AUTOMATED AND NON-INVASIVE MENTAL FATIGUE ASSESSMENT OF CONSTRUCTION EQUIPMENT OPERATORS

MEHMOOD IMRAN

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Building and Real Estate

AUTOMATED AND NON-INVASIVE MENTAL FATIGUE ASSESSMENT OF CONSTRUCTION EQUIPMENT OPERATORS

MEHMOOD Imran

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

July 2023

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 acknowledgement has been made in the text.

___________________________________ (Signed)

MEHMOOD Imran (Name of student)

ABSTRACT

The construction industry plays a vital role in enhancing and advancing the infrastructure, contributing significantly to its development. However, at the same time it is labour-intensive and famous for poor safety records around the globe. Despite the positive effects it brings, ensuring the safety of construction site workers remains an unresolved issue and a top priority. Construction workers have to perform repetitive and mentally demanding tasks. In construction industry, various heavy equipment, such as excavators, tower cranes, trucks, loaders, is used for construction tasks. The operations of these construction equipment are repetitive and mentally demanding. The equipment operators have to work for prolonged hours to complete ongoing equipment operations that require constant attention from the operator and can be mentally challenging. Extended periods of operating construction equipment can induce mental fatigue as it demands continuous attention from the operator. This can lead to an elevated risk of accidents caused by human errors and compromised health for operators due to lapses in attention. To mitigate the risk of accidents and safeguard the well-being of operators, it is crucial to consistently and promptly monitor their mental fatigue in real time.

Mental fatigue poses a notable risk factor for on-site incidents and accidents, as it impairs operators' ability to sustain their focus during construction equipment operations. The published literature states that the mental fatigue can result in poor decision-making, human errors, or underperformance, potentially creating hazardous situations for the operators. Recognizing its high prevalence and profound effect on construction workers and machinery operators, extensive research has been conducted to detect and identify mental fatigue early. Recently, invasive technologies have been used to solve this problem. However, these methods require the use of physical sensors, which can irritate and annoy operators, thereby impeding normal construction site work. This study addresses these issues by introducing a non-invasive method for assessing mental fatigue using geometric measures of facial features, rather than having operators wear sensors on their bodies. The proposed method was further validated by comparing it with wearable electroencephalography (EEG) technology, establishing its ecological validity for construction equipment operators. The following are the primary objectives of this research study: (1) to study the non-invasive detection of mental fatigue in construction equipment

operators through geometric measurements of facial features; (2) to examine the validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators; (3) to develop deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data; (4) to examine multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals.

Although the application of facial features has been widespread in other domains, such as drivers and other occupation scenarios, its ecological validity for construction excavator operators remains a knowledge gap. Consequently, there is a dearth of knowledge about creating a contactless and non-intrusive system for detecting mental fatigue in construction equipment operators. To start with, a study was conducted to investigate whether there are variations in the geometric measurements of facial features owing to mental fatigue. An excavation experiment was conducted and simultaneously with the task, the operators were video recorded to collect the data on their facial features via mobile camera. Based on geometric measurements, facial features (eyebrow, mouth outer, mouth corners, head motion, eye area, and face area) were extracted. The results found that there was a significant difference in the measured metrics for high fatigue as compared to low fatigue. Specifically, the most noteworthy variation was for the eye and face area metrics, with respective mean differences of 45.88% and 26.9%. The results indicate that the use of geometric measurements of facial features is an effective and non-intrusive method for detecting mental fatigue in construction equipment operators. Secondly, the proposed method was further validated through investigations that involved a comparison with flexible headband-based wearable electroencephalography (EEG) technology. The aim was to establish the ecological validity of the proposed method for construction equipment operators. Ground truth data, including brain activity captured by wearable EEG, along with geometric measurements of facial features, were extracted and analysed at baseline and at 20-minute intervals over the course of one hour. The results revealed significant temporal variation in the measured metrics such as eye aspect ratio, eye distance, mouth aspect ratio, face area, and head motion. These metrics were also found to have a significant correlation with both the ground truth data and the EEG metrics. Additionally, the patterns observed in the brain visualizations obtained from EEG were associated with variations in the facial features. Overall, the findings of this study demonstrate that mental fatigue among construction equipment operators can be effectively monitored in a non-invasive manner using geometric measurements of facial features.

Thirdly, a study was conducted to investigate mental fatigue in construction equipment operators as a multimodal problem. Previous studies classified mental fatigue using single modal data with acceptable accuracy. However, mental fatigue is a multimodal problem, and no single modality is superior. Moreover, none of the previous studies in construction industry have investigated the multimodal data fusion for classifying mental fatigue, and whether such an approach would improve mental fatigue detection. This study proposes a novel approach using three machine learning models and multimodal data fusion to classify mental fatigue states. Electroencephalography, electrodermal activity, and video signals were acquired during an excavation operation, and the decision tree model using multimodal sensor data fusion outperformed other models with 96.2% accuracy and 96.175% to 98.231% F1 score. Multimodal sensor data fusion can aid in developing a real-time system to classify mental fatigue, improving safety and health management on construction sites. Finally, a study was conducted to propose the feasibility of a construction site strategy that utilizes flexible headband-based sensors to capture raw EEG data, and deep learning networks to recognize operators' mental fatigue. Previous approaches, such as machine learning using EEG-based wearable sensing systems, have been proposed to detect mental fatigue accurately. However, implementing these strategies on actual construction sites remains a challenge. The limited mobility and systemic instability of EEG sensors restrict their application to laboratory settings rather than to real construction environments. In addition, machine learning classifiers relying solely on manually engineered EEG features may compromise their performance in practical construction scenarios. To address these issues, this study employed the NASA-TLX score as the ground truth for measuring mental fatigue. Brain activity patterns were recorded using a wearable EEG sensor, and raw EEG data were used to develop the deep learning-based classification models. The performances of different deep learning models, including long short-term memory (LSTM), bidirectional LSTM, and one-dimensional convolutional networks, were assessed using metrics such as accuracy, precision, recall, specificity, and F1-score. The findings revealed that the bidirectional LSTM (Bi-LSTM) model outperformed other deep learning models, achieving a remarkable accuracy of 99.941% and an F1-score ranging from 99.917% to 99.993%. These results demonstrate the feasibility of implementing the Bi-LSTM model and contribute to the recognition and classification of mental fatigue by using wearable sensors. Ultimately, this advancement enhances the health and safety of operations at construction sites.

LIST OF PUBLICATIONS

*An asterisk * indicates corresponding author.*

List of Relevant Publications during PhD (2020-2024)

(a) Refereed Journal Papers (Published)

1. **Imran Mehmood**, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". Journal of Safety Research, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291
2. **Imran Mehmood***, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". Developments in the Built Environment, Volume 15, 100198
3. Shahnawaz Answer, Heng Li, Maxwell Fordjour Antwi-Afari, Aquil Maud Mirza, Mohammed Abdul Rahman, **Imran Mehmood**, Runhao Guo, Arnold YL Wong (2023), "Evaluation of data processing and artifacts removal approaches used for physiological signals captured using wearable sensing devices during construction tasks: A State-of-the-Art Review". Journal of Construction Engineering and Management Volume 150, Issue 1
4. **Imran Mehmood***, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". Advanced Engineering Informatics, Volume 56, 101978
5. Shahnawaz Anwer, Heng Li, Waleed Umer, Maxwell Fordjour Antwi-Afari, **Imran Mehmood***, Yantao Yu, Carl Haas, Arnold Yu Lok Wong (2023) "Identification and classification of physical fatigue in construction workers using linear and nonlinear heart rate variability measurements". Journal of Construction Engineering and Management, Volume 147, Issue 7, 04023057

6. Jie Ma, Heng Li, Xingcan Huang, Bo Fang, Zeyu Zhao, **Imran Mehmood**, Yiming Liu, Guo Zhang, Xin Fang, Mehrdad Arashpour, Shahnawaz Answer (2023) "A Sweat-based Lactate Biosensor for assessing fatigue of construction equipment operators". International Journal of Industrial Ergonomics, Volume 96, 103472
7. **Imran Mehmood***, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" Advanced Engineering Informatics, Volume 54, 101777
8. Shahnawaz Anwer, Heng Li, Maxwell Fordjour Antwi-Afari, Waleed Umer, **Imran Mehmood**, Mohamed Al-Hussein, Arnold Yu Lok Wong (2021) "Test-retest reliability, validity, and responsiveness of a textile-based wearable sensor for real-time assessment of physical fatigue in construction bar-benders". Journal of Building Engineering, 44, 103348. https://doi.org/10.1016/j.jobe.2021.103348

(b) Refereed Journal Papers (Under Review)

1. Jie Ma, Heng Li, Bo Fang, Zeyu Zhao, **Imran Mehmood***, Aamir Arsalan, Lei Wang, (2024) "Assessing physical and mental fatigue in construction workers: validating sweat lactate as a biomarker". International Journal of Occupational Safety and Ergonomics, JOSE-2024-0045.

(c) Refereed Conference Papers

1. **Imran Mehmood***, Heng Li, Shahnawaz Anwer, Waleed Umer (2023) "Machine learning-based recognition of mental fatigue in construction equipment operators using facial features". 13th International Conference on Construction in 21st Century (CITC-13), May 8-11, 2023, Netherlands

List of Other Publications during PhD (2020-2024)

(a) Refereed Journal Papers

1. Weihao Sun, Maxwell Fordjour Antwi-Afari, **Imran Mehmood**, Shahnawaz Anwer, Waleed Umer (2023) "Critical Success Factors for Implementing Blockchain Technology in the Construction Industry: Review, Stage Framework, and Future Research Directions". Automation in Construction, Volume 156, December 2023, 105135

2. Haitao Wu, Wenyan Zhong, Botao Zhong, Heng Li, Jiadong Guo, **Imran Mehmood** (2023) “Barrier Identification, Analysis, and Solutions of Blockchain Adoption in Construction: A Fuzzy DEMATEL and TOE integrated Method”. Engineering Construction and Architectural Management (Published on 20 July 2023)
3. Saimin Huang, Hongchang Wang, Waqas Ahmad, Ayaz Ahmad, Nikolai Ivanovich Vatin, Abdeliazim Mustafa Mohamed, Ahmed Farouk Deifalla, **Imran Mehmood** (2022) “Plastic Waste Management Strategies and Their Environmental Aspects: A Scientometric Analysis and Comprehensive Review”. International Journal of Environmental Research and Public Health, Vol. 19, Issue 8, https://doi.org/10.3390/ijerph19084556
4. Shahnawaz Anwer, Heng Li, Maxwell Fordjour Antwi-Afari, Waleed Umer, **Imran Mehmood**, Arnold Yu Lok Wong (2021) “Effects of load carrying techniques on gait parameters, dynamic balance, and physiological parameters during a manual material handling task”. Engineering, Construction and Architectural Management, https://doi.org/10.1108/ecam-03-2021-0245
5. Ayaz Ahmad, Krzysztof Ostrowski, Mariusz Maslak, Furqan Farooq, **Imran Mehmood**, Afnan Nafees (2021) “Comparative Study of Supervised Machine Learning Algorithms for Predicting the Compressive Strength of Concrete at High Temperature”. Materials, Vol. 14, Issue 15, https://doi.org/10.3390/ma14154222

(b) Refereed Journal Papers (Under Review)

1. Yanxue Li, Maxwell Fordjour Antwi-Afari, Shahnawaz Answer, **Imran Mehmood**, Waleed Umer, Saeed Reza Mohandes, Ibrahim Yahaya Wuni, Heng Li (2024) “Artificial intelligence in net-zero carbon emissions for sustainable building projects: a systematic literature and science mapping review”. Energy & Buildings, ENB-D-24-00348.

(c) Refereed Conference Papers

1. Ridwan Taiwo, Kwok Chin Wang, Oludolapo Ibrahim Olanrewaju, Salman Tariq, Owolabi Titilayo Abimbola, **Imran Mehmood**, Tarek Zayed (2022) “An Analysis of Employee Motivation in the Construction Industry: The Case of Hong Kong”. Engineering Proceedings, Vol. 22, Issue 1, https://doi.org/10.3390/engproc2022022011

ACKNOWLEDGEMENTS

In the name of Allah, the Most Beneficent, the Most Merciful, may peace and blessings be upon the final Prophet Muhammad (Peace be upon him), his family, and his companions. First of all, I offer my profound gratitude to Allah, the Most Merciful and the Most Compassionate, for bestowing upon me the strength, patience, and perseverance to embark on this academic journey. I am deeply grateful for the blessings and opportunities provided throughout the process. Ultimately, a protracted yet exciting journey of research studies reached its conclusion. The attainment of success in this endeavour would have been unfeasible without the vital assistance of the individuals I am honoured to acknowledge and express my gratitude.

I would like to begin by extending my utmost gratitude to Chair Professor Heng Li, my chief supervisor, for his exceptional guidance and unwavering support throughout my PhD journey. His passion for research has been an immense source of motivation, inspiring me to strive for excellence in my scholarly pursuits. Professor Li's constant encouragement, both in and outside the realm of academia, has played a crucial role in making this challenging journey more manageable. His insightful advice, and willingness to provide guidance have been invaluable in shaping my academic growth. I am deeply appreciative of his commitment to not only my research progress but also my long-term professional development.

I am immensely grateful for the invaluable support and guidance provided by Dr Waleed Umer, throughout my research. Dr Waleed's expertise in experimental methods has been instrumental in shaping my understanding and application of scientific techniques. His meticulous review and editing of my research papers, word by word, have significantly improved the clarity and precision of my work. Furthermore, I would like to acknowledge the support from Dr Shahnawaz Anwer. His unwavering support during challenging times when my research encountered obstacles. His resourcefulness and willingness to assist in finding solutions have been invaluable in overcoming difficulties and keeping my research on track.

I would also like to extend my gratitude to Dr Muhammad Saad Shakeel, Dr Maxwell Antwi-Affari, Dr Yazan Qarout, Dr Aamir Arsalan, Dr Khursheed Ahmed, Dr Mehran Khan, Dr Muhammed Aquil Mirza, Mr. Aamir Mehmood, Mr. Muhammad Rehan Naseer, Mr. Husnain Tansar, Mr. Zahid Javed, Mr. Muhammad Zeeshan, Mr. Umar Mansoor, Mr. Abasal Hussain, Mr. Muhammad Ayaz Akbar, Mr. Hafiz Muhammad Habib and all my colleagues in the Smart Construction Lab for their extended help and encouragement throughout my research studies.

Furthermore, I would like to acknowledge the research funding provided by the following three research funds to complete this research work: 1. General Research Fund (GRF) Grant (15201621) titled "Monitoring and managing fatigue of construction plant and equipment operators exposed to prolonged sitting"; 2. General Research Fund (GRF) Grant (15210720) titled "The development and validation of a noninvasive tool to monitor mental and physical stress in construction workers"; and 3. Research Institute for Intelligent Wearable System (RI-IWEAR) - Strategic Supporting Scheme (CD47) titled "An Automated Assessment of Construction Equipment Operators' Mental Fatigue based on Facial Expressions".

I would like to dedicate this work to my late father's loving memory, whose unwavering support and belief in my abilities has been a guiding light throughout my academic journey. His wisdom, encouragement, and sacrifices shaped the person I am today, and I am forever grateful for his unconditional love. Finally, I would like to acknowledge the unconditional affection and guidance of my parents, sibling, wife, and son. Without their assistance, I would have been unable to complete the journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction[1]

1.1. Introduction

This chapter provides context of the research, defines research topic, provides limitation to the current research, specifies aim and research objectives, research scope, explains research design and structure of the thesis.

1.2. Background

The construction industry (CI) has made substantial contributions to the development of countries with over 350 million workers worldwide (Birhane et al., 2022). Despite its significant positive impact, workforce safety in this industry remains a persistent and unresolved challenge. The construction industry has gained a reputation for its poor safety performance (Ke et al., 2021b) and is recognized as one of the most hazardous sectors (Hinze and Teizer, 2011) , characterized by complexity, uncertainty, and disorderliness. Construction projects are particularly challenging because of the unpredictable and uncertain environment in which they are conducted (Choi et al., 2020, Laitinen and Päivärinta, 2010). The dynamic nature of construction sites, with the daily activities of workers, materials, and equipment, creates unique conditions (Zhu et al., 2016) that make construction workers more vulnerable to accidents than other industries (Albert et al., 2020). Consequently, accidents in the construction industry occur frequently (Koc and Gurgun, 2022), leading to an excessively high accident rate globally (ILO, 2022). In the United States, approximately 20% of fatal accidents and 40% of fatal accidents in

---

[1] This chapter is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" *Advanced Engineering Informatics*, Volume 54, 101777
Imran Mehmood, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". *Advanced Engineering Informatics*, Volume 56, 101978
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". *Developments in the Built Environment,* Volume 15, 100198
Imran Mehmood, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". *Journal of Safety Research*, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291

Singapore occur in the construction industry (Feng et al., 2015, OSHA, 2019). Similarly, the construction industry in Hong Kong reported 2947 and 2532 accidents in 2019 and 2020, respectively, with the highest number of fatalities and accident rates among all industrial sectors in the first three months of 2022 (Labor, 2022). Pakistan has also experienced a significant increase in accidents in the construction industry, making it the second most accident-prone industry compared to others, with a rising percentage of accidents in recent years (PBS, 2015, PBS, 2018, PBS, 2021). Safety concerns persist in the Chinese construction industry, which accounts for over a third of all recorded incidents, and the number of accidents and deaths has remained high (CLB, 2020). Additionally, the People's Republic of China's Ministry of Emergency Management reported in 2018 that the total number of accidents had increased year-on-year and has remained high. Furthermore, the accident and death rates increased by 7.8 percent in the first half of 2018 to 1,732 accidents and 1.4 percent to 1,752 deaths, respectively (MEM, 2018). These accidents not only result in serious injuries and fatalities but also disrupt work progress at construction sites (Sarkar et al., 2020). Among these incidents, construction equipment-related accidents constitute a significant portion and are undeniably one of the most common types of construction accidents (Li et al., 2021a, Li et al., 2017b). (Li et al., 2021a, Li et al., 2017b). In the United States, construction equipment is the leading cause of work-related fatalities and injuries (Vahdatikhaki et al., 2019). According to Hinze and Teizer (2011), one in four construction industry fatalities is caused by accidents involving equipment. Likewise, OSHA also found that struck-by accidents are among the four major causes of fatalities in the construction industry. Consequently, reducing the occurrence of equipment-related incidents at construction sites is crucial for mitigating fatalities and injuries in the construction industry.

1.3. Mental fatigue and construction industry

Human behavior is a leading cause of construction equipment-related accidents (Ma et al., 2021b), and fatigue states have a significant influence on such behavior (Behrens et al., 2023, Yang et al., 2021, Molan and Molan, 2021, Bucsuházy et al., 2020). Bai and Qian (2021) reported that more than 65% of accidents can be attributed to human error. As defined by Brown (1994), fatigue is a state of energy depletion that hinders task-directed efforts and diminishes attentiveness. This poses risks to workers' health and safety (Williamson et al., 2011), leading to reduced energy levels and increased fatigue during and after work (Frone and Tidwell, 2015). There are two primary types of fatigue: physical and mental fatigue (Villani et al., 2022). Physical fatigue is characterized by feelings of tiredness, weakness, or lack of energy resulting from physical exertion such as exercise or manual labor (Alghadir and Anwer, 2015). It increases the risk of construction accidents and occupational injuries owing to impaired worker judgment in dynamic environments (Wu et al., 2018, Umer et al., 2018, Adane et al., 2013, Chan, 2011). Similarly, mental fatigue occurs when the brain engages in intellectually demanding tasks for extended periods, leading to decreased behavioral and cognitive performance (Borragán et al., 2016, Boksem and Tops, 2008, van der Linden et al., 2003). Mental fatigue is particularly relevant in occupations that require cognitive activity and vigilance, such as long-distance driving (Hu and Lodewijks, 2020), airport baggage screening (Chavaillaz et al., 2019), and nursing during prolonged shifts (Farag et al., 2022). Both physical and mental fatigue have negative effects on performance, safety, and well-being (Chen and Hsu, 2020). Although physical and mental fatigue have different sources, their symptoms may be similar, including decreased energy, motivation, and impaired performance (Behrens et al., 2023, Van Cutsem et al., 2017). In the construction industry, various challenging tasks such as excavation, material lifting, and compaction rely on construction equipment. These tasks require cognitive effort and equipment operators to maintain sustained attention and alertness (Li et al., 2020d). Wagstaff and Sigstad Lie (2011) highlighted that the prolonged operations and demanding tasks in construction lead to mental fatigue among equipment operators, resulting in their inability to sustain the necessary attention for equipment operations. This impaired judgment and focus (Das et al., 2020) leads to decreased productivity and performance (Masullo et al., 2020), making equipment operators more vulnerable to equipment-related incidents, injuries, and fatalities at construction sites. Therefore, preventing inattention among construction equipment operators is crucial to enhance site safety (Han et al., 2019). Consequently, the continuous monitoring of mental fatigue in construction equipment operators is imperative to enable prompt responses from safety personnel when needed.

