion of the state-of-the-art. Geomechanics for Energy and the Environment, 32, 100343. https://doi.org/10.1016/j.gete.2022.100343

Yong, H., Jianqiang Huang, Xiansheng Hua, & Zhang, L. (2020). Gradient Centralization: A New Optimization Technique for Deep Neural Networks. In The 16th European Conference on Computer Vision (ECCV 2020) (pp. 1-20). (The 16th European Conference on Computer Vision (ECCV 2020)).

You, Y., Li, J., Reddi, S., Hseu, J., Kumar, S., Bhojanapalli, S., ... & Hsieh, C. J. (2019). Large batch optimization for deep learning: Training bert in 76 minutes. arXiv preprint arXiv:1904.00962.

Yueshun, H., & Wei, Z. (2013). The reseach on wireless sensor network for landslide monitoring. International Journal on Smart Sensing and Intelligent Systems, 6(3), 867-887. https://doi.org/10.21307/ijssis-2017-570

Yunus, M. A. M., Ibrahim, S., Khairi, M. T. M., & Faramarzi, M. (2015). The application of WiFi-based wireless sensor network (WSN) in hill slope condition monitoring. Jurnal Teknologi, 73(3), 75-84. https://doi.org/10.11113/jt.v73.4250

Zaheer, M., Reddi, S., Sachan, D., Kale, S., & Kumar, S. (2018). Adaptive methods for nonconvex optimization. Advances in neural information processing systems, 31. Retrieved from https://papers.nips.cc/paper_files/paper/2018/hash/90365351ccc7437a1309dc64e4db32a3-Abstract.html

Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701.

Zhang, L. L., Zhang, J., Zhang, L. M., & Tang, W. H. (2011). Stability analysis of rainfall-induced slope failure: a review. Proceedings of the Institution of Civil Engineers - Geotechnical Engineering, 164(5), 299-316. https://doi.org/10.1680/geng.2011.164.5.299

Zhang, L., Cui, Y., Zhu, H., Wu, H., Han, H., Yan, Y., Shi, B. (2023). Shear deformation calculation of landslide using distributed strain sensing technology considering the coupling effect. Landslides, 20, 1583–1597.

Zhang, S., Wang, K., Hu, K., Xu, H., Xu, C., Liu, D., Lv, J., & Wei, F. (2020). Model test: Infrasonic features of porous soil masses as applied to landslide monitoring. Engineering Geology, 265, 105454. https://doi.org/10.1016/j.enggeo.2019.105454

Zhang, T., Li, Y., Wang, T., Wang, H., Chen, T., Sun, Z., ... & Han, L. (2022). Evaluation of different machine learning models and novel deep learning-based algorithm for landslide susceptibility mapping. Geoscience Letters, 9(1), 26. https://doi.org/10.1186/s40562-022-00236-9

Zhang, W., Goh, A. T. C., & Zhang, Y. (2016). Multivariate Adaptive Regression Splines Application for Multivariate Geotechnical Problems with Big Data. Geotechnical and Geological Engineering, 34(1), 193-204. https://doi.org/10.1007/s10706-015-9938-9

Zhao, B., Dai, Q., Han, D., Dai, H., Mao, J., & Zhuo, L. (2019a). Probabilistic thresholds for landslides warning by integrating soil moisture conditions with rainfall thresholds. Journal of Hydrology, 574, 276-287. https://doi.org/10.1016/j.jhydrol.2019.04.062

Zhao, B., Dai, Q., Han, D., Dai, H., Mao, J., Zhuo, L., & Rong, G. (2019b). Estimation of soil moisture using modified antecedent precipitation index with application in landslide predictions. Landslides, 16(12), 2381-2393. https://doi.org/10.1007/s10346-019-01255-y

Zhao, B., Marin, R. J., Luo, W., Yu, Z., & Yuan, L. (2025). Rainfall thresholds for shallow landslides considering rainfall temporal patterns. Bulletin of Engineering Geology and the Environment, 84(3), 132. https://doi.org/10.1007/s10064-025-04144-y

Zhao, L., Liu, X., Mao, J., Shao, L., & Li, T. (2020). Three-dimensional distance potential discrete element method for the numerical simulation of landslides. Landslides, 17(2), 361-377. https://doi.org/10.1007/s10346-019-01282-9

Zhou, W.G., (2008). Experimental and theoretical study on pullout resistance of grouted soil nails. Ph.D. Thesis, The Hong Kong Polytechnic University.

Zhu, Y., Ishikawa, T., Zhang, Y., Nguyen, B. T., & Subramanian, S. S. (2022). A FEM-MPM hybrid coupled framework based on local shear strength method for simulating rainfall/runoff-induced landslide runout. Landslides, 19(8), 2021-2032. https://doi.org/10.1007/s10346-022-01849-z

Zivot, E., & Wang, J. (2006). Rolling Analysis of Time Series. In: Modeling Financial Time Series with S-PLUS®. Springer, New York, NY. https://doi.org/10.1007/978-0-387-32348-0_9