1.4. Mental fatigue assessment in the construction industry

Safety is an essential requirement for individuals engaged in construction. Proactive safety management has become indispensable in ensuring the well-being and protection of construction equipment operators with the aim of preventing accidents (Hallowell et al., 2013, Carbonari et al., 2011). Previous studies have focused on monitoring and analysing mental fatigue at construction sites using psychological or physiological techniques. Initially, subjective assessments of mental fatigue relied on questionnaires, with the NASA-TLX tool being widely used (Li et al., 2019b). However, this assessment is intrusive in nature and time-consuming (Umer et al., 2020). Further, it lacks accuracy as it is prone to bias (Han et al., 2019). As a result, researchers were motivated to develop a more objective assessment of mental fatigue. As such, this assessment was not suitable for continuous monitoring of mental fatigue since it hampers the routine work of operators, it is intrusive in nature, time-consuming, and is based on biased self-reporting of workers; hence, it lacks accuracy (Umer et al., 2020, Han et al., 2019). Consequently, researchers have sought to develop objective methods for assessing mental fatigue. Wearable sensors have gained considerable attention owing to technological advancements that enable objective monitoring of mental fatigue at construction sites. Researchers have made efforts to evaluate workers' physiological signals, such as electroencephalograms (Jeon and Cai, 2022, Ke et al., 2021a, Wang et al., 2019b), electrodermal activity (Umer, 2022, Lee et al., 2021, Choi et al., 2019), eye tracking (Noghabaei et al., 2021, Li et al., 2020d, Han et al., 2020) and electrocardiograph (Umer et al., 2022, Zhao et al., 2012), to assess mental fatigue. These signals have shown a significant correlation with workers' mental states and can be reliably used to identify fatigue among construction workers. Physiological signals and their associated parameters change when an individual's mental state fluctuates (Dziuda et al., 2021). Therefore, these signals have the potential to effectively monitor mental fatigue in a construction setting.

1.5. Limitations of current assessment methods

Although these technologies have demonstrated promising results in the diagnosis of mental fatigue, several drawbacks are associated with their use. One issue is that these devices must be worn by equipment operators, which makes them invasive and can cause discomfort during their tasks (Li et al., 2020d). Moreover, these techniques rely on electrical conductivity and are susceptible to the harsh conditions found at construction sites In addition, Chen et al. (2015) stated that these wearables often require skin preparation for sensor placement and may require limited physical activity to minimize artifacts. Another limitation is the poor spatial resolution of certain technologies such as electroencephalography (Kaur et al., 2022). Because electrodes measure surface activity, it is uncertain whether the signals originate from superficial or deep brain regions. Additionally, many studies have been conducted in simulated scenarios, which restricts their applicability and reliability in real construction environments and equipment operators, for instance the studies by Liu et al. (2021a) and Li et al. (2019b). This limits their practical use in fatigue detection (Shi et al., 2017). Therefore, there is a need to bridge this knowledge gap by developing automated, non-invasive methods that can detect mental fatigue in equipment operators without interfering with their ongoing tasks. Implementing a cost-effective and automated warning system for monitoring the mental fatigue of construction equipment operators would contribute to enhancing the safety of construction sites.

1.6. Problem solving approach.

Accordingly, this study proposes the use of non-invasive and contact-free measurements of construction equipment operators' facial features as a means of detecting mental fatigue. Previous research by Ma et al. (2021a) has shown that the human face not only reveals personal information but also indirectly reflects emotions. Dziuda et al. (2021) demonstrated the effective and contactless detection of fatigue through a continuous analysis of facial images of drivers while they are driving. Similarly, Cheng et al. (2019) concluded that observing facial expressions and cues can provide insight into stress and fatigue levels. Previous studies have highlighted the utility of facial features for detecting fatigue. In the early 1990s, the percentage of time spent with closed eyes ranging from 80% to 100% was used to study driver fatigue (Daza et al., 2014, Zhang and Zhang, 2010b). Subsequent studies considered eye closure ranges of 70–100% (Lin et al., 2015) and 75–100% (Henni et al., 2018). Other indicators of mental fatigue related to eyes include eye aspect ratio (Kuwahara et al., 2022, El Kerdawy et al., 2020), blinking rate (Bachurina and Arsalidou, 2022, Zargari Marandi et al., 2018), and eye distance (Giannakakis et al., 2017). Wang et al. (2018) have emphasized that a significant amount of information in our brains originates from our eyes, making eye behaviour a potential tool for evaluating mental state. Additionally, Chew et al. (2021) analysed gaze behaviour patterns to assess perceived workload. Nevertheless, eye blinks were also considered in recent studies on driver fatigue (Aravind et al., 2019). Similarly, Li et al. (2021b) used self-report, eye blinking rate, and R-value as indicators to substantiate the driver’s fatigue state. Furthermore, tracking the position of the driver's head provides additional information about mental fatigue, as stressful conditions result in more frequent and quicker head motion (Ansari et al., 2022, Giannakakis et al., 2018). Furthermore, research has also shown that mouth-related features such as lip movement can be influenced by fatigue (Iwasaki and Noguchi, 2016). Similarly, Giannakakis et al. (2017) reported an increased mouth activity observed in stressful situations.

Despite the potential of using automated facial features to assess the mental fatigue of construction equipment operators, there is a lack of research utilizing geometric measurements of facial features to understand operators' mental fatigue in real construction environments. Additionally, it is challenging to apply findings from other occupations, such as drivers, to monitor fatigue in excavator operators owing to substantial differences in their work patterns. For instance, excavator operators continuously move their heads to track the excavator bucket (Liu et al., 2021a), raising questions regarding whether geometric measurements of facial traits can still be effective in detecting mental fatigue in construction equipment operators under such circumstances. Therefore, the ecological validity of using geometric measures of facial features for monitoring mental fatigue in construction operators remains uncertain, highlighting the need for research in the development and testing of an objective, automated, and non-invasive method for assessing operators' mental fatigue.

To address this research gap, this study proposes a non-invasive assessment of temporal geometric measurements of facial features as a means of detecting mental fatigue. Additionally, this study compared these geometric measurements with wearable electroencephalography (EEG) measurements, which are commonly used as an invasive method to assess mental fatigue in construction workers. Previous studies by Lee and Lee (2022), Wang et al. (2022), Jeon and Cai (2022), Ke et al. (2021a), Xing et al. (2020b), Li et al. (2019a), Wang et al. (2019b), Jebelli et al. (2019d), Jebelli et al. (2018a), Hwang et al. (2018b), and Wang et al. (2017) have extensively employed EEG to monitor the mental fatigue and stress of construction workers. This comparison aims to ecologically validate the use of geometric facial features in assessing mental fatigue, specifically in construction equipment operators, ensuring their applicability during operators' routine on-site operations without interference. While previous studies in the construction industry have made efforts to address this issue effectively, advancements in wearable sensing technology now enable the continuous and appropriate monitoring of mental fatigue. However, the question remains as to which physiological indicator should be measured to yield the most reliable mental fatigue assessment findings for workers on real construction sites, which is where safety experts come into play. Furthermore, previous methods typically focused on individual physiological indicators to assess workers' mental states. In contrast, Tao et al. (2019), Charles and Nixon (2019) and Young et al. (2015) have shown that no single approach is superior when it comes to assessing mental fatigue using physiological indicators. In this context, the present study explores the feasibility of a multimodal approach for assessing mental fatigue in equipment operators during prolonged excavation operations by integrating data from multiple sensors through machine learning techniques. Consequently, the proposed study aims to improve the current non-invasive assessment of mental fatigue through contact-free measurements. Moreover, this study can contribute to the development of a real-time system for classifying mental fatigue, thereby enhancing safety and health management at construction sites.

1.7. Research Scope

Fatigue is a state of energy depletion that hinders task-directed efforts and diminishes attentiveness (Brown, 1994). This poses risks to workers' health and safety (Williamson et al., 2011), leading to reduced energy levels and increased fatigue during and after work (Frone and Tidwell, 2015). There are two primary types of fatigue: physical and mental fatigue (Villani et al., 2022). Physical fatigue is characterized by feelings of tiredness, weakness, or lack of energy resulting from physical exertion such as exercise or manual labor (Alghadir and Anwer, 2015). However, mental fatigue occurs when the brain engages in intellectually demanding tasks for extended periods, leading to decreased behavioral and cognitive performance (Borragán et al., 2016, Boksem and Tops, 2008, van der Linden et al., 2003). In the construction industry, various challenging tasks such as excavation, material lifting, and compaction rely on construction equipment. These tasks require cognitive effort and equipment operators to maintain sustained attention and alertness (Li et al., 2020d). Wagstaff and Sigstad Lie (2011) highlighted that the prolonged operations and demanding tasks in construction lead to mental fatigue among equipment operators, resulting in their inability to sustain the necessary attention for equipment operations. This impaired judgment and focus (Das et al., 2020) lead to decreased productivity and performance (Masullo et al., 2020), making equipment operators more vulnerable to equipment-related incidents, injuries, and fatalities at construction sites. Therefore, preventing inattention among construction equipment operators is crucial to enhance site safety (Han et al., 2019). Hence, this study focuses on mental fatigue monitoring of equipment operators at construction sites. The mental fatigue monitoring in operators is very important at construction sites. It helps the workplace safety supervisors and managers to understand the development of mental fatigue during prolonged operations and act promptly to avoid incidents on construction sites.

1.8. Aims and objectives.

Given the background and scope outlined above, the current study aims to explore the automated, non-invasive and contactless method by using geometric measurements of facial features for assessing mental fatigue in construction equipment operators.

1.9. Research Objectives

The specific objectives set to achieve the above aim of this research are as follows:

(a) To study non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features.

(b) To investigate the validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators.

(c) To explore the use of deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data.

(d) To study the multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals.

1.10. Research Design

To fulfil the research objectives, a methodology was designed to collect relevant data from construction sites and generate results. The diagram of the research approach is shown in Figure 1.1 below.

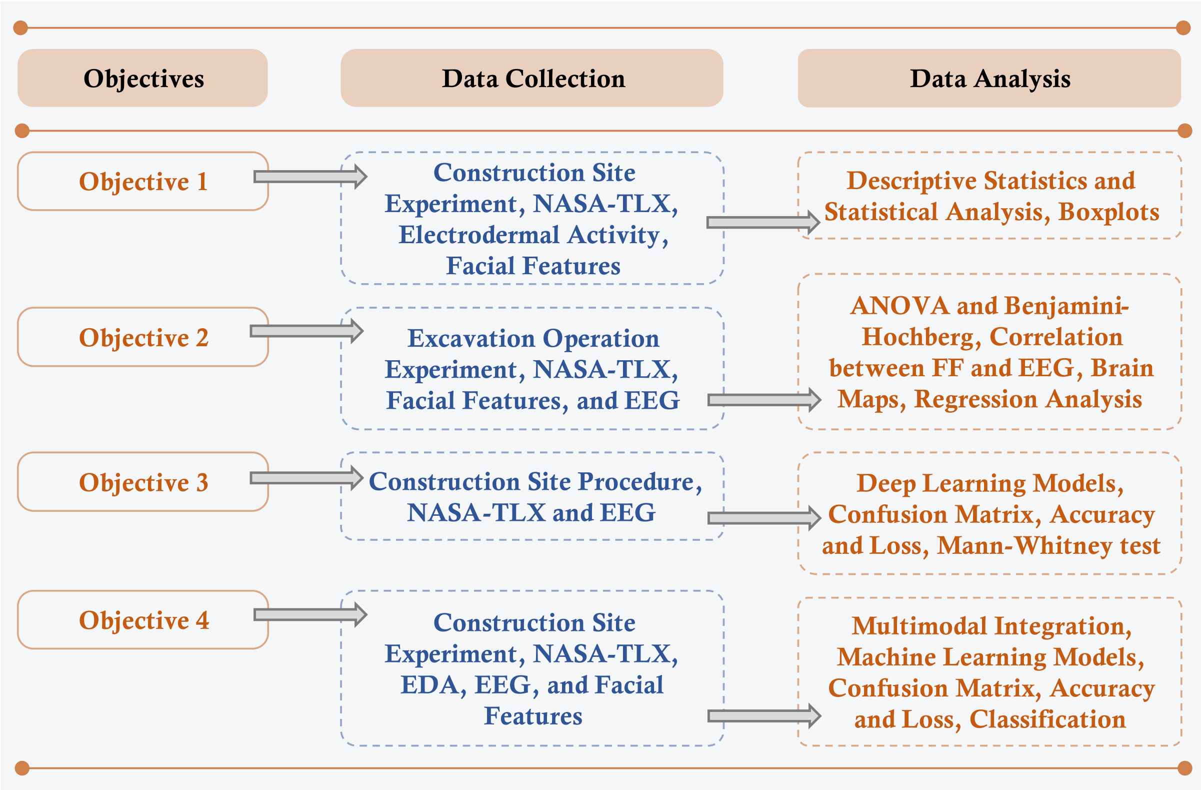


Figure 1.1: Research Design

1.11. Structure of thesis

This thesis is a collection of published papers used to accomplish the specified research objectives. This thesis is divided into six sections. Each of the papers that correlates to a study objective is introduced in Chapters 2–5 as follows.

Chapter 1: *Introduction*. The chapter discusses the research background, mental fatigue and its assessment in the construction industry, research problem, research scope, aim and objectives, research design and structure of the thesis.

Chapter 2: *Literature Review and Related Work*. This chapter comprehensively describes in detail the previous literature related to the mental fatigue assessment in the construction industry. It starts with the role of construction industry in country's economy and as well as safety issues pertaining to poor safety performance. It also describes the accidents, injuries and reason of these incidents. Moreover, an overview of mental fatigue assessment methods has been provided followed by the limitations of current assessment techniques. Furthermore, it also describes the application of current approach in other industrial domains. In addition, it also explains, how every research objective was defined by explaining the knowledge gap.

Chapter 3: *Methodology*. This chapter describes the methodology adopted to achieve every research objective. This chapter has been divided into four sections pertaining to each research objective. First, it proposes a contactless and non-intrusive method for detecting mental fatigue in construction equipment operators. A study was conducted to investigate whether there are variations in the geometric measurements of facial features owing to mental fatigue. An excavation experiment was conducted and simultaneously with the task, the operators were video recorded to collect the data on their facial features. Based on geometric measurements, facial features (eyebrow, mouth outer, mouth corners, head motion, eye area, and face area) were extracted. Second, it describes the validation of proposed method through investigations that involved a comparison with flexible headband-based wearable electroencephalography (EEG) technology. Third, it discusses methodology related to the feasibility of using deep learning techniques and raw EEG data from equipment operators to classify mental fatigue. Furthermore, it shows how the performances of different deep learning models, including long short-term memory (LSTM), bidirectional LSTM, and one-dimensional convolutional networks, were assessed using metrics such as accuracy, precision, recall, specificity, and F1-score. Lastly, this chapter explains the methodology of a novel approach of using machine learning and multimodal data fusion was proposed to classify mental fatigue. Electroencephalography, electrodermal activity, and video signals were acquired during an excavation operation.

Chapter 4: *Results*. This chapter provides the experimental results and has been divided into four sections. First, the results found that there was a significant difference in the measured metrics for high fatigue as compared to low fatigue. Second, the findings indicate a significant correlation between proposed method and electroencephalography and demonstrate that the proposed method can be employed for mental fatigue monitoring in construction equipment operators. Third, these results demonstrate the feasibility of implementing the Bi-LSTM model and contribute to the recognition and classification of mental fatigue by using wearable sensors. Lastly, the findings indicated that multimodal sensor data fusion can aid in developing a real-time system to classify mental fatigue.

Chapter 5: *Discussion*. This chapter discusses in detail the findings described in the chapter 3. It explains how the findings are in the line with the studies conducted in other domains with better performance metrics. Furthermore, this chapter also describes the limitations and future research applications.

Chapter 6: Conclusions and Contributions. The chapter provides summarized information of the current research. The contributions have also been provided. Lastly, this chapter also provides a framework that describes a systematic, step-by-step approach for applying the research outcomes on construction sites to assess mental fatigue.

1.12. Summary

This chapter provided an overview of the research, covering the study's context and the research problem. It has also highlighted the aim and objectives of the research, research scope, research design. At the end, structure of the thesis was provided.

Chapter 2

Literature Review and Related Work[2]

2.1. Introduction

Over 350 million people are employed by the construction industry worldwide, which has made significant contributions to the economic growth of many countries (Birhane et al., 2022). Regardless of their significance in boosting the economy, the health and safety on construction sites should not be underestimated (Jaafar et al., 2018). Owing to its poor safety performance, the construction industry remains one of the most hazardous industries despite ongoing efforts for improvement (Choi and Lee, 2017). For instance, the Hong Kong construction industry reported 2,947 and 2,532 accidents in 2019 and 2020, respectively (HKOSHS, 2020). Similarly, construction accounted for 20.5% of fatal workplace incidents in the EU-27 in 2018 (Eurostat, 2020). Likewise, severe construction accidents remain a key concern for other parts of the world as well, including Australia (Allison et al., 2019), China (Shao et al., 2019) and Canada (Chen et al., 2017b). Accidents occur frequently because of the unique and dynamic environment of construction projects (Koc and Gurgun, 2022), causing injuries and fatalities at construction sites (Sarkar et al., 2020). Among these, accidents related to construction equipment constitute a significant proportion (Li et al., 2021a). Statistics from the United Kingdom's construction industry show that "struck by moving equipment" accidents were the fourth-highest cause of worker injury (HSE, 2020). Furthermore, according to Vahdatikhaki et al. (2019), equipment is also a major cause of work-related fatalities and injuries in the United States construction industry. For this

---

[2] This chapter is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" *Advanced Engineering Informatics*, Volume 54, 101777
Imran Mehmood, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". *Advanced Engineering Informatics*, Volume 56, 101978
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". *Developments in the Built Environment,* Volume 15, 100198
Imran Mehmood, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". *Journal of Safety Research*, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291

reason, it is imperative to eliminate equipment-related events at construction sites by addressing the underlying causes. The construction industry typically involves a range of heavy equipment such as trucks, excavators, loaders, and tower cranes for performing construction tasks. The operations of these equipment are cognitively arduous, needs constant attention from the operator, and are a major source of mental fatigue for construction equipment operators (Li et al., 2019b). Mental fatigue is an inability of equipment operators to continue construction equipment operations due to prolonged attention. It has been associated with attention failure and working for extended hours without taking proper rest (Fang and Cho, 2017). It has been regarded as a disorder that impairs affecting states, behaviors, body responses (Goetz et al., 2022, Anwer et al., 2021), and attention functions (Lazaro et al., 2022). It has been identified as a key constraint that hampers the equipment operators' judgement and concentration, potentially resulting in accidents (Das et al., 2020). Thus, it is vital to monitor mental fatigue of construction equipment operators automatically to assist safety personnel and construction managers to act promptly when necessary.

2.2. Assessment of mental fatigue in the CI

Considering the prevalence of mental fatigue and severe consequences for construction industry workers and equipment operators, its monitoring has been extensively studied. These studies include subjective as well as objective approaches. Subjective approaches include the use of questionnaires to assess fatigue at construction sites (Techera et al., 2018) with NASA-TLX being the most extensively 19 used evaluation tool (Hart, 2006b). Moreover, unsafe behaviors and social factors leading to fatigue-led accidents were also determined using surveys and predicted using machine learning (Niu et al., 2021). However, it lacks precision because it is susceptible to individual bias (Han et al., 2019). Their usage is hindered because they are intrusive in nature, time-consuming, and are not practical for continuous fatigue monitoring (Umer et al., 2020). This prompted a search for a more quantitative method. Consequently, researchers have been encouraged to establish objective measures of mental fatigue. In recent years, wearable sensors have attracted considerable interest from researchers because of the technological advancements that enable more objective monitoring of mental fatigue at construction sites. Therefore, research has been conducted to assess mental fatigue by studying physiological signals of workers. Moreover, wearable technology has allowed researchers to overcome the limitations of subjective assessments. These technologies such as electroencephalography (EEG), photoplethysmography (PPG), electrocardiography (ECG), electrodermal activity (EDA), and eye-tracking technology (Umer et al., 2022, Noghabaei et al., 2021, Han et al., 2020, Li et al., 2019a, Ahn et al., 2019). Compared to questionnaires, physiological indicators have better performance in terms of sensitivity, diagnostic ability, and non-intrusiveness (Zhao et al., 2018). It has been shown that physiological signals can be used to reliably identify worker fatigue because of their strong correlation with workers' mental fatigue states. In recent years, the activity monitoring method has been applied to detect and evaluate mental fatigue, which mainly involves eye-tracking and psychomotor vigilance tests (Noghabaei et al., 2021, Aryal et al., 2017). Additionally, several studies have sought to automatically recognize workers' stress by applying different machine learning algorithms, such as Jebelli et al. (2018a), who studied EEG-based workers' stress recognition on construction sites. Other studies have applied eye-tracking technology to detect mental fatigue among operators, such as hazard recognition by Li et al. (2021a), situation awareness by Hasanzadeh et al. (2018), and visual search patterns by Jeelani et al. (2019). Although these technologies have provided promising results for mental fatigue detection, there are many problems associated with their application. Firstly, these technologies are invasive in nature, requiring sensors to be attached to the construction equipment operator's body. Secondly, the construction equipment operator has to wear these devices during operations at a construction site, which causes annoyance while performing operations and requires skin preparation for wearable devices (Li et al., 2020d). Thirdly, the use of these wearable biosensor technologies such as EEG requires precise positioning of the sensors, which requires equipment operators to maintain a low level of physical activity to minimize artifacts (Chen et al., 2015). Fourthly, most of these wearable devices require chargeable batteries. Lastly, studies entailing actual equipment operators on jobsites are scarce, since most of the studies are conducted in a simulated environment, such as the studies by Liu et al. (2021a) and Li et al. (2020d). Thus, there exists a knowledge gap in developing an automatic and contact-free mental fatigue detection and warning system. In contrast to the wearable sensors, this study proposes an analysis of geometric measurements of facial features of construction equipment operators with remote and contactless measurements using a mobile camera. As such, this enables non-invasive assessment of equipment operators' mental fatigue during their ongoing tasks.

2.3. Non-invasive detection of mental fatigue through geometric measurements of facial features

2.3.1. Facial features and mental fatigue assessment

Human faces carry a lot of information. Ma et al. (2021a) stated that the human face not only displays personal information directly but also indirectly conveys a variety of emotions. Mehrabian (2017) reported that facial expressions transmit 55% of the information in human interactions. Furthermore, facial expressions of emotion are a significant mode of interpersonal communication. Hence, observing and interpreting a person's facial expression in relation to their present emotional state is extremely important (Giannakakis et al., 2017). Similarly, Cheng et al. (2019), Sharma and Gedeon (2014) and Lee and Chung (2012) concluded that facial signs and expressions can provide insights into the related mental stress and fatigue. The main manifestations of mental stress on human face involve changes in the metrics related to the eyes, mouth, and the behavior of the head (Holgado et al., 2020, Zargari Marandi et al., 2018). Giannakakis et al. (2017) and Bevilacqua et al. (2018) studied head mobility features to foresee the presence of stress. It has been reported that under stressful conditions, head motions are more frequent and quicker, with a greater overall amount of head motion (Giannakakis et al., 2018, Wenhui et al., 2005, Dinges et al., 2005). It has been reported that mental stress is thought to affect eye blink patterns, which are reflected in the behavior of the ocular area (Giannakakis et al., 2019). Wang et al. (2018) also reported that behavior of the eyes can be used to detect mental states since 80% of the information received by brains comes from the eye. Likewise, Xu et al. (2018) found that eye-related features have been linked to emotional states such as anxiety. Maffei and Angrilli (2019) concluded that an individual’s blink rate could indicate how emotionally and cognitively engaged they are, especially if they are under stress. Similarly, Li et al. (2021b) used R-value, eye blink frequency, and self-report as indicators to validate the driver’s fatigue state. Furthermore, a study by (Chew et al., 2021) evaluated the perceived workload by analyzing gaze behavior patterns. There is also sufficient evidence in the research literature that stress, and anxiety have been shown to impact mouth-related aspects, such as lip movement (Iwasaki and Noguchi, 2016, Dinges et al., 2005). It has been reported that there is a faster mouth movement during stressful situations (Giannakakis et al., 2017). Metaxas et al. (2004) found that asymmetric lip deformations are an indicator of high stress. Likewise, Liao et al.

(2005a) presented that stress, as measured by increased cognitive workload, has an inverse relationship with the frequency of mouth openings. The literature suggests that stress and fatigue are related to each other, however, they are not equivalent. For example, Desmond and Matthews (2009) reported that stress and fatigue co-exist and prolonged operations may invoke stress and fatigue simultaneously during equipment operations. Similarly, Desmond et al. (2012) stated that fatigue can be considered a an output of sustained stress. Similarly, a study by Doerr et al. (2015) reported that during a period of increased stress, the levels of fatigue increased, showing that high stress raises the risk of accumulating fatigue. The association between stress and fatigue is therefore reciprocal. Our research focuses on the fatigue state of operators i.e., task specific fatigue, that develops during prolonged and monotonous time-on-task excavation operations. Such state is a major contributor towards attention failure of operators and consequently, lead to unsafe excavation operations on construction sites. Thus, excavator operators' mental fatigue is crucial in contributing to excavator operation-related on-site safety problems. Therefore, this study hypothesizes that the facial feature measures used to detect mental stress can also be used to detect mental fatigue.

2.3.2. Research objective and knowledge gap.

Unlike other industries where the working conditions are stable, construction is a dynamic and complex industry with distinct working circumstances (Xing et al., 2020a). Such conditions require an approach that monitors operators' mental health without disrupting their ongoing equipment operations. Despite the potential of automated facial features for mental fatigue assessment of construction equipment operators, there is a scarcity of research using geometric measurements of facial features to understand equipment operators' mental fatigue on real construction sites. To the author's best knowledge, the only study in the construction industry was conducted by Liu et al. (2021a). The study used a computer vision-based mental fatigue detection approach for tower crane operators in a simulated environment. However, this previous study by Liu et al. (2021a) has few drawbacks. Firstly, it was based on facial expressions such as yawning, nodding, etc. Noteworthy, in the case of construction equipment operators or drivers, these facial expressions might not show up until just before an accident (Ghoddoosian et al., 2019). Additionally, while frequent yawning is an important precursor, it is not always observed among operators in a fatigued state (Weng et al., 2017). Secondly, it was conducted in a simulated environment which cannot capture the dynamics and complexity of a real construction site. Hence, the ecological validity of this study is questionable. Thirdly, although the application of facial features has been widespread in other domains such as drivers and other occupation scenarios, its validity for construction excavator operators remains a knowledge gap. Because there is no study to apply geometric measurements of facial features to monitor construction equipment operators' mental fatigue.

Application of temporal changes in facial features for fatigue detection among excavator operators is more than re-evaluation of a method under a different context because of the nature of tasks of the operators. It is challenging to consider results from drivers for fatigue detection under excavator operations due to significant difference in the working pattern of drivers and excavator operators. According to Liu et al. (2021a), construction equipment operators' work in a fundamentally different manner than drivers. For example, during equipment operations, excavator operators move their heads continuously to track the excavator's bucket. So, it remains unknown whether such frequent head movement still leads to significant changes in facial features because of fatigue or not. As a result, it is questionable whether the geometric measurements of facial features, that have been shown to be ecologically useful for detecting mental fatigue in drivers, also apply to excavator operators in the field. Hence, their validity for construction industry is still unknown. Consequently, there exists a research gap to develop and test an approach that is objective, automatic, and non-invasive for monitoring operators' mental fatigue. To fill the research gap, the study suggests a computer vision-based facial features approach to detect construction equipment operators' mental fatigue during their site work. The approach uses the geometric measurements of facial features to detect mental fatigue. The geometric measurements of the face regions using a camera can be a non-invasive, efficient, and practical approach for real-time monitoring of excavator operators' mental fatigue. The approach is expected to improve mental fatigue detection in a non-invasive way, during ongoing equipment operations and will help safety personnel with effective and proactive responses.

2.4. Ecological validity of facial features' geometric measurements for assessment of mental fatigue in construction equipment operators

Despite the potential of automated facial features for the mental fatigue assessment of construction equipment operators, there is a scarcity of research using geometric measurements of facial features to understand equipment operators' mental fatigue on real construction sites. Additionally, it is challenging to use findings from other occupations, such as drivers, for fatigue monitoring in excavator operators due to the substantial differences between the work patterns of drivers and excavator operators. For example, during equipment operations, excavator operators move their heads continuously to track the excavator's bucket (Liu et al., 2021a). Therefore, it remains unknown whether geometric measurements of facial traits under such circumstances can still be used to detect construction equipment operators' mental fatigue. Thus, the ecological validity of the geometric measures of facial features for mental fatigue monitoring of construction operators is still questionable. Consequently, a research gap exists for the development and testing of an objective, automatic, and non-invasive method for assessing operators' mental fatigue. To fill this gap, firstly, the study proposes a non-invasive assessment of temporal geometric measurements of facial features to detect mental fatigue. Secondly, the study compares geometric measurements to wearable electroencephalography measurements, which is an established invasive method for mental fatigue assessment of construction workers. Many researchers have utilized it extensively to monitor the mental fatigue and stress of construction workers, for instance studies by Lee and Lee (2022), Wang et al. (2022), Jeon and Cai (2022), Ke et al. (2021a), Xing et al. (2020b), Li et al. (2019a), Wang et al. (2019b), Jebelli et al. (2019d), Jebelli et al. (2018a), Hwang et al. (2018b), and Wang et al. (2017). This comparison serves to ecologically validate the geometric measurement of facial features in terms of their applicability to construction equipment operators' as well as their effective use during routine operations by operators without interfering with their on-site operations. As a result, the proposed study is expected to improve the current assessment of mental fatigue in a non-invasive way through contact-free measurements.

2.5. Deep learning and EEG sensor data for mental fatigue classification during construction equipment operations

EEG is an electrophysiological monitoring system that records the electrical activities generated by cortical neurons (Sanei and Chambers, 2013). It is recognized as a potent technique in the field of construction research since it detects brain activity rapidly, cost-effectively, with a high temporal resolution, and in a portable manner (Saedi et al., 2022). There has been extensive research into the construction industry's use of EEG data gathered from wearable devices for the purpose of analyzing distinct mental states among construction workers, such as fatigue (Tehrani et al., 2021, Xing et al., 2020b, Li et al., 2019a), stress (Lee and Lee, 2022, Jebelli et al., 2019a), distraction (Ke et al., 2021b, Ke et al., 2021a), workload (Chen et al., 2017a), vigilance (Wang et al., 2019b), emotion (Xing et al., 2019, Hwang et al., 2018a), and hazard identification (Wang et al., 2022, Liao et al., 2022, Jeon and Cai, 2022). In the aforementioned studies, different mental states were analyzed and computed using either statistical methods or machine learning. Several statistical significance tests, including the Kruskal-Wallis test, the analysis of variance (ANOVA), the Mann-Whitney U test, the Wilcoxon signed-rank test, the Spearman rank-order correlation test, and the paired sample t-test, have been used to draw conclusions between experimental and control groups in EEG-based studies to compute the cognitive status of construction workers (Ke et al., 2021b, Chae et al., 2021, Xing et al., 2020b). The purpose of these analyses was to ascertain whether or not there was a statistically significant relationship between construction workers' EEG signals and their performance on the task pertaining to their mental states. Even if these assessments fared well, they have significant shortcomings when it comes to drawing reasonable and trustworthy inferences about the mental wellbeing of construction workers. Cheng et al. (2022) reported these limitations: that conventional statistical approaches are inadequate for modeling complex mapping, whereas the relationships between EEG patterns and cognitive ability are rather complex. This is mostly because the sample data used to test these statistical approaches is typically subject to stringent requirements. Therefore, the researchers turned to machine learning methods, which offer a high degree of adaptability (Rajula et al., 2020).

2.5.1. Machine learning-based mental fatigue classification

Machine learning may be utilized to compute the mental state of construction workers using their EEG signals. Various machine learning models have been developed by researchers to estimate the mental state of construction workers. For instance, using a supervised learning algorithm, Jebelli et al. (2019a) proposed a framework that can identify stress levels among construction workers and achieved an accuracy of 84.5%. Aryal et al. (2017) predicted the fatigue of construction workers with an accuracy of 82% using a boosted tree classifier. Furthermore, Hwang et al. (2018a) demonstrated that two aspects of construction workers' emotional states (arousal and valence) could be measured and quantified using EEG signals as they performed various construction-related tasks. To identify mental stress in construction workers, Jebelli et al. (2018a) compared the efficacy of K-nearest neighbors (KNNs), support vector machines (SVM), and gaussian discriminant analysis (GDA). When compared to other methods, they discovered the highest accuracy of 80.32% with SVM. In another study, Ke et al. (2021b) proposed a distraction monitoring method for construction workers, and validation was done using SVM classifier. Similarly, Jeon and Cai (2022) explored multi-class classification for hazard identification in construction workers using EEG signals in a virtual reality environment and achieved 82.3% accuracy. Selecting a suitable model and optimizing its hyperparameters are key phases in machine learning for achieving optimal results. Regardless of these advances, robust and accurate detection of construction workers' cognitive performance by EEG remains a challenge. The EEG-based studies conducted in the construction industry using machine learning were conducted offline by first measuring the data and then downloading the raw electroencephalography data for analysis (Cheng et al., 2022). In such a case, the model development may be suitable for implementation on real-time monitoring of the mental states of construction workers while they are facing dynamic site conditions (Cesa-Bianchi and Orabona, 2021). Furthermore, it is generally understood that EEG manifestations are very non-stationary and change over time within and between subjects (Thodoroff et al., 2016). Hence, recognizing overarching trends in EEG data is difficult since the signals are constantly changing (Zeng et al., 2018).

2.5.2. Deep learning applications in the construction industry

According to Türk and Özerdem (2021) and Li et al. (2020c), the ability of deep learning to analyze raw data and identify key features is its major strength. Deep learning approaches are actively applicable to different signal processing because they can learn the features from raw data and have cutting-edge performance and robust skills in creating trustworthy features in time-series data analysis (Rastgoo et al., 2019, Liu et al., 2017, Zheng et al., 2014). They have been utilized in several fields, including computer vision, natural language processing, and speech recognition (LeCun et al., 2015). Subsequently, construction-related research domains have recently shown a significant deal of interest in deep learning networks due to their outstanding performance in a variety of research areas, such as image classification (Yeşilmen and Tatar, 2022, Duan et al., 2022, Del Savio et al., 2022, Zhong et al., 2020, Yang et al., 2018), object identification and recognition (Wu et al., 2021b, Fang et al., 2018b, Fang et al., 2018a), natural language processing (Wu et al., 2022, Moon et al., 2022, Ding et al., 2022, Zhong et al., 2020, Zhang et al., 2019), and recognition of work-related risk factors (Zhao et al., 2022b, Antwi-Afari et al., 2022, Zhao and Obonyo, 2021, Wang et al., 2021, Seo and Lee, 2021, Zhao and Obonyo, 2020, Yang et al., 2020, Lee et al., 2020, Kim and Cho, 2020, Yu et al., 2019, Zhang et al., 2018). Although EEG analysis and decoding of data with deep learning algorithms have become hot research topics in recent years, unfortunately, EEG-based classification of mental fatigue using deep learning approaches has not previously been investigated for construction equipment operators on real construction sites.

2.5.3. Objective of the research

Therefore, the objective of the current research is to evaluate the feasibility of using deep learning techniques to classify construction equipment operators' mental fatigue using raw EEG data and has two major contributions. The present study represents the first attempt to acquire and analyze EEG data from construction equipment operators in real-world construction sites, thus demonstrating the feasibility and applicability of the proposed method for construction site settings. This approach enabled the authors to collect data in a natural environment, providing a more authentic and realistic context. Moreover, the study is likely to have higher external validity, which refers to the extent to which the findings of a study can be generalized. Many prior investigations on mental fatigue have been carried out in controlled laboratory settings with student participants, as exemplified by studies conducted by Li et al. (2020d) and Li et al. (2019b). However, such laboratory experiments may face challenges related to generalization and validity since they lack the dynamics and complexity of actual construction sites (Xing et al., 2020b). Therefore, the current study collected EEG data from construction equipment operators during an on-site excavation operation to support the study's findings, resulting in more comprehensive, accurate, and realistic results.

Secondly, the current study evaluates the usefulness and performance of deep learning models in detecting and classifying mental fatigue in construction equipment operators using EEG sensor data. Hypothetically, these models are more suitable for time-dependent data such as EEG signals, as they account for temporal dependencies and trends that cannot be captured using traditional classification machine learning algorithms. To the best of the authors' knowledge, no previous research in the construction industry (Cheng et al., 2022) has demonstrated the innovative approach of using deep learning models and EEG signals for detecting and classifying mental fatigue in construction workers. This is attributed to the difficulty in collecting EEG data in the field due to various factors such as noise, motion artifacts, and safety concerns, as highlighted in previous studies (Ke et al., 2021b, Ahn et al., 2019). Furthermore, the limited availability of large EEG datasets in the construction industry, as observed in studies by Wang et al. (2019b) and Jebelli et al. (2018a), may constrain the training and validating of deep learning models. These challenges have hindered the application of deep learning models in the construction industry. To overcome these challenges, the current study collected an EEG dataset using a four-channel EEG sensor and recorded one hour of EEG data from each equipment operator. This resulted in more than 18 million data points for the entire experiment, enabling the effective application of deep learning models. This gap was also filled by the current study.

There is a plethora of deep learning architectures to choose from in the literature; nevertheless, choosing the right one is crucial for EEG data processing. Recent studies by Nakagome et al. (2022), Roy et al. (2019), and Craik et al. (2019) have examined the latest trends in EEG research and identified that convolutional neural networks (CNN) and recurrent neural networks (RNN) are gaining popularity for processing EEG data. According to Nakagome et al. (2022), more than half of EEG studies used CNN or RNN, particularly with raw EEG data as input, to analyze EEG data end-to-end, eliminating the need for time-consuming feature extraction processes. Moreover, both these deep learning architectures have been effectively used in studies involving individuals exposed to external stimuli (Nakagome et al., 2022). Subsequently, this study employed and investigated the performance of three deep learning techniques, i.e., long short-term memory, bidirectional long short-term memory, and one-dimensional convolutional networks, for mental fatigue recognition in construction equipment operators. Therefore, the findings of this study are expected to provide a better understanding of the application of electroencephalography technology for mental fatigue detection in construction equipment operators in real construction scenarios based on field tests. Furthermore, using this approach, operators of construction equipment might have their mental fatigue continuously monitored without having to be observed or watched by a supervisor. Having said that, this study will also contribute to classifying construction equipment operators' mental fatigue using raw EEG data, without any human intervention for manual crafting of features. As a result, the suggested method has the potential to improve the standardization of safety management within the construction industry.

2.6. Multimodal integration for data-driven classification of mental fatigue during 14 construction equipment operations.

2.6.1. Mental fatigue as a multimodal problem

Addressing mental fatigue in construction workers is a multifaceted challenge (Ding et al., 2020). This is due to the fact that the unregulated nature of the labour-intensive construction industry poses a significant threat to workers' well-being (Ojha et al., 2023). While previous studies in the construction sector have attempted to address this problem, recent advancements in wearable sensing technology have opened new possibilities for the continuous and accurate monitoring of mental fatigue. However, determining the physiological indicator that yields the most reliable assessment of mental fatigue for workers at construction sites remains an important question to be answered by safety experts. Moreover, previous studies have assessed workers' mental states by individually investigating various physiological indicators. In contrast, Tao et al. (2019), Charles and Nixon (2019) and Young et al. (2015), have reported, no single approach has proven to be superior to others when it comes to assessing mental fatigue using physiological indicators. The same uncertainty applies to the geometric measurements of the facial features. Consequently, it remains unclear whether one physiological indicator is superior to another, or whether geometric measurements of facial features are more dependable than physiological indicators in determining a construction worker's level of mental fatigue. In light of this uncertainty, the objective of the current research is to investigate the feasibility of a multimodal data fusion approach to recognize mental fatigue in equipment operators during prolonged excavation operations.

2.6.2. Knowledge gap

First, the present study is the first to attempt to investigate a novel approach of integrating data from multiple sensors, such as electroencephalography, electrodermal activity, and geometric measurements of facial features, and machine learning techniques to classify various levels of mental fatigue. Although the concept of merging data streams from multiple sources may seem straightforward, combining data from different sensors has proven to be more accurate in predicting outcomes (Walambe et al., 2021).

Each of the aforementioned unimodal measures has its strengths and limitations in terms of accuracy and suitability for detecting worker fatigue. Hence, the integration of multiple sensor data is intended to enhance mental fatigue recognition accuracy and reduce false warnings, facilitating comprehensive and holistic monitoring of mental fatigue. The literature also supports the effectiveness of combining data from multiple sensors to assess outcomes (Zhao et al., 2022a). While research on multimodal approaches is ongoing in other industrial domains, studies investigating the classification of equipment operators' mental fatigue through the integration of multimodal sensor data, such as physiological indicators and facial features' geometric measurement, are scarce within the construction industry (Hu et al., 2023). Second, the current study acquired multimodal data in a natural setting, which provided a more realistic and authentic perspective for research. This aspect is crucial, as it enhances the study's external validity, which refers to the extent to which the findings can be generalized. Previous investigations of mental fatigue have primarily relied on controlled laboratory settings, for instance by Liu et al. (2021a), Li et al. (2020d), and Li et al. (2019b). However, conducting experiments in laboratory environments presents challenges in terms of generalization and validity, mainly because they lack the dynamic nature and complexity of construction sites (Xing et al., 2020b). To address this limitation, this study collected multimodal sensor data directly from construction equipment operators during on-site excavation operations. By capturing data in a realistic environment, the study's outcomes are more likely to reflect the complexities and nuances associated with mental fatigue in construction settings, and also hold practical relevance for understanding and managing mental fatigue among construction workers. Therefore, the current research is motivated by the need to learn more about how to recognize the mental fatigue of equipment operators holistically.

2.7. Summary

The is chapter reviewed the literature and related work done for mental fatigue assessment in the construction industry. The stats regarding construction accidents were described and mental fatigue in construction equipment is a contributing factor towards site related accidents. Furthermore, it was discussed how mental fatigue was assessed among construction workers in the previous studies. Likewise, it was explained that the current study will establish the ecological validity of the proposed method for mental fatigue assessment. Similarly, it was explained how the deep learning techniques can be employed using EEG data for fatigue classification in the construction equipment operators. Lastly, it was reviewed how multimodal analysis can be useful for fatigue classification.

Chapter 3

Methodology

3.1. Introduction

In this chapter the research methodological approach has been presented in four parts, i.e., for each research objective: (a) to study non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features (b) to investigate the validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators (c) to explore the use of deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data, and (d) to study the multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals.

3.2. Non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features[3]

Figure 3.1 depicts an overview of the proposed approach for detecting mental fatigue in construction equipment operators by using facial features collected from video recordings. To collect related data for detecting mental fatigue of construction equipment operators, an excavator operating experiment was conducted at construction site as explained in Figure 3.2. On different days, the experiment was conducted at the same time i.e., from 9:00am to 11:00am (Li et al., 2019b, Zhao et al., 2012) in the morning under similar weather conditions i.e., clear weather on all data collection days. The operators were assigned a one-hour monotonous task. It was a time-on-task excavation and discharge experiment, which is the most typical type of earthwork excavation operation. It comprised of ground excavation and material movement from pits to transportation vehicles The conditions for each excavator operator were the same, requiring them to continuously operate the equipment in the manner of a cyclic operation.

---

[3] The methodology presented in section 3.2 is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". *Journal of Safety Research*, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291

The amount of earth excavated or moved, as well as the number of vehicles filled, were not fixed, since it was a time-on-task experiment. Mental fatigue was induced using a time-on-task procedure (Hopstaken et al., 2016). Simultaneously with their tasks, the operators were video recorded to collect data on their facial features via a mobile camera, and their electrodermal activity (EDA) was measured by an Empatica E4 sensor, which is an objective indicator of fatigue. Studies have shown that electrodermal activity can be a more useful and objective index of perceived workload such as Choi et al. (2019) found that EDA can be effectively used to quantitatively assess the continuous fatigue of construction workers.

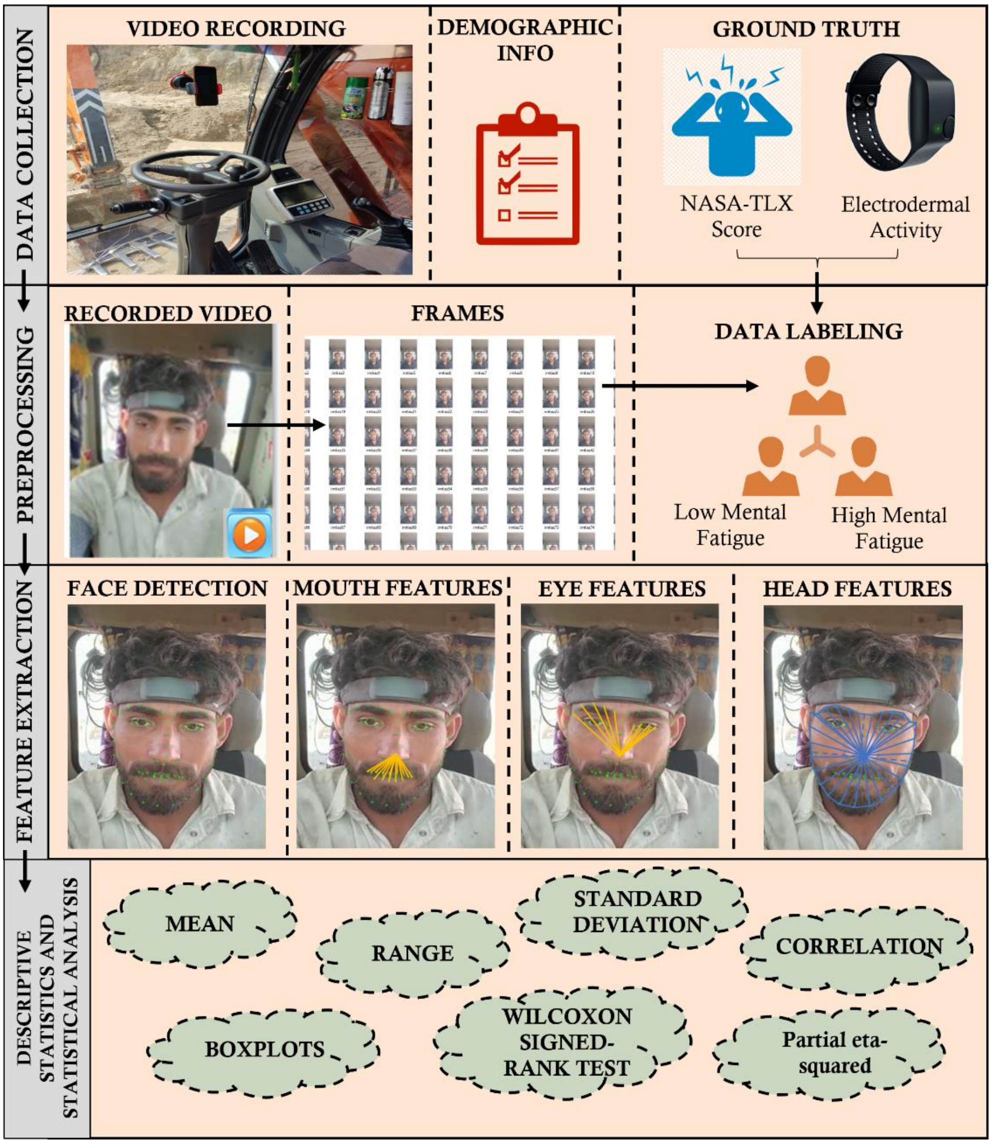


Figure 3.1: An overview of the process for detecting equipment operators' mental fatigue.

Besides, NASA-TLX score was utilized to quantify equipment operators' mental workload. It has been widely used in various research investigations since its development, and its reliability and sensitivity have been tested in a consistent number of independent tests (Hart, 2006b). The NASA-TLX is intended to measure operators' perceived workload in six dimensions: mental demand, physical demand, effort, own performance, temporal demand, and frustration. An overall NASA-TLX score was computed by adding the scores from each of the six dimensions of the scale. Overall NASA-TLX scores were used, with no weight applied to the individual categories. Adding together the subscale scores to get an overall score is a common way to simplify the original scale. (Hart, 2006b, Byers, 1989). In our study, an increase in the NASA-TLX score is result of increased mental demand owing to mental fatigue. Furthermore, studies by Das et al. (2020), Li et al. (2019b), Chen et al. (2017a) and Bitkina et al. (2021) reported that an increase in NASA-TLX score is an indicator of fatigue. Moreover, a temporal increase in NASA-TLX scores for the same task can be effectively considered as a subjective indicator of mental fatigue (Zhang et al., 2021, Kaduk et al., 2021, Li et al., 2020d, Mitropoulos and Memarian, 2013). Both the objective and subjective assessments were used as a ground truth for construction equipment operators' mental fatigue levels. The subjective assessment was collected ten minutes after the task started and then at the end of the complete task. Based on the NASA-TLX score and EDA values, the first ten minutes of operators' excavation operation were labeled as low mental fatigue whereas the same for the last ten minutes were labeled as high mental fatigue. The NASA-TLX score, and EDA values were lowest for the first ten minutes and highest for the last ten minutes of excavation operation. Low and high mental fatigue are not absolute states. They merely depict two fatigue-related temporal conditions. Furthermore, the high fatigue in our study does not imply that the operators were exhausted after a one-hour experiment. It pertains to a significant difference in the two temporal conditions of fatigue in excavation operators. Similarly, the subjective assessment (NASA-TLX score) had a broad range for the high fatigue group, ranging from 53 to 72, as contrary to 13 to 19 for low fatigue group. Additionally, the timeframe for the classification of low and high mental fatigue is not established in the literature. It can be first 5, 10 or 15 minutes and vice versa for high fatigue. For instance, Zhao et al. (2012) detected mental fatigue in the drivers when the measurements were carried out at the start and end of a 90-minute driving task. Similarly, Liang et al. (2009) studied driver fatigue by collecting physiological data for 9 minutes before and after a driving task. As a result, we investigated facial features data at two timepoints because the focus of the research was to see if there was a significant difference in geometric measures of face features between low and high fatigue states of operators. Likewise, previous studies were considered when deciding how long a monotonous construction operation task would take. Since it has been established that performing such tasks could induce mental fatigue (Thiffault and Bergeron, 2003). For instance, Li et al. (2019b) conducted a one-hour stimulated excavation operation where participants felt an increasing level of mental fatigue with the progress of experiment. Likewise, Zhao et al. (2012) reported that a driving mental fatigue was caused by a 90-minute driving task, as indicated by all physiological symptoms. However, 11 out of 13 participants mentioned that they felt tired just after first 30 minutes of the driving task. Similarly, Liang et al. (2009) monitored fatigue as a result of a driving task and concluded that having a break after one hour of driving substantially lower induced fatigue. Facial features were then extracted from each frame and artifacts were removed using normalization coefficient as shown in Figure 3.3(a). Finally, mental fatigue was detected by evaluating temporal changes in facial features using computer vision techniques as explained in the following sections.

Based on the aim of the study, we followed a within-subjects design, and all the excavator operators were treated with the same independent variable that was a time-on-task procedure. As dependent variables, we evaluated eye-related features, mouth related-features, head-related features, subjective and objective assessments of mental fatigue. Consequently, temporal changes in the geometric measurements of facial features within operators were investigated. Although the feeling of fatigue can differ amongst operators. The disparities across individuals were not the focus of this investigation.

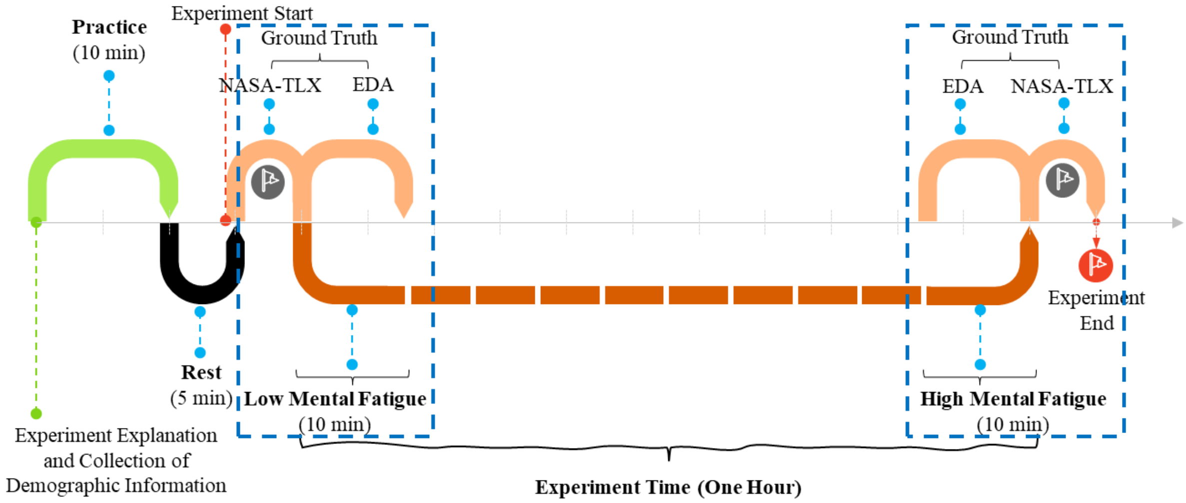


Figure 3.2: Experiment procedure for data collection

3.2.1. Participants

The experiment included seventeen male construction equipment operators with a mean age of 32.65 years (SD = 3.02), shown in Table 3.1. We determined the sample size of excavator operators to recruit for our research investigations based on sample sizes from previous studies. In earlier studies with similar purposes, 12 excavator operators (Li et al., 2019b), 12 crane operators (Das et al., 2020), 11 drivers (Ahn et al., 2016), 6 excavator operators (Li et al., 2020d), and 5 crane operators (Liu et al., 2021a) were recruited. Considering previous research in the literature, we decided that more than fifteen operators would be sufficient for our investigation and to justify our results. The participants were excavator operators with prior experience in excavator operations at construction sites. All the excavator operators had slept at least eight hours during the previous night and abstained from alcoholic drinks for at least 24 hours before experimentation. The operators were required to directly come for experiments on their designated day, and they were not involved in any other tasks or activities before the start of the experiment. Under such conditions, variations in subjective (e.g., NASA-TLX) and objective parameters (e.g., facial features) can be termed because of mental fatigue. In addition, we ensured that each operator remained fully engaged during the length of the task. All the operators had normal vision. The study was approved by the ethics subcommittee of the university (Reference Number: HSEARS20210927008) and conducted in accordance with the Declaration of Helsinki. Participants provided their informed consent.

Table 3.1: Operators' Demographic Information

	Mean	SD	Range (Min-Max)
Age (Years)	32.65	3.02	15 (26-41)
Job Experience (Years)	6.24	3.49	10 (2-12)
Height (cm)	171.47	4.24	15 (165-180)
Weight (kg)	76.41	7.66	27 (65-92)
Body Mass Index (kg/m²)	25.96	2.05	7.61 (21.80-29.41)

3.2.2. Equipment and Data Collection

3.2.2.1. Camera- based video recording

The operators were recorded while sitting in an equipment cabin with a color video camera positioned on the interior side of the equipment. The camera and operator were approximately 0.6m apart. The camera was mounted on the windscreen of equipment in such a way that there was no chance of visual interference in the routine work of the operator. The sampling frequency of the color video camera was

30 frames per second (24-bit RGB with three channels or 8-bit RGB per channel), with a resolution of 1440 x 1440 pixels.

3.2.2.2. Electrodermal activity

Electrodermal activity was intended to measure mental fatigue in excavator operators. To measure EDA, a photoplethysmography (PPG) wristwatch (Empatica E4) as shown in Figure 3.1 was used. The PPG wristwatch comprises four light-emitting diodes and four photoreceptors. Empatica E4 uses two sensors to automatically monitor fluctuating changes in the actual electrical properties of the skin, which is used to derive EDA (Milstein and Gordon, 2020). The EDA datasheet contains one column, which indicates EDA data in MicroSiemens sampled at 4 Hz (Milstein and Gordon, 2020). A MATLAB-based software, Ledalab, which is freely available, was used in the current study to derive cleaned, scaled, and meaningful EDA data. Because EDA recording is prone to the existence of different sources and forms of noise such as electrodes noise and operators' movement (Boucsein, 2012). A low-pass filter was used to reduce the most prevalent artifacts in EDA signals (Taylor et al., 2015). In addition, the EDA signals were smoothed using a high-pass filter with a cut frequency of 0.5 Hz (Braithwaite, 2013). However, large-magnitude artifacts, such as excessive electrode pressure and body motions, are not effectively filtered out by these methods (Taylor et al., 2015). To achieve this goal, a rolling filter was implemented on the EDA signals (Fitzpatrick and Kuo, 2016, Jovanovic et al., 2009). The EDA was estimated in MicroSiemens for every 500 ms with a rolling filter of 500 data points (Posada-Quintero and Chon, 2020). For further analysis, the EDA was first separated into its tonic (EDL) and phasic (EDR) components. The prior is indicative of individual variation in sympathetic arousal, whereas the latter alludes to a dynamic component of EDA that reflects quick changes in response time to external stimuli (Greco et al., 2015, Braithwaite, 2013). In our study, we used electrodermal response as a ground truth of mental fatigue. Poh et al. (2010) stated that electrodermal response is evoked by attention-demanding tasks. Furthermore, investigations by Collet et al. (2014) stated that electrodermal response can be effectively used to detect mental fatigue.

3.2.3. Data Pre-processing

Pre-processing video recordings entailed extracting the video segments comprising the actual operator's tasks. To begin with, each equipment operator's captured video was transformed into frames using OpenCV (an open-source computer vision library in Python). Since the camera's frame rate was 30 frames per second, each operator's experiment lasted for one hour. This resulted in 108,000 frames for an operator. After that, 18,000 frames from the first and last ten minutes each were identified for further analysis of the operators' mental fatigue. The frames were denoted as $I_{r,i}$ where r is the excavation operator, i represents the mental fatigue labeling for each operator and expressed as vector, $i \in \{ET_l, ET_h\}$, l for low mental fatigue and h for high mental fatigue. Hence, the pre-processing resulted in 34 segments of frames, two for each operator, keeping in view that the total number of excavator operators was 17. The labeling of equipment operators resulted into 17 $V_{r,l}$ segments (306,000 frames) for low mental fatigue and 17 $V_{r,h}$ segments (306,000 frames) for high mental fatigue. Following successful labeling, the next stage was to recognize the operator's face and extraction of facial features from each frame.

Face detection is a critical stage in facial analysis using pictures or video records. It is a necessary step for the identification of facial landmarks, the extraction of numerous facial features, face modeling, and normalization (Zhang and Zhang, 2010a, Ming-Hsuan et al., 2002). A constrained local neural field model was used to perform the facial detection process on each frame from video recordings (Zadeh et al., 2017, Baltrušaitis et al., 2016). For this model, a local neural field patch expert trains neural networks to take non-linear relationships and spatial coherence into account while assigning pixel values to every landmark. As a result, the process of facial feature identification is greatly aided by difficult and complex conditions (Johnston and Chazal, 2018, Bevilacqua et al., 2016). This model was applied to detect the operators' faces in each frame and produced a vector L of 68 landmarks identified using a face landmark detector, on the operators' faces in every frame, shown in Figure 3.3(a) and Eq 1. To localize facial regions, several face landmark detectors are being used. In our method, we utilized a pre-trained facial landmark detector based on dlib (computer vision library) to locate the 68 (x, y) coordinates of key facial parts, including eyes, eyebrows, mouth corners, and so forth (Rosebrock, 2017). The facial landmark detector is trained on the iBUG-300W dataset [X] (Sagonas et al., 2016), that was created by the manual annotation and labelling of each of the 68 coordinates on a total of 7,764 images. To begin, a training set of labelled facial landmarks on an image is used. These photos are labelled by hand, with exact (x, y)-coordinates of regions surrounding each face structure and priors based on the probability of distance between pairs of input pixels specified. As a result, a trained facial landmark detector model recognizes precise landmarks on excavator operators' face that correspond to the features such as eyes, mouth, etc. The sample landmarks detected on the excavator operators' face is shown in Figure 3.3(a). The facial parts on which the landmarks are assigned are shown in Table 3.2.

$$L = [p_1, p_2, p_3, \dots\dots\dots, p_i]^T \qquad Eq.\ 1$$

Where, p_i is a detected face landmark in any frame with coordinates (x_i, y_i), T is the number of any frame and i is the index of detected landmarks in one frame i.e., between 1 to 68. Since construction equipment operators may move towards or away from the camera during the experiment. This may be due to the vibrations produced by equipment operations and the natural movements of the operators when tracking the excavator's bucket. Such movements may affect the collected data by causing artifacts. To avoid the effects of artifacts on the collected data, it must be normalized. To tackle this issue, the facial landmarks which are more stable were used as a reference and to normalize the affected data. Based on the results of Giannakakis et al. (2017) and Bevilacqua et al. (2018), the landmarks in the nose region (i.e., the length of nose line) were used as a reference to normalize the collected data. First, the landmarks along the nose line, expressed as $Q = [p_{28}, \dots\dots, p_{32}]^T$ shown in Figure 3.3(a), were used to calculate the normalization coefficient Q as Euclidean distance of nose line, using Eq 2. Afterwards, all the features were then divided by Q and were expressed as normalized features. The normalization coefficient was calculated for every frame. The Eq 2 shows formula to calculate and Euclidean distance between any two facial landmarks.

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \qquad Eq.\ 2$$

3.2.4. Feature Extraction

The proposed approach of automated facial analysis used 68 facial landmarks to quantify six facial features. Figure 3.3(a to g) illustrates the facial traits listed in Table 3.2. In our study we used two eye related features; eye area and eyebrows, two mouth related features; mouth corners and mouth outer and two head related facial features; head motion and face area. Euclidean distances between landmarks were used to calculate these facial traits, similar to a previous study by Bevilacqua et al. (2018). It is a common and easy approach for geometric measurements of facial features (Samara et al., 2016) which is more flexible than the traditional approach since it doesn't use predefined six universal facial expressions for mental fatigue assessment. The features were calculated for each frame separately.

Table 3.2: Details of extracted facial features.

Feature	Description
Eye Area (E1)	Area of a closed polygon formed by joining the eye landmarks.
Eyebrow (E2)	Sum of the distance between anchor and eyebrow landmarks.
Mouth Outer (M3)	Sum of the distance between anchor and outer landmarks of mouth.
Mouth Corner (M4)	Sum of the distance between anchor and corner landmarks of mouth.
Face Area (H5)	Area of a closed polygon formed by joining the external landmarks on the face
Head Motion (H6)	Sum of the distance between anchor to external landmarks of face, per frame

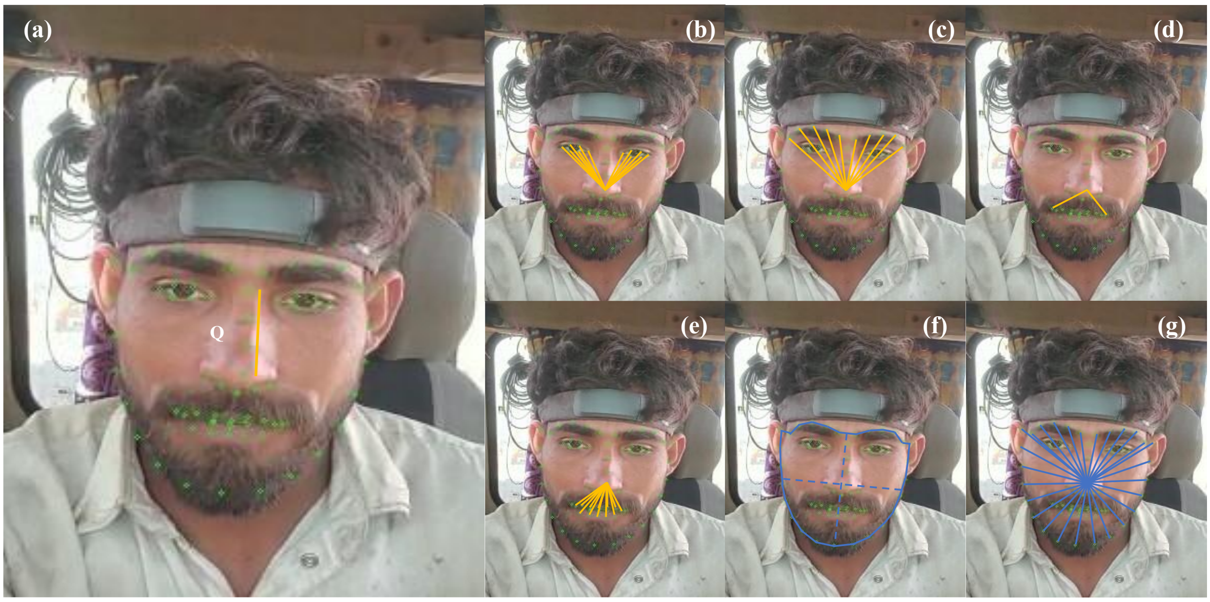


Figure 3.3: (a) identification of facial landmarks and calculation of (b) eye area (c) eyebrow (d) mouth corners (e) mouth outer (f) face area (g) head motion

3.2.4.1. Features related to eye.

These features were intended to detect the variations in the eye region involving the eye area and eyebrow as shown in Figures 3.3(b) and 3.3(c). The first feature, the eye area is the area of a closed polygon formed by joining the eye landmarks in the right (*r*) and left (*l*) eye. The area was calculated using OpenCV's function i.e., contourArea(), which is based on Green's theorem (Stewart, 2011). The area for the right and left eyes was calculated using landmarks $E1_r = [p_{43}, p_{44}, \dots \dots \dots, p_{48}]^T$ and $E1_l = [p_{37}, p_{38}, \dots \dots \dots, p_{42}]^T$ respectively, where p_i is the i^{th} landmark point of any eye at an

Euclidean distance A_j from the anchor landmark on the nose line. The second feature, eyebrow was calculated as the sum of the distance between anchor (p_{31}) on the nose line and eyebrow landmarks. The landmarks for left and right eyebrows were $E2_r = [p_{23}, \dots \dots \dots, p_{27}]^T$ and $E2_l = [p_{18}, \dots \dots \dots, p_{22}]^T$ respectively. The eyebrow feature was calculated using Eq. 3, where $E2_{r,i}$ or $E2_{l,i}$ and A_j are the i^{th} and j^{th} landmark points of $E2$ and A.

$$E2 = \frac{1}{Q}\sum_{i=1}^{E2}\sum_{j=1}^{A}\left[d\left(A_j, E2_{r,i}\right) + d\left(A_j, E2_{l,i}\right)\right] \qquad Eq.\ 3$$

3.2.4.2. Features related to mouth.

These features were intended to detect the variations in relation to mouths such as mouth corners and mouth outer as shown in Figures 3.3(d) and 3.3(e). The mouth outer was divided into the landmark points showing the outer boundary of the mouth. The mouth outer feature was calculated as the sum of the distance between anchor (p_{31}) landmark on the nose line and outer landmarks of the mouth ($M3_i$), and expressed as a vector $M3 = [p_{49}, p_{50}, \dots \dots \dots, p_{68}]^T$ using Eq. 4, where A_j and $M3_i$ are the j^{th} and i^{th} landmark point of A and $M3$.

$$M3 = {}^{1}\!/_{Q}\sum_{i=1}^{M3}\sum_{j=1}^{A} d\left(A_j, M3_i\right) \qquad Eq. 4$$

The mouth corner feature represents the corner points of the mouth as shown in Figure 3.3(d) and is expressed as a vector $M4 = [p_{49}, p_{55}]^T$. This feature is the sum of the distance between anchor (p_{31}) landmark and corner landmarks of the mouth ($M4_i$), expressed in Eq. 5, where A_j and $M4_i$ are the j^{th} and i^{th} landmark point of A and $M4$.

$$M4 = {}^{1}\!/_{Q}\sum_{i=1}^{M4}\sum_{j=1}^{A} d\left(A_j, M4_i\right) \qquad Eq. 5$$

3.2.4.3. Features related to head.

These features were intended to detect variations in the pose of the head and temporal motion of the face/head during the excavation task. The two head-related features that were calculated include the face area and head motion, shown in Figures 3.3(f) and 3.3(g). The face area is the area of a closed polygon formed by joining the outer landmarks on the face expressed as a vector $H5_i =$

$[p_1, p_2, p_3, \dots, p_{18}, \dots, p_{27}]^T$. The normalization coefficient was used to normalize the face area calculated from each frame. The head motion feature was calculated as number of pixels per frame. Firstly, the sum of Euclidean distance between the anchor to outer landmarks of the face was calculated. Afterward, the head motion feature was calculated as the difference between the two consecutive frames as given by Eq. 6, where Q is normalization coefficient, A is the length between anchor landmark and face's outer landmarks, p_a and p_b denote the Euclidean distance's sum between outer landmarks on the face and anchored landmark for current and previous frame.

$$H_{mot} = \frac{1}{Q}\sum_{i=1}^{A}|p_a - p_b| \qquad Eq.\ 6$$

3.2.5. Statistical Analysis

SPSS version 22 (IBM Inc., Chicago, IL) was used to analyse the data. This study conducted a statistical analysis based on six facial features for mental fatigue detection including eye area (E_1), eyebrow activity (E_2), mouth outer (M_3), mouth corners (M_4), face area (H_5), head motion (H_6) and NASA-TLX score, and EDA values. Since for each equipment operator, eighteen thousand frames were extracted, each for low and high mental fatigue states, one value of each facial feature was calculated from each frame, which resulted in a dataset of eighteen thousand values for each facial feature. The normality of the collected geometric measurements of facial features was tested by conducting Shapiro-Wilk test assuming it to be parametric. However, the results were otherwise, indicating that they confirm to non-parametric analysis. Therefore, Wilcoxon Signed-ranked test was conducted to compare the mean values of low and high mental fatigue. Additionally, the outliers if any, were processed and removed from the collected data using Z-scores (Venkataanusha* et al., 2019) which is the most common tool to determine the usual and unusual data points in the collected data. After that, the mean value of facial features was calculated along with standard deviation, and range of mean values for each mental fatigue group. The effect size of mean values for each feature and ground truth, between mental fatigue levels was analyzed using partial eta-squared (η^2). Correlation between the ground truth i.e., NASA-TLX score, and EDA, was also calculated for low and high mental fatigue, respectively. Moreover, mean values for each facial feature were compared between low and high mental fatigue using Wilcoxon signed-rank test. In addition, Pearson correlation coefficients were computed between the mean values of geometric face feature measures and subjective mental fatigue scores.

3.3. Validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators[4]

The overview of the research process and experiment procedure is depicted in Figure 3.4 and Figure 3.5, respectively. It shows the proposed approach for identifying mental fatigue in construction equipment operators by using geometric measurements of facial features collected through video recordings. An excavator operating experiment was conducted at a construction site to collect related data for detecting the mental fatigue of construction equipment operators. On different days, the experiment was conducted at the same time, i.e., from 9:00am to 11:00am (Li et al., 2019b, Zhao et al., 2012) in the morning under similar weather conditions, i.e., clear weather on all data collection days. The experiment was based on a monotonous and prolonged excavating and discharge task on a

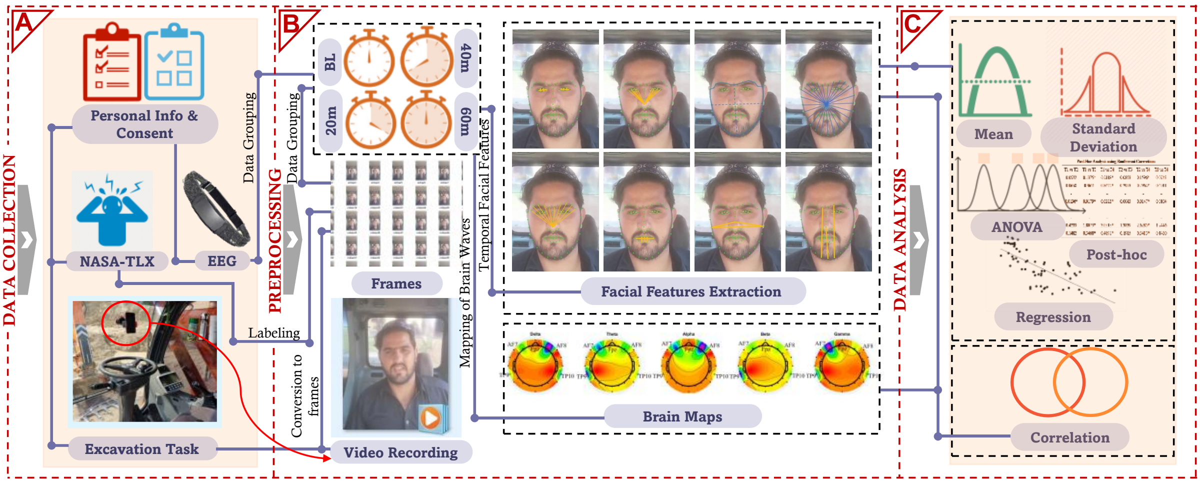


Figure 3.4: Overview of the research process

construction site. All the excavator operators were directed to complete a monotonous and prolonged excavation task for an hour, which included ground excavation and moving the material from pits to transport vehicles. Mental fatigue was induced using the time-on-task procedure. Simultaneously with their tasks, the operators were video recorded to collect data on their facial features via a mobile camera. Besides, the NASA-TLX score was utilized to quantify the subjective assessment of equipment

[4] The methodology presented in section 3.3 is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" *Advanced Engineering Informatics*, Volume 54, 101777

operators' mental workload. The subjective mental fatigue levels were assessed at the start as a baseline measurement and every 20 min for the one-hour experiment (i.e., at 20, 40, and 60 min). Geometric measurements of facial features were then extracted from each frame, and artifacts were removed using a normalization coefficient Q. It is a Euclidean distance along the nose line. Apart from visual cues, EEG data for each equipment operator was also collected for every experiment phase. For the purpose of statistical analysis, since the subjective mental fatigue levels were assessed at baseline and every 20-min experiment phase, the continuous real-time data of facial features from video frames and EEG sensor data was averaged for the respective time points (i.e., at 20, 40, and 60 min), as shown in block-B of Figure 3.1. Mental fatigue was detected by evaluating temporal changes in facial features and through EEG sensors between the time points. Finally, the detected mental fatigue with EEG and

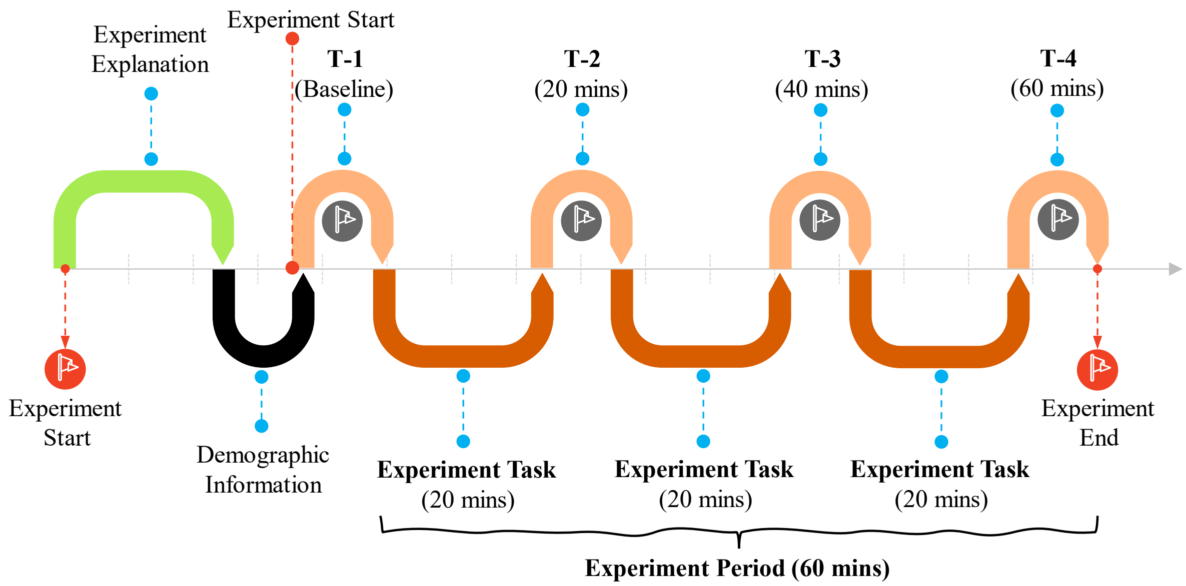


Figure 3.5: Experiment procedure; Assessments through NASA-TLX, facial features, and EEG at T-1, T-2, T-3,

geometric measurements of facial features were correlated to develop ecological validity for construction equipment operators.

3.3.1. Equipment Operators

Sixteen construction equipment operators with a mean age of 32.63 years (SD = 4.11) were included in the on-field data collection. We determined the sample size of excavator operators to recruit for our research investigations based on sample sizes from previous studies. In earlier studies with similar purposes, 12 excavator operators (Li et al., 2019b), 12 crane operators (Das et al., 2020), 11 drivers (Ahn et al., 2016), 6 excavator operators (Li et al., 2020d), and 5 crane operators (Liu et al., 2021a) were recruited. Considering previous research in the literature, we decided that more than fifteen operators would be sufficient for our investigation and to justify our results. In addition, the results showed statistically significant differences, demonstrating that the sample size was adequate to infer valid conclusions. Furthermore, all the excavator operators who participated in the study were male. All the equipment operators were excavator operators, with prior experience of excavation operations at construction sites. The excavator operators indicated in their self-report that they were well rested and in good health. All the excavator operators reported having slept at least eight hours during the previous night and abstained from alcoholic drinks for 24 hours before experimentation. On their assigned day, the operators were to report directly to the experiments and perform no other duties or activities prior to the commencement of the experiment. The recruited excavation operators had normal vision and provided informed consent before the data collection. The study was approved by the ethics subcommittee of the Hong Kong Polytechnic University (Reference Number: HSEARS20210927008) and conducted in accordance with the Declaration of Helsinki. Table 3.3 provides the demographic information of the excavation operators.

Table 3.3: Construction equipment operators' demographic information

	Mean	**SD**	**Range (Min-Max)**
Age (Years)	32.63	4.11	13 (26-39)
Job Experience (Years)	7.44	2.90	9 (2-11)
Height (cm)	174.50	5.06	18 (166-184)
Weight (kg)	77.31	5.99	23 (68-91)
Body Mass Index (kg/m^2)	25.43	2.29	8.30 (21.46-29.76)

3.3.2. Equipment and Measurement

3.3.2.1. Subjective assessment scales

The NASA-TLX score was used for the labeling of construction equipment operators by assessing their individual subjective feelings of mental fatigue. The NASA-TLX score was utilized to quantify equipment operators' mental workload. It has been widely used in various research investigations since its development, and its reliability and sensitivity have been tested in a consistent number of independent tests (Hart, 2006b). Likewise, studies by Liu et al. (2016) and Puspawardhani et al. (2016) also stated that NASA-TLX is a popular component of research studies since it is reliable and easy to use. Furthermore, temporal increase in NASA-TLX scores for the same task is considered as a subjective indicator of mental fatigue (Li et al., 2020d). The subjective assessment was used as a ground truth for construction equipment operators' mental fatigue levels and was used to compare temporal outcomes of facial features' geometric measurements.

3.3.2.2. Video recordings

A color video camera was mounted on the inner side of the excavator to film the operators while they sat in the cabin. The approximate distance between the operator and the camera was 0.6m. The camera was installed on the windscreen of the equipment in such a manner that the operator's usual work was not disrupted by its presence. The sampling frequency of the color video camera was 30 frames per second (24-bit RGB with three channels or 8-bit RGB per channel), with a resolution of 1440 x 1440 pixels. Furthermore, unlike other industries where the working conditions are stable, construction is a dynamic and complex industry with distinct working circumstances (Xing et al., 2020a). In this case, variations in illumination or non-uniform lighting conditions can impair facial detection performance. As discussed in the manuscript, the performance of our method depends heavily on the accurate localization of facial landmarks, which are hard to detect in low-light environments. Furthermore, we collected data from the real construction site at the same time on separate days while keeping weather forecasts in mind to avoid the extreme impacts of illumination. As a result, the overall effect of illumination and temperature was comparable for all operators. Furthermore, on days during data collection, the average minimum and maximum temperatures were 29.1°C and 30.4°C, respectively. Additionally, on all days, the weather was clear.

3.3.2.3. EEG Data Collection

We used the Muse headband, which is a flexible and easy-to-use EEG recording system, to acquire EEG signals. It is a headband with four channels and dry electrodes at AF7, AF8, TP9, and TP10. FPz, being the reference electrode, is placed at the forehead position. The material used for the electrodes is silver. The Muse headband records EEG data at a sampling rate of 256 Hz. The Muse headband was linked to a smart phone through Bluetooth so that data could be transmitted. Using an app called "Mind Monitor," EEG data was recorded on a smart phone and then sent to a PC to be processed later (Arsalan et al., 2019).

3.3.3. Pre-processing of the data

3.3.3.1. Data Labelling and Facial Feature Extraction

All the operators were video recorded for one hour while performing excavation operations at the construction site. Initially, each operator's captured video was transformed into frames using OpenCV (an open-source computer vision library in Python). This resulted in 108,000 frames for each operator during the whole experiment since the frequency of the camera was 30 frames per second. Subsequently, these frames for each operator were divided into four groups as per the experiment phases, i.e., baseline, 20, 40, and 60 min for further analysis. The frames were then denoted as $F_{o,p}$ where o is the excavation operator, p represents each experiment phase and expressed as vector, $p \in \{ET_1, ET_2, ET_3, ET_4\}$, 1 for baseline, 2 for data at 20 min, 3 for data at 40 min and 4 for data at 60 min. Hence, the pre-processing resulted in 16 segments of frames for each experiment phase, owing to the number of operators being 16 and each operator's data being divided into four groups. Thus, the total number of frames processed was 1,728,000. Following the successful division of frames into experiment phases, the next stage was to recognize the faces in each frame and extract the respective facial features for further analysis. The facial detection process was performed on each frame from the video recording using a local constrained neural field model (Baltrušaitis et al., 2016). This model was applied to detect the operators' face in each frame and produced a vector L of 68 landmarks identified on the operators' face in every frame using Dlib (King, 2009) and expressed as a vector $L = [q_1, q_2, q_3, \dots \dots \dots, q_i]^T$. Where q_i is a detected face landmark in any frame with coordinates (a_i, b_i), T is the number of any frame, and i is index of detected landmarks in any frame, i.e., between 1 to 68. Eq. 1 was then used to compute the Euclidean distance between any two landmarks. This Euclidean distance was eventually used to determine the geometric measurement of eight facial features, as in the previous studies conducted by Cech and Soukupova (2016) and Bevilacqua et al. (2018). The proposed eight facial features were computed separately from each individual frame, and the details of the eight facial features have been listed in Table 3.4 and shown in Figure 3.6.

Table 3.4: Details of extracted facial features.

Feature	Equation
Eye Aspect Ratio (EAR): Ratio of height and width of an eye	$EAR = \frac{\lVert p_{42} - p_{38} \rVert + \lVert p_{41} - p_{39} \rVert}{2\lVert p_{40} - p_{37} \rVert}$

<u>Eye Distance (ED)</u>: Sum of the distance between anchor and eye landmarks.	$ED = \lVert p_{37} - p_{31} \rVert + \lVert p_{38} - p_{31} \rVert + \lVert p_{39} - p_{31} \rVert + \lVert p_{40} - p_{31} \rVert + \lVert p_{41} - p_{31} \rVert + \lVert p_{42} - p_{31} \rVert$
<u>Eyebrow Distance (EBD)</u>: Sum of the distance between anchor and eyebrow landmarks.	$EBD = \lVert p_{23} - p_{31} \rVert + \lVert p_{24} - p_{31} \rVert + \lVert p_{25} - p_{31} \rVert + \lVert p_{26} - p_{31} \rVert + \lVert p_{27} - p_{31} \rVert$
<u>Mouth Aspect Ratio (MAR)</u>: Ratio of height and width of mouth	$MAR = \frac{\lVert p_{68} - p_{62} \rVert + \lVert p_{67} - p_{63} \rVert + \lVert p_{66} - p_{64} \rVert}{3\lVert p_{55} - p_{49} \rVert}$
<u>Nose to Jaw Ratio (NJR)</u>: Distance between anchor landmark and jaws	$NJR = \frac{\lVert p_{31} - p_{3} \rVert}{\lVert p_{15} - p_{3} \rVert}$
<u>Nose to Chin Ratio (NCR)</u>: Distance between anchor landmark and chin	$NCR = \frac{2\lVert p_{31} - p_{9} \rVert}{\lVert p_{22} - p_{8} \rVert - \lVert p_{23} - p_{10} \rVert}$
<u>Face Area (FA)</u>: Area of a closed polygon formed by joining the external landmarks on the face	$FA = \frac{1}{Q}\sum_{i=1}^{N=27}\left(S\left(S - d(p_1, p_{31})\right)^2\left(S - d(p_2, p_{31})\right)^2\left(S - d(p_1, p_2)\right)^2\right), \quad \therefore S = \frac{d(p_1, p_{31}) + d(p_2, p_{31}) + d(p_1, p_2)}{2}$
<u>Head Motion (HM)</u>: Sum of the distance between anchor to external landmarks of face, per frame	$H_{mot} = \frac{1}{Q}\sum_{i=1}^{A}\lvert p_a - p_b \rvert$

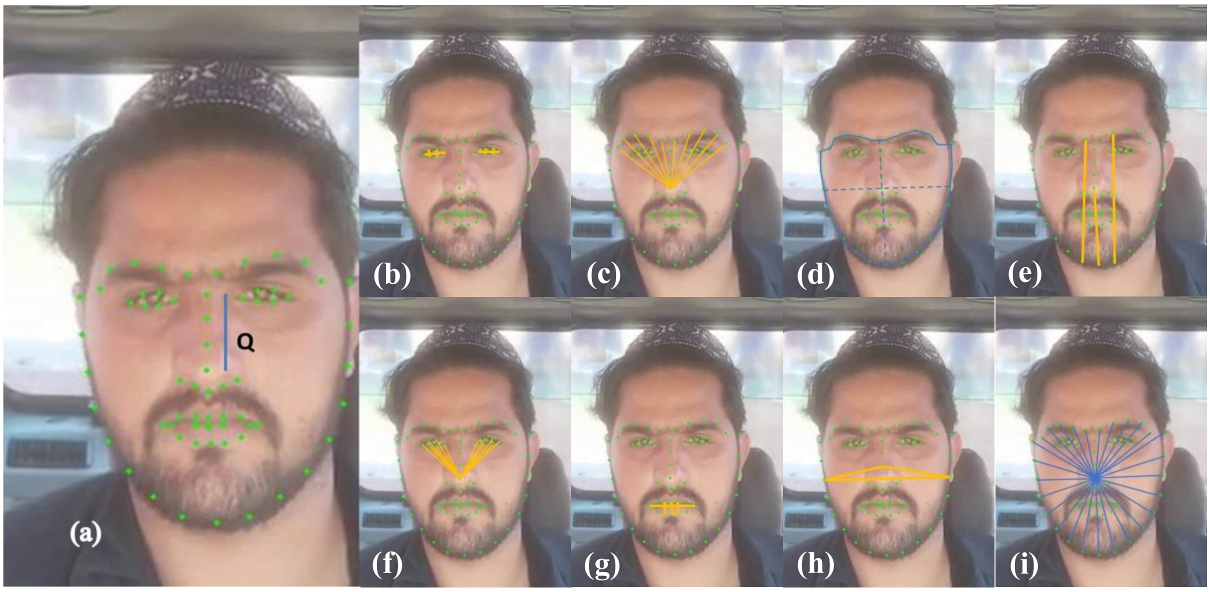


Figure 3.6: Extraction of facial features; (a) 68 landmarks detection, (b) eye aspect ratio, (c) eyebrows (d) face area (e) nose-to-chin ratio (f) eye distance (g) mouth aspect ratio (h) nose-to-jaw ratio (i) head motion

3.3.3.2. Artifacts Removal

The data collected even in the experimental setting contains artifacts, which are undesired variations in the collected data due to external sources (Sweeney et al., 2012). These artifacts need to be removed since their existence within the data may easily misinterpret it and create skewness in analysis (Jebelli et al., 2018b, Hwang et al., 2018b). In the case of excavator operators, they undergo continuous excessive and extreme movements during ongoing excavation operations. These movements are due to equipment vibrations as well as the movements of operators when tracking bucket to excavate and dump the earth. Such movements cause artifacts that need to be removed from the collected data. In the case of facial recognition and facial feature extraction, the facial regions having stable values are used for artifact removal. As reported in the research by Bevilacqua et al. (2018) and Giannakakis et al. (2017), the length of the nose line formed by joining the nose landmarks expressed by vector $Q = [q_{28}, \dots \dots, q_{32}]^T$ was used to remove artifacts, shown in Figure 3.6(a). Firstly, the landmarks shown by the vector Q were used to calculate the Euclidean distance (expressed as Eq 1) of the nose line. After that, all the facial features were then divided by Q to get normalized facial features from each frame.

$$d(q_1, q_2) = \sqrt{(a_2 - a_1)^2 + (b_2 - b_1)^2} \qquad Eq.\ 1$$

The recorded EEG signals are subjected to artifact removal techniques to remove muscular artifacts, power line noise, and other artifacts. Before analyzing the EEG data, it was subjected to preprocessing in which all the possible artifacts (muscular, power line, head motion, and eye movement artifacts) that could contaminate the EEG signal were removed as follows. Firstly, the MUSE EEG headband has an on-board noise cancellation mechanism to filter out the noise based on the statistical properties of the data. The statistical properties used by the MUSE headband include amplitude, variance, and kurtosis. An EEG signal is considered clean if its statistical properties are below a predetermined threshold; otherwise, the signal is considered noisy and discarded. Furthermore, an SG filter was used to smooth out the EEG signals that were recorded while keeping the strength of the signals. The Savitzky-Golay (SG) filter is a good way to smooth out data because it is based on the least square polynomial approximation principle (Savitzky and Golay, 1964). Different frequency (delta (0–4 Hz), theta (4–7 Hz), alpha (8–12 Hz), beta (12-30 and beta-30) bands were used to translate the pre-processed EEG data into different frequency bands using the MUSE on-board signal processing module. The mechanism used in this study for the noise cancellation of the EEG signal has been found quite effective in several EEG studies in the literature (Raheel et al., 2021, Raheel et al., 2020, Abd Rahman and Othman, 2016).

3.3.4. Data Analysis

The data was analyzed using SPSS version 22 (IBM Inc., Chicago, IL) and statistical analysis was performed based on eight facial features for mental fatigue detection, including eye aspect ratio (EAR), eye distance (ED), eyebrow distance (EBD), mouth aspect ratio (MAR), Nose to Jaw ratio (NJR), Nose to Chin ratio (NCR), Face Area (FA), Head Motion (HM), NASA-TLX score, and EEG signals. Twenty-seven thousand frames were extracted from each equipment operator's face during each experiment phase, and one value of each facial feature was calculated from each frame, culminating in a dataset of twenty-seven thousand facial features for each equipment operator during any experiment phase. After that, for descriptive representation, standard deviation (SD) and mean (M) values of facial features for each phase of the experiment were computed. To analyze the variations in facial features due to mental fatigue, we used general linear models for repeated measures. Four geometric measurements of each facial feature were added as within-subjects factors: at baseline (T1), at 20 min (T2), at 40 min (T3), and at 60 min (T4). Using partial eta-squared (η^2), we calculated the amount of the effect on the mean values of each characteristic and the ground truth. Within-subject repeated measures analysis of variances (ANOVAs) was used for data analysis. Consequently, the F distributions with degree of freedom was reported in the results. Furthermore, Benjamini-Hochberg was also applied for multi-comparison corrections (Izmirlian, 2020) with a 5% false discovery rate (FDR) or $q = 0.05$. Benjamini-Hochberg procedure is the most widely used statistical tool that increases the statistical power and decreases the false discovery rate (Palejev and Savov, 2021). Pearson correlational coefficients were used to assess the associations between the mean changes in geometric measurements of facial features throughout the course of the experiment and the NASA-TLX scores to validate the proposed method. Furthermore, to develop ecological validity for construction equipment operators, Pearson correlation coefficients were computed between mean values of geometric measurements of facial features and EEG metric $[(\theta + \alpha) / (\alpha + \beta)]$. Because Tyas et al. (2020) reported that such an EEG metric is the most used for computation of mental fatigue.

3.4. Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data[5].

Figure 3.7 shows an overview of the research process. It demonstrates the proposed method for detecting mental fatigue in construction equipment operators by analysing brain activity patterns

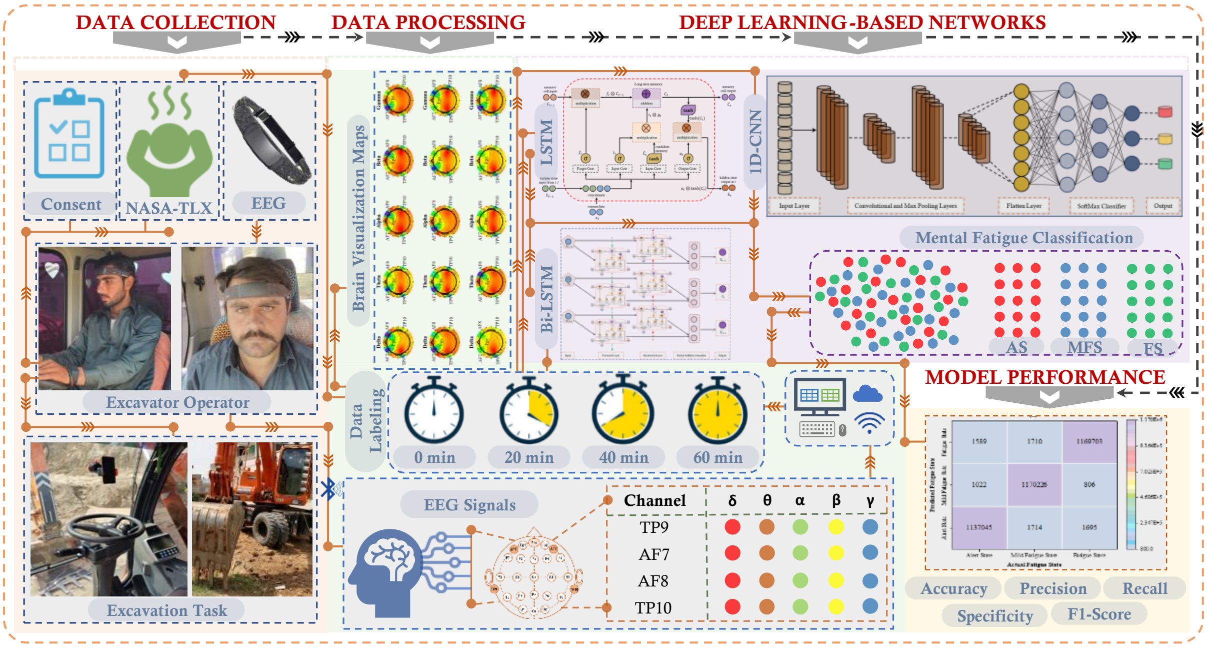


Figure 3.7: Overview of the research methodology

acquired using an EEG device. The research process consists of four steps. In the first step, an experiment was conducted to acquire relevant data. A headband was mounted on the head of construction equipment operators to obtain EEG data, and data related to subjective feelings of mental fatigue was gathered using a questionnaire. In the second step, the EEG data was analysed and labelled into mental fatigue levels using subjective scores, artifacts were removed, and the data was down sampled. In the third step, detection of multiple mental fatigue levels in construction equipment operators was done based on deep learning techniques. Each deep learning model was trained using raw EEG data from an EEG device as input data. In the last step, the performance of each deep learning architecture was assessed using metrics.

[5] The methodology presented in section 3.4 is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". *Advanced Engineering Informatics*, Volume 56, 101978

3.4.1. Experiment design and data collection.

3.4.1.1. Subjects

Fifteen male construction equipment operators were voluntarily recruited to participate in the experiments. The operators' mean age was 33.07 years (SD = 3.95). Construction equipment operators were recruited and participated in this study because excavation operation tasks (ground excavation and moving the material from the pits to the transport vehicles) are repetitive, cognitively demanding, and often involve prolonged working hours, which require the operators to have a significant level of sustained attention (Li et al., 2020d). Furthermore, we determined the sample size of excavator operators to recruit for our research investigations based on sample sizes from previous studies. In earlier studies with similar purposes, 12 crane operators (Das et al., 2020), 12 excavator operators (Li et al., 2019b), 11 drivers (Ahn et al., 2016), 6 excavator operators (Li et al., 2020d), and 5 crane operators (Liu et al., 2021a) were recruited. Considering previous research in the literature, we decided that more than fifteen operators would be sufficient for our investigation and to justify our results. All the participants were excavator operators with prior experience in excavator operations at construction sites. All the excavator operators had slept at least eight hours the previous night and abstained from alcoholic drinks for at least 24 hours before experimentation. The operators were required to directly come for experiments on their designated days, and they were not involved in any other tasks or activities before the start of the experiment. In addition, we ensured that each operator remained fully engaged during the length of the task. The experimental protocol for data collection was reviewed and approved by the ethics subcommittee of the Hong Kong Polytechnic University (Reference Number: HSEARS20210927008) and conducted in accordance with the Declaration of Helsinki. In addition, written consent was obtained from each participant after a verbal explanation of the experimental procedures. Table 3.5 provides the demographic information of the construction equipment operators who participated in the study.

Table 3.5: Construction equipment operators' demographic information.

	Mean	**SD**	**Range (Min-Max)**
Age (Years)	33.07	3.95	13 (26-39)
Job Experience (Years)	7.27	2.58	8 (3-11)
Height (cm)	175.87	5.32	18 (166-184)
Weight (kg)	77.86	7.72	22 (68-90)

Body Mass Index (kg/m²)	25.16	2.06	7.48 (21.91-29.39)

3.4.1.2. Subjective Assessment of mental fatigue

The NASA-TLX score was used for the labeling of construction equipment operators by assessing their individual subjective feelings of mental fatigue. The subjective assessment was used as a ground truth for construction equipment operators' mental fatigue levels. It has been widely used in various research investigations since its development, and its reliability and sensitivity have been tested in a consistent number of independent tests. The NASA-TLX is intended to measure operators' perceived workload in six dimensions: mental demand, physical demand, effort, own performance, temporal demand, and frustration. An overall NASA-TLX score was computed by adding the scores from each of the six dimensions of the scale. Overall NASA-TLX scores were used, with no weight applied to the individual categories. Adding the subscale scores to calculate an overall score is a common approach to simplifying the original scale (Hart, 2006a, Byers, 1989). Additionally, several recent studies by Kaduk et al. (2021), Mehmood et al. (2022), Bitkina et al. (2021), Das et al. (2020), Li et al. (2019b), and Chen et al. (2017a) reported that an increase in the NASA-TLX score over time during the same task can serve as a reliable indicator of mental fatigue. Moreover, in the construction industry, studies by Mehmood et al. (2022), Li et al. (2020d), and Li et al. (2019b) have employed the increase in the overall NASA-TLX score for the same task as a subjective indicator of mental fatigue. Likewise, in our study, an increase in the NASA-TLX score was the result of an increase in mental fatigue.

3.4.1.3. EEG Recording

To capture EEG signals, we employed the Muse headband, a flexible and user-friendly EEG recording device. Dry electrodes are located at AF7, AF8, TP9, and TP10 on a four-channel headband, with the FPz serving as the reference electrode. Electrodes are typically made of silver. The Muse headband has a sampling rate of 256 Hz, which makes it suitable for capturing EEG data. Through a Bluetooth connection, data was transmitted from the Muse headband to a smartphone. The construction equipment operators' EEG data was gathered on a smartphone using an app called Mind Monitor, then transferred to a computer for post-processing. The recorded EEG signals are subjected to artifact removal techniques to remove muscular artifacts, power line noise, and other artifacts. The Muse EEG headband has an on-board noise cancellation mechanism to filter out the noise based on the statistical properties of the data. The statistical properties used by the MUSE headband include amplitude, variance, and kurtosis. An EEG signal is considered clean if its statistical properties are below a predetermined threshold; otherwise, the signal is considered noisy and discarded (Arsalan et al., 2019). Although the on-board noise cancellation method has been successful in various fields, including research by Cannard et al. (2021) and Arsalan et al. (2019), construction site tasks are demanding and dynamic (Xing et al., 2020b). It involves the continual body movement of workers to perform these tasks on construction sites (Mehmood et al., 2022). Hence, it is crucial to remove artifacts that cause noise in the acquired EEG data. Therefore, the acquired data underwent further pre-processing techniques for artifact removal, including the third-order one-dimensional median filter (Krauss et al., 1994) and the Savitzky-Golay (SG) filter (Orfanidis, 1995). The classical SG filter is designed based on the least-squares polynomial approximation phenomenon (Savitzky and Golay, 1964) and is used to remove inappropriate and large spikes in the EEG sensor data. The goal was to smooth the data while retaining the quality of the signal. To achieve this, we applied an overlapping window of 50% (Krauss et al., 1994). Previous studies in the construction industry by Aryal et al. (2017) have effectively used this noise cancellation method to smooth the data while retaining the quality of the acquired EEG data. Once the artifacts were removed from the data, it was down sampled to 128 Hz by selecting each second sample and effectively reducing the number of data points by half. It is a common method to reduce the dimensionality of the data (Frydenlund and Rudzicz, 2015). According to Roy et al. (2019), 72% of the studies employing EEG sensors have used the down sampling technique to preprocess their EEG data. In our investigation, doing down sampling did not affect the data model's predictive power, yet it improved the training time of the models significantly. Figure 3.8 demonstrates the electrode positioning system on the scalp of construction equipment operators as well as the various views of the EEG device used in the study.

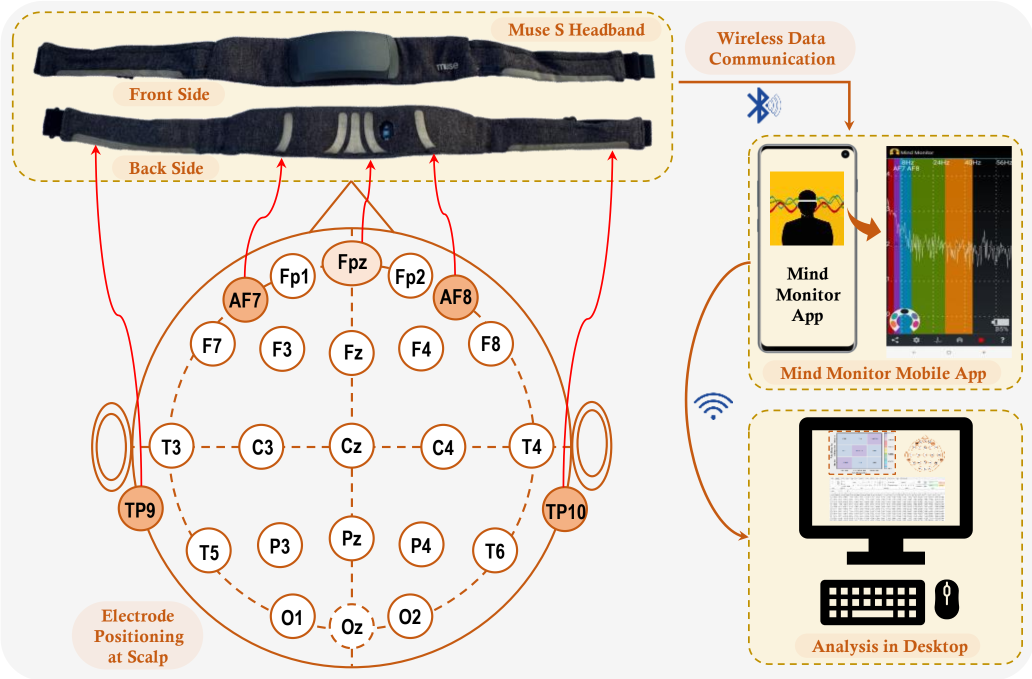


Figure 3.8: Overview of headband-based EEG device, 10-20 system of electrode positioning,

3.4.1.4. *Experiment Procedure*

Figure 3.9 shows an overview of the experiment's procedure. At a construction site, an excavator operation experiment was carried out to collect data for detecting the mental fatigue of construction equipment operators. The experiment was carried out on different days at the same time, from 9:00 a.m. to 11:00 a.m. (Li et al., 2019b) in the morning, under similar weather conditions, particularly clear weather on all data collection days. The experiment involved a repetitive and time-consuming excavation and discharge task on a construction site. It was a time-on-task approach, which is a common approach to induce mental fatigue (Li et al., 2020d, Morales et al., 2017, Hopstaken et al., 2016). For an hour, the excavator operators were required to conduct a repetitive and protracted excavation operation that included ground excavation and transporting material from pits to transport vehicles. The conditions for each excavator operator were the same, requiring them to continuously operate the equipment in the manner of a cyclic operation. The amount of earth excavated or moved, as well as the number of vehicles filled, were not fixed since it was a time-on-task experiment. Furthermore, no prior practice session was scheduled for the operators, as they already had experience with excavation operations. During the experiment, the operators were wearing a headband-based EEG device to collect data on their brain activity regarding active brain areas for mental fatigue while doing their tasks. Furthermore, the NASA-TLX score was used to quantify the subjective evaluation of equipment operators' mental fatigue. It has been used in various previous studies to subjectively assess mental fatigue in operators (Das et al., 2020). For the one-hour experiment, the subjective mental fatigue levels were recorded every 20 minutes, i.e., at 20, 40, and 60 min. Accordingly, the acquired EEG data was then labelled as per the subjective assessment into three mental fatigue states, i.e., alert state, mild fatigue state, and fatigue state (Prabaswari et al., 2019, Grier, 2015). There was no practice session included in the experiment because all the operators were professional excavator operators with prior experience in excavation operations. Furthermore, the exact duration of the experiment was not revealed to the operators. The purpose was to avoid the end-spurt effect reactivation that occurs when participants realize the experiment is nearing its conclusion (Morales et al., 2017).

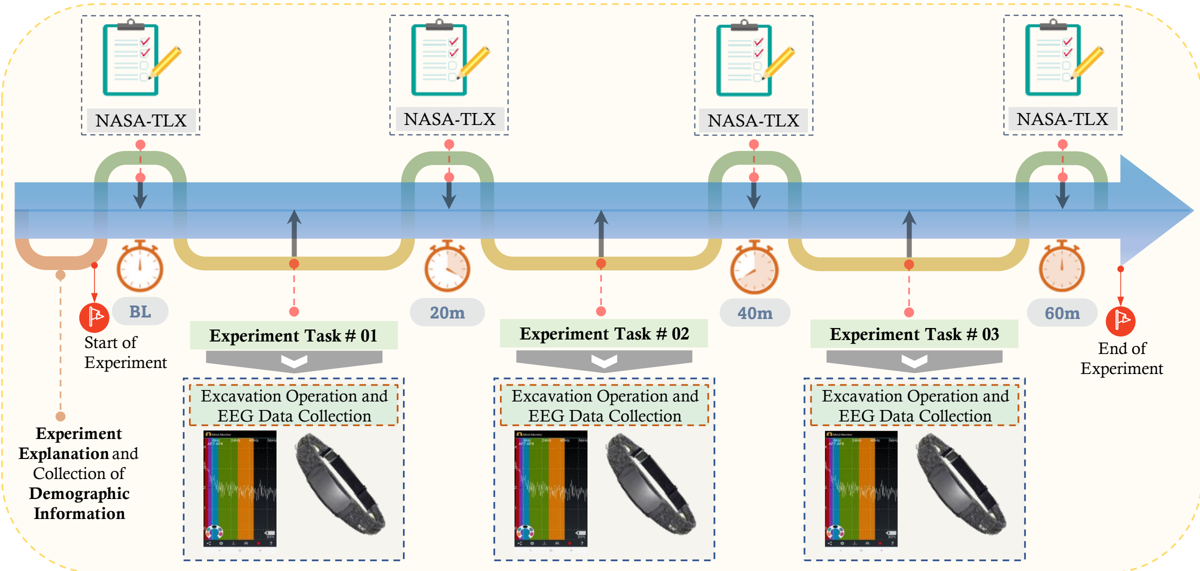


Figure 3.9: Experimental procedure for temporal assessment through NASA-TLX score and electroencephalography.

3.4.2. Deep Learning-based Networks

The aim of our research was not to develop new and unique models, but rather to evaluate the innovative approach of utilizing deep learning techniques and headband-based wearable EEG sensor data to identify and classify mental fatigue states in construction equipment operators. Hence, the primary purpose was to contribute to the advancement of knowledge in the construction field by providing a deeper understanding of the cognitive processes and mental states of construction workers through a more sophisticated analysis of EEG data. This, in turn, could pave the way for improving safety and productivity, reducing accidents and injuries, and enhancing the overall well-being of construction workers. To achieve this, we employed three types of deep learning models: long short-term memory, bidirectional long short-term memory, and one-dimensional convolutional networks to train raw EEG data acquired by a wearable sensor. The sub-sections explain the details about the structures of the deep learning architectures we adopted in our research.

3.4.2.1. Long Short-Term Memory (LSTM)

In the last decade of the twentieth century, Hochreiter and Schmidhuber (1997) presented the first examples of LSTMs. These networks have the unique ability to learn long-term dependencies. Since it also has a memory component, it is one of the finest algorithms for processing sequence data. As a result of its memory component, LSTM can recall its prior actions in a process. With just a little structural tweak, it can solve the problem of the vanishing gradient that plagues RNN. The basic layout of an LSTM cell is depicted in Figure 3.10 (Olah, 2015). Because of this cell state, LSTM can only allow specific sets of information to pass through it. To implement this function, three logic gates are used. Input to these gates is provided by the sigmoid activation function. The first gate to determine what data can be safely erased from the cell is known as the Forget Gate f_t and is described in Eq. 1:

$$f_t = \sigma\left(x_t W^f + h_{t-1} U^f + b_f\right) \qquad Eq.1$$

The result is either 0 or 1, with 0 indicating forget and 1 indicating keep. The second phase is the input gate, which determines which data will be added to the cell state or saved. As indicated in Eq. 2, the input gate also includes a second sigmoid layer for determining fresh candidate inputs that may be utilized to modify the cell's status.

$$i_t = \sigma\left(x_t W^i + h_{t-1} U^i + b_i\right) \qquad Eq.2$$

In the following phase of LSTM, the old cell is replaced with a new one. As demonstrated in Eq. 3, the tanh function generates a vector of possible values that could be appended to the state.

$$\hat{C}_t = tanh(x_t W^g + h_{t-1} U^g + b_c) \qquad Eq.3$$

Then, the new cell state replaces the previous one in C_{t-1} by discarding the information created by the forget gate in Eq. 1. The current cell state, denoted by C_t in Eq. 4, has been modified.

$$C_t = \sigma(f_t \times C_{t-1} + i_t \times \hat{C}_t) \qquad Eq. 4$$

Finally, a sigmoid layer and subsequently a $tanh$ layer is employed to classify the output, as stated in Eq. 5 and 6.

$$\sigma_t = \sigma(x_t W^o + h_{t-1} U^o + b_o) \qquad Eq. 5$$

$$h_t = tanh(C_t) \times \sigma_t \qquad Eq. 6$$

where, i_t, f_t, and σ_t denotes the input gates, forget gates, and output gates, respectively. W^i, W^f, and W^o denotes the weights for the input gate, forget gate, and output gates at time step t, respectively. W^g is the weight of the candidate layer. U^i, U^f, and U^o are the weights for the input gate, forget gate, and output gates at time step $t-1$. U^g is the weight for the candidate layer. x_t is the input at the current time step t. h_t and h_{t-1} are the cell outputs at the current time step t and the previous time step $t-1$, respectively. C_t and C_{t-1} are the states of the cell at time steps t and $t-1$, respectively. b_i, b_f, and b_o

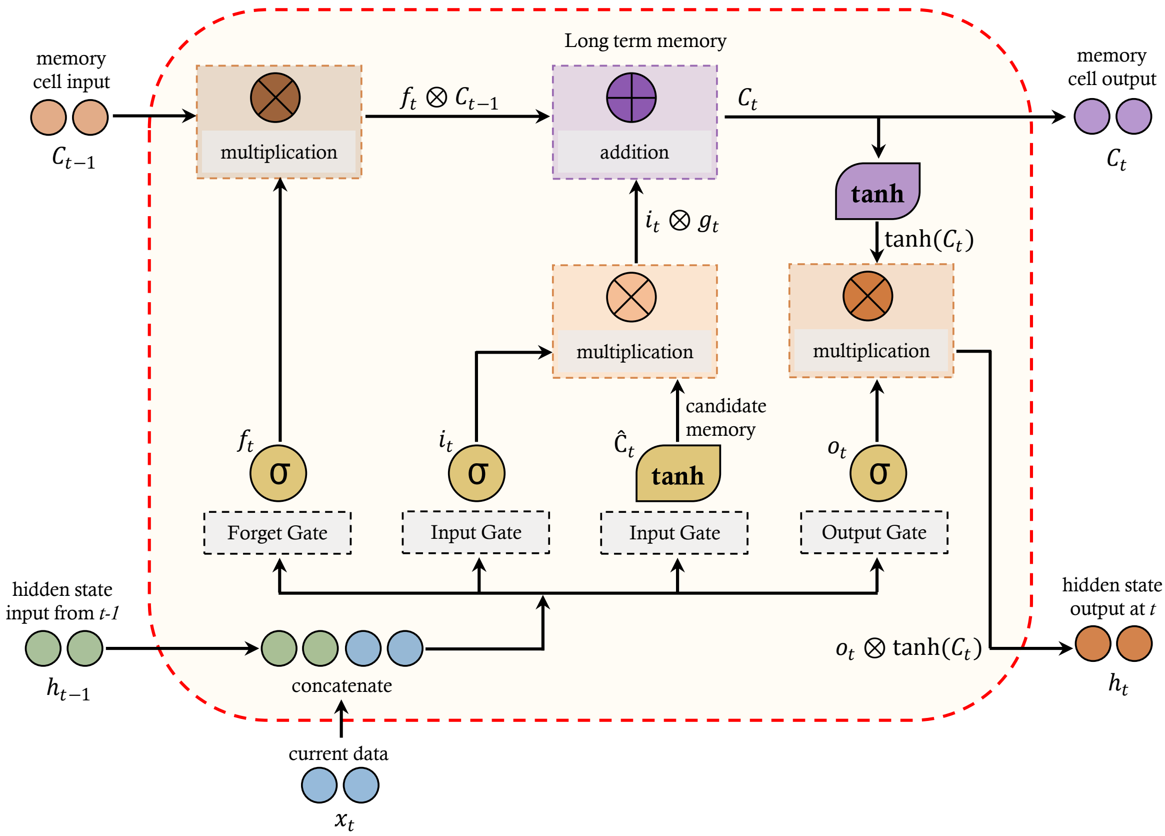


Figure 3.10: Long short-term memory (LSTM) cell architecture

denotes the biases for the input gate, forget gate, and output gates, respectively. b_c is the bias for the candidate layer, and σ is the sigmoid function.

3.4.2.2. *Bidirectional Long Short-Term Memory (Bi-LSTM)*

The Bi-LSTM layer structure is shown in Figure 3.11, and it consists of three independent layers that share the same input sequence and whose outputs are combined and displayed in the sequence. The state cells of a standard LSTM are split into a forward layer that controls the forward time path and a

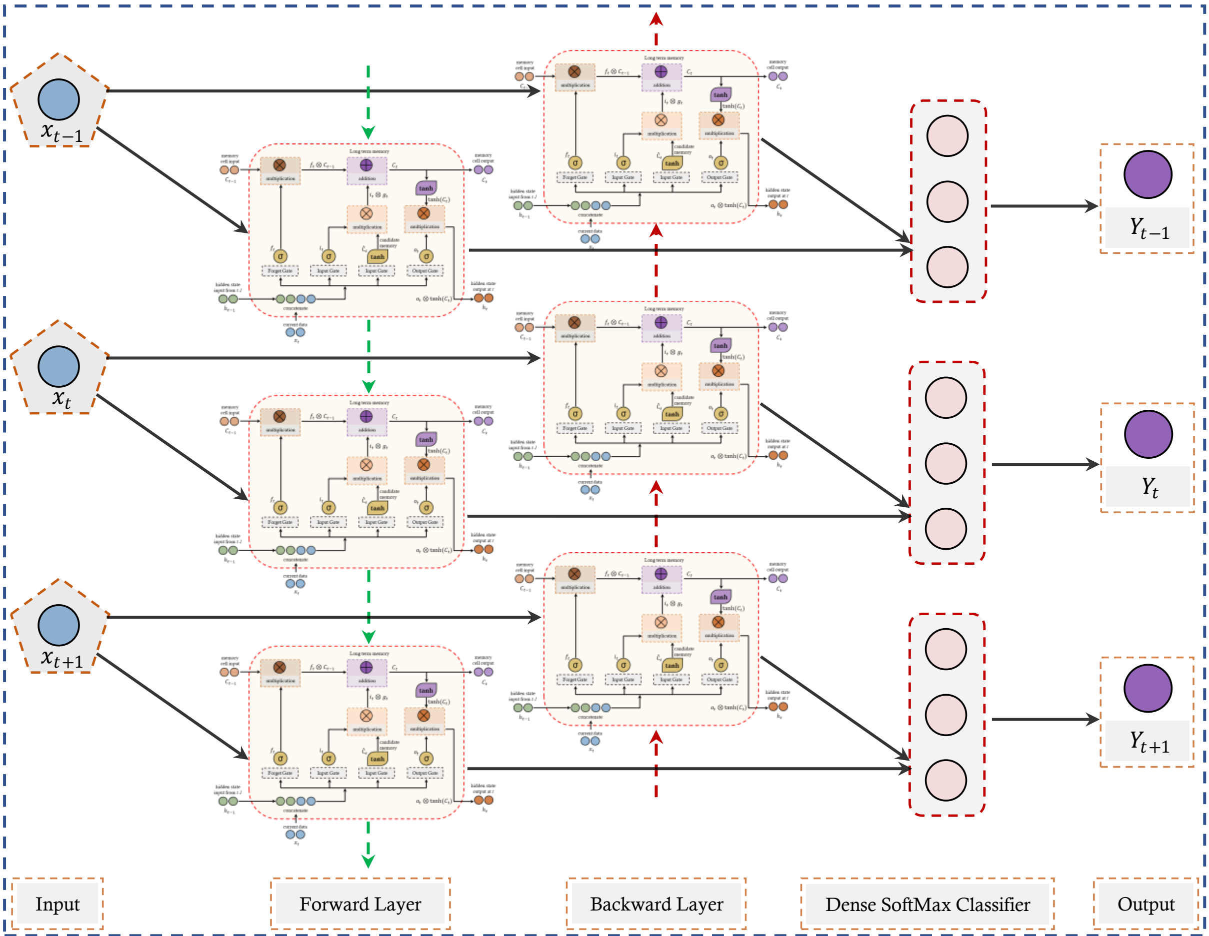


Figure 3.11: Bidirectional long short-term memory (Bi-LSTM) layer architecture

backward layer that controls the backward time direction in a Bi-LSTM model. For each time step forward and backward, information can be obtained by concatenating the outputs of the forward and backward layers. Given the established dependency between adjacent data pairs, this method improves the learning process.

3.4.2.3. 1-Dimensional Convolutional Network

Deep convolutional neural networks, as they have been traditionally described in the literature, were developed with a focus on processing only two-dimensional data, such as images and recordings (Kiranyaz et al., 2021). For this reason, 2D-CNNs have become commonplace. Recently, however, 1D convolutional neural networks (1D-CNN) have been designed to work on one-dimensional data and have been applied to a wide variety of scenarios instead of 2D-CNN, such as by Eren et al. (2019), Kiranyaz et al. (2018), and Abdeljaber et al. (2018). Typically, specialized hardware is required for training deep 2D CNNs (e.g., cloud computing or GPU farms). Conversely, training small 1D CNNs with few hidden layers is practical and can be done quickly on any CPU implementation on a desktop machine (Kiranyaz et al., 2021). As a result of their minimal processing requirements, small 1D CNNs are ideal for real-time and low-cost scenarios (Eren, 2017). The 1D-CNN structure of the time-series prediction models used in this study is depicted in Figure 3.12. The network has several layers, including input, convolution, pooling, flattening, fully connected, and output layers. The features of the input are passed into a convolution layer. A feature map is generated by filtering an input feature in the convolution layer. The outcomes are then activated using the provided function. To shrink the feature map, the convolution layer's output is fed into a pooling layer. After that, to prepare the merged feature map for further processing, it is given to a flattening layer, which transforms it into a one-dimensional array. The completely linked layer then receives input from the layered-flattened layer. The weights are

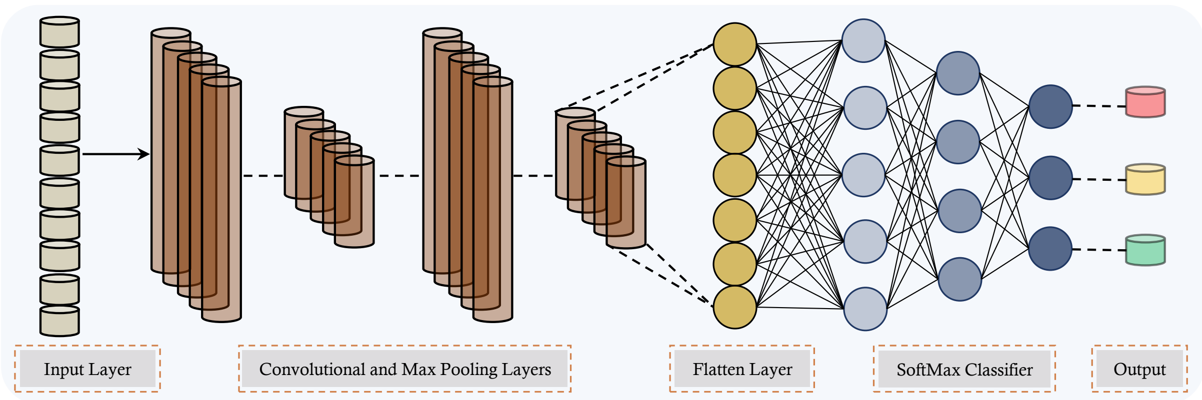


Figure 3.12: A sample one-dimensional convolutional network layer, flatten layer and SoftMax layer architecture.

used in the fully connected layer to process the data. The output layer receives the signal from the layer with all connections made. When it comes to activation functions, ReLU is used in the convolution layer of this research. All other layers are ignored by the activation function (Chaerun Nisa and Kuan, 2021).

3.4.3. Training and performance evaluation of deep learning models

The EEG data of brain activity patterns was trained using three different deep learning techniques in the present study: long short-term memory (LSTM), bidirectional long short-term memory (Bi-LSTM), and one-dimensional convolutional networks (1D-CNN). To ensure consistency among the deep learning models, they were all constructed with the same dataset for training and evaluation. Based on EEG analysis, each designated class represents a single construction equipment operator. The electroencephalography data vector for each excavator operation task done by each operator has a dimensionality of 20 vectors (5 brain waves from each electrode x 4 electrodes of the EEG device) x 256 data samples. As a result, 5120 values serve as data samples in total. Input data for the current investigation consisted of 6,971,010 sample values from each electrode for every brain wave from fifteen equipment operators, since each window size contained 256 data samples and data was gathered for one hour. A sliding window approach was utilized with a window size of nine seconds to split the EEG sensor data into smaller segments, in order to capture long-term dependencies in the data. Overlapping of consecutive windows was then employed to ensure that no relevant data was missed. Specifically, a 50% overlap of adjacent data segment lengths was used in this study, as described by Liu et al. (2021c). However, there is no consensus on the optimal percentage of overlap, as previous studies have reported a range of overlapping percentages from 1% to 95% (Roy et al., 2019). Each deep learning model consists of three layers, with the number of hidden units varying from one hundred to five hundred. A similar architecture was utilized in a previous study, also using 200 hidden units per layer. When assessing the accuracy of our models, we employed a cost function based on the cross-entropy losses (the log loss function). In a classification problem, the loss function is what ultimately decides the model's performance. It is more indicative of reality when the loss value is lower. The optimization function is responsible for making the necessary adjustments to the model's weights and biases. An adaptive form of stochastic gradient descent was utilized for model training (Kingma and Ba, 2014), in addition to the Adam optimization function. Adam is a trustworthy optimizer that provides precise and quick updates to the network's settings. This research utilized the dropout technique (Srivastava et al., 2014), a popular stochastic regularization method, to prevent model overfitting. When the loss function is extremely small on the training data and extremely large on the testing data, overfitting occurs. The primary objective of the dropout method is to inhibit neurons in the system from over-adapting to one another, which leads to poor model generalization.

Table 3.6: Dataset and hyperparameters of proposed deep learning models

Dataset and Parameters	**Value**
Number of classes	3 (Alert State, Mild Fatigue State, Fatigue State)
Number of EEG sensors	4 (TP9, AF7, AF8, TP10)
Window size	9 s
Overlap of adjacent windows	50 %
Sampling rate	128 Hz
Epoch	30
Dropout	5%
Batch size	1000
Learning rate	0.001 (Adam optimizer: provides adaptive optimization)
Number of sample data	6,971,010 data samples

During the evaluation of the model's performance, the available data was partitioned into two subsets, with 70% being allocated for training purposes and the remaining 30% for testing. The original training dataset was split into two parts, with 80% going to the training phase and 20% going to the validation phase. We used the validation data to fine-tune our hyper-parameters and find the perfect spot for each of our three deep learning models' unit counts. Analogous to earlier research using deep learning networks (Antwi-Afari et al., 2022, Yang et al., 2020, Kim and Cho, 2020), the 10-fold cross-validation method was utilized to evaluate the classification performance of deep learning models. The optimum hyper-parameters can be chosen by 10-fold cross-validation, and the deep learning models can be tested as generalized models that exhibit acceptable classification performance with an unseen dataset. Based on the model, we chose the parameter values that achieved the highest level of accuracy with the least amount of time spent in training. The findings demonstrate that by adjusting the parameters of epoch, dropout, batch size, learning rate, and hidden units to 30, 0.5, 64, 0.001, and 200, respectively, our tuning procedure yielded the best accuracy for the datasets. To run the tests and train the models, we used a computer outfitted with a 2.50 GHz Quad-Core Intel Core i7-9750H CPU, 16 GB of RAM, a 64-bit operating system (Windows 10 Pro), and an Intel Iris Plus Graphics 650, 1,536 MB GPU running

MATLAB R2020b. Table 3.6 displays the fine-tuned hyperparameters of the proposed deep learning models and the detailed dataset.

Accuracy, precision, recall, specificity, and the F1-score were employed to evaluate the three different types of deep learning models' performance in terms of evaluation and classification (Phutela et al., 2022, Attal et al., 2015). Each metric's breakdown for evaluation may be seen in Table 3.7. The most widely utilized metric to sum up classification performance across all classes is accuracy. Specifically, it is the ratio of instances that were correctly labeled relative to the total number of instances. Precision is the rate at which positive cases are correctly identified as such. In this sense, it is a quantitative indicator of precision. It is the ratio of positive instances that were correctly labeled compared to the total number of positive instances classified. Recall, also referred to as sensitivity, is a measure of how accurately positive examples were identified as such. Correctly classifying positive instances as a percentage of all positive instances is the definition of this metric. Whereas specificity is measured by how many times negative examples are correctly labeled as negative. To put it simply, it is the ratio of false-negatives that were identified compared to the total number of false-negatives. Precision and recall are combined into a single number called the F1-score, which is then used to evaluate the efficacy of the classification model without introducing any systematic bias (Ordóñez and Roggen, 2016). In addition to these measures, the confusion matrix was used to evaluate the performance of each model in particular classes, and the accuracy and loss curves were plotted to determine which model performed the best. The confusion matrix describes the discrepancies between the data's true labels and the model-generated labels. Elements on the diagonal of this matrix represent correctly classified fatigue states, whereas those off the diagonal represent incorrectly classified fatigue states. Furthermore, the Mann-Whitney test was conducted to analyze the results obtained from the deep learning models. While previous studies on EEG data and deep learning models for mental fatigue classification have compared models based on their achieved accuracy or training time, they have not statistically evaluated the difference in accuracy between models. To address this, we chose the Mann-Whitney test as it is a non-parametric test that does not require any assumptions about the distribution (Mat Roni et al., 2021), resulting in more conservative results. Velarde et al. (2022) and Phutela et al. (2022) employed analogous techniques in their investigations of the significance of predicted outcomes for time-series data using deep learning models. Table 4.17 shows the inferences from classifiers with a p-value of less than 0.01 were considered significant, while the others were considered insignificant. If the p-value was less than 0.01, it was deduced that the classifier used for analysis is significant; otherwise, it is insignificant.

Table 3.7: Performance evaluation metrics for deep learning models

Performance metric	Equation
Accuracy	$(TP+TN)/(TP+TN+FP+FN)$
Precision	$(TP)/(TP+FP)$
Recall	$(TP)/(TP+FN)$
Specificity	$(TN)/(TN+FP)$
F1-Score	$2 \text{ x } \frac{Precision \text{ x } Recall}{Precision + Recall}$

3.5. Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals[6].

Figure 3.13 presents an outline of the research process, which details a proposed approach for detecting mental fatigue in construction equipment operators through the integration of physiological and facial feature data obtained from EEG, EDA sensors, and a video camera. The research methodology comprises four distinct steps. The initial step entailed conducting an excavation operation on the construction site to gather pertinent data. This involved mounting a headband on the heads of construction equipment operators to capture EEG data, positioning an E4 watch on the wrist of operators to collect EDA data, mounting a video camera on the inside of the front screen of the excavator to capture facial feature data, and administering a questionnaire to elicit data related to subjective feelings of mental fatigue. In the second stage, the acquired data from multiple sensors was analyzed, and mental

[6] The methodology presented in section 3.4 is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". *Developments in the Built Environment,* Volume 15, 100198

fatigue levels were designated using subjective scores. The data was subjected to artifact removal, and relevant features were extracted. The third stage involved using supervised machine learning techniques to detect multiple levels of mental fatigue in construction equipment operators. Each machine learning technique was trained utilizing extracted features from multiple sensors as input data. Finally, in the last step, the performance of each supervised machine learning technique was evaluated using metrics.

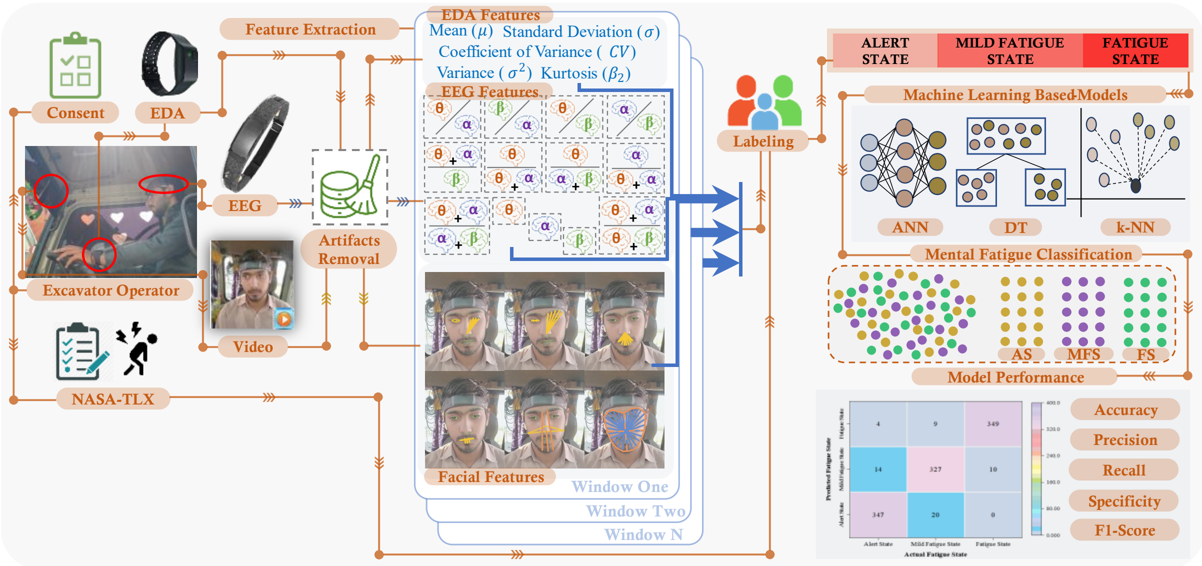


Figure 3.13: Outline of research methodology

3.5.1. Experiment procedure for collection of data.

The experiment conducted at a construction site to gather data on the mental fatigue of construction equipment operators is presented in Figure 3.14. The study was conducted at a construction site, where a time-on-task approach was employed to induce mental fatigue in the operators. Studies by (Li et al., 2020d) and (Morales et al., 2017) indicate that time-on-task is a common approach to induce mental fatigue. The experiment was conducted on multiple days, at the same time in the morning, and with consistent weather, with clear skies on each day of data collection. It involved repetitive and time-consuming excavation and discharge tasks that were carried out by excavator operators over the course of an hour. The repetitive task was an excavation operation that involved excavating the ground and transporting the excavated material from pits to vehicles. All excavator operators were subjected to the same conditions, which involved operating the equipment continuously in a cyclical manner. As this was a time-on-task experiment, the amount of earth excavated or moved, and the number of vehicles filled were not predetermined. Additionally, no practice session was arranged for the operators since they already had prior experience with excavation operations. During the experiment, the excavator operators wore an E4 watch on their wrist and a headband-based wearable EEG device to record their electrodermal activity and brain waves, respectively. Additionally, a video camera was attached to the excavator's windscreen to capture the operators' facial expressions while operating the equipment. The video footage was later converted into frames and analyzed to extract geometric measurements of facial features. To evaluate the operators' mental fatigue levels, the NASA-TLX score was used, which was recorded every 20 minutes during the one-hour experiment. The collected data was then transferred to a desktop computer, where noise removal techniques were applied to eliminate any artifacts in the data. The electrodermal activity, EEG data, and geometric measurements of facial features were labeled according to subjective measurements into three mental fatigue states: alert state, mild fatigue state, and figure state (Prabaswari et al., 2019, Grier, 2015). The duration of the experiment was not disclosed to the operators to avoid the end-spurt effect reactivation that can occur when participants realize the experiment is ending.

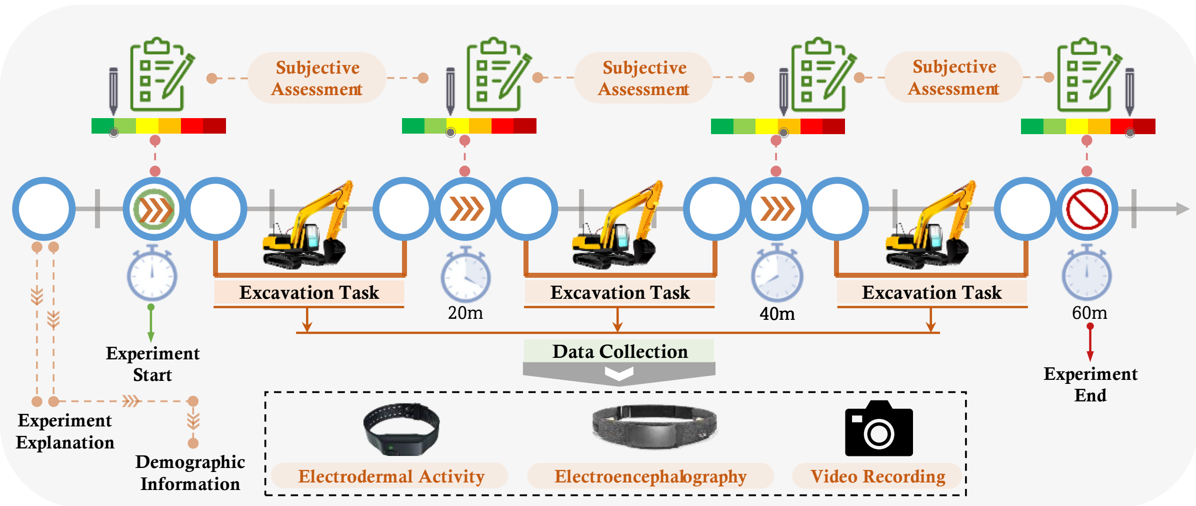


Figure 3.14: Experiment design and procedure

3.5.2. Construction operators

Sixteen male construction equipment operators were recruited voluntarily to participate in this study, with a mean age of 32.65 years (SD = 3.02). The study focused on excavator operators because excavation operation tasks, such as ground excavation and material transport, are repetitive, cognitively demanding, and often involve prolonged working hours that require operators to maintain sustained attention (Li et al., 2020d). All participants were experienced excavator operators with prior experience in excavator operations at construction sites, as demonstrated in Table 3.8. The operators were required to report directly to the experiment on their designated days and were not involved in any other tasks or activities before the start of the experiment. Furthermore, we ensured that each operator remained fully engaged during the length of the task. They had slept for at least eight hours the previous night and abstained from alcoholic drinks for at least 24 hours before the experiment. The experimental protocol was reviewed and approved by the ethics subcommittee of the Hong Kong Polytechnic University (Reference Number: HSEARS20210927008) and conducted in accordance with the Declaration of Helsinki. Additionally, written consent was obtained from each participant after a verbal explanation of the experimental procedures. Table 3.8 provides demographic information on the construction equipment operators who participated in the study.

Table 3.8: Demographic information of construction operators

	Mean (Standard Deviation)	**Range (Minimum-Maximum)**
Height (cm)	171.47 (5.32)	15 (165-180)
Age (Years)	32.65 (3.02)	15 (26-41)
Weight (kg)	76.41 (7.66)	27 (65-92)
Job Experience (Years)	6.24 (3.49)	10 (2-12)
Body Mass Index (kg/m2)	25.96 (2.05)	7.61 (21.80-29.41)

3.5.3. Apparatus and Measurement

3.5.3.1. Ground truth of mental fatigue

The NASA-TLX score was utilized to evaluate construction equipment operators' subjective feelings of mental fatigue and to provide a ground truth for their mental fatigue levels. Since its inception, the NASA-TLX has been widely utilized in numerous research studies, and its reliability and sensitivity have been established through a significant number of independent assessments. Moreover, a growing body of research has demonstrated that an increase in the NASA-TLX score during the same task over time can reliably indicate mental fatigue, as reported in studies by Kaduk et al. (2021), Bitkina et al. (2021), Das et al. (2020), Li et al. (2019b), and Chen et al. (2017a). Additionally, studies by Mehmood et al. (2022), Li et al. (2020d), and Li et al. (2019b) have also utilized the increase in the NASA-TLX score for the same task as a subjective indicator of mental fatigue in construction equipment operators. In line with these findings, our study considered an increase in the NASA-TLX score to be a reliable indicator of an increase in mental fatigue.

3.5.3.2. Electroencephalogram Recording

In this study, EEG signals were acquired using the Muse headband, a flexible and user-friendly recording system. The Muse headband has four channels with dry electrodes positioned at the AF7, AF8, TP9, and TP10 sites, while the reference electrode FPz is located at the forehead position. The electrodes are made of silver, and the sampling rate of the Muse headband for EEG signal acquisition is 256 Hz. The Muse headband was worn by all excavator operators during the excavation operation for an hour. The EEG data was transmitted in real-time from the Muse headband to a smartphone via Bluetooth, where the "Mind Monitor" app was used for recording the EEG signals. After recording, the data in the form of comma-separated value file, was transferred to a PC for further processing, as described by Mehmood et al. (2022) and Arsalan et al. (2019), and demonstrated in Figure 3.15.

3.5.3.3. Electrodermal Activity

The study utilized a photoplethysmography (PPG) wristwatch, specifically the Empatica E4, to measure electrodermal activity (EDA) in excavator operators for the purpose of assessing their mental fatigue. The Empatica E4 wristwatch includes four light-emitting diodes and four photoreceptors that automatically monitor changes in the electrical properties of the skin to derive EDA. The Empatica E4 watch was worn by all the operators during the excavation operation for an hour. EDA data was collected in real-time and transmitted from the Empatica E4 to a smartphone via Bluetooth, where the "E4 Realtime" app was utilized to record the EDA signals. The recorded data was subsequently downloaded and transferred to a PC for further processing. The EDA datasheet included a single column that indicated EDA data in MicroSiemens sampled at 4 Hz. These methods were consistent with the approaches taken by Milstein and Gordon (2020). Figure 3.15 shows an example of the Empatica E4 PPG wristwatch.

3.5.3.4. *Video Signals*

The study recorded operators' facial behavior using a color video camera placed inside the equipment cabin. The camera was positioned on the interior side and was approximately 0.6m away from the operator. The placement of the camera was carefully chosen so that it did not interfere with the operator's routine work. It was mounted on the windscreen of the equipment with no chance of visual obstruction. The color video camera had a sampling frequency of 30 frames per second, capturing 24-bit RGB with three channels or 8-bit RGB per channel. It had a high-resolution of 1440 x 1440 pixels, providing an intricate view of the operator's facial behavior for the study.

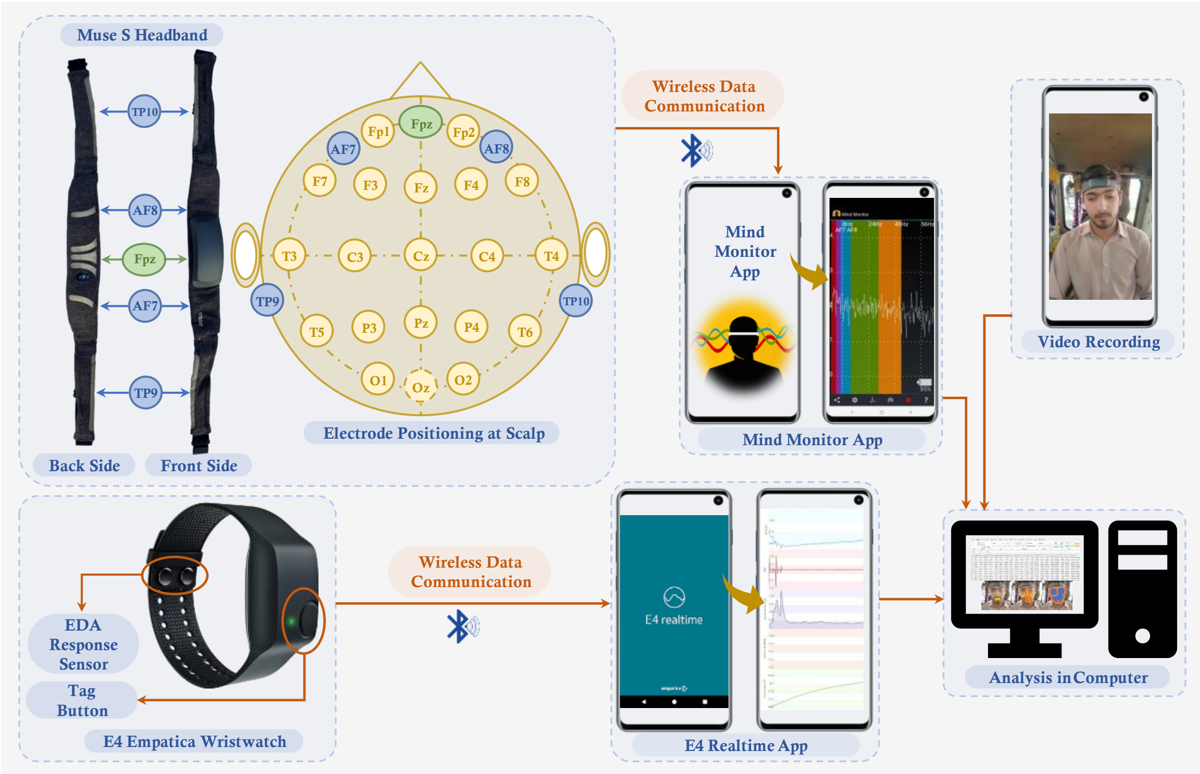


Figure 3.15: Overview of apparatus utilized to collect and transfer the acquired data.

3.5.4. Extraction of Features

3.5.4.1. *EEG features*

In the current study, ten distinct EEG metrics from each channel, including theta, alpha, and beta, were computed and analyzed to evaluate and classify mental fatigue in construction equipment operators. The investigation did not include Delta and Gamma activities since they are not expected to exhibit any activity during mental fatigue assessment. Previous research, such as Eoh et al. (2005), has reported that delta activity corresponds to a person's sleeping state. Therefore, our study concentrated on generating EEG metrics for the other three EEG bands as an indication of mental fatigue. The process involved generating band ratios from EEG channels over time, following the methodology used in Dasari et al. (2013) and Borghini et al. (2012). For example, the θ/α EEG metric was computed as the ratio of the average power spectral density value from the theta band with the average power spectral density value from the alpha band. Table 3.9 outlines all the computed EEG metrics utilized in the current study.

Table 3.9: Description of extracted EEG features

EEG Metric	Previous Research
(i) θ, (ii) α, (iii) β,	Liu et al. (2021b), Li et al. (2020a), Jap et al. (2009)
(iv) θ/α, (v) β/α, (vi) θ/β, (vii) α/β	Raufi and Longo (2022), Dissanayake et al. (2022), Stancin et al. (2021), Fan et al. (2015), Jap et al. (2009), Eoh et al. (2005)
(viii) $(\theta+\alpha)/\beta$, (ix) $\theta/(\theta+\alpha)$, (x) $\alpha/(\theta+\alpha)$, (xi) $\theta/(\alpha+\beta)$	Dissanayake et al. (2022), Wu et al. (2021a), Wang et al. (2019a), Fan et al. (2015), Eoh et al. (2005)
(xii) $(\theta+\alpha)/(\alpha+\beta)$, (xiii) $(\theta+\alpha)/(\theta+\beta)$	Mehmood et al. (2022), Stancin et al. (2021), Tyas et al. (2020),

3.5.4.2. Geometric measurement of facial features

When performing excavation operations at the construction site, all operators were video recorded for one hour on camera. OpenCV, a freely available open-source computer vision toolkit developed with Python, was initially utilized to convert each operator's video footage into frames. Subsequently, face recognition was done on each frame of the video recording using a local constrained neural field model (Baltrušaitis et al., 2016). The operator's face was detected in each frame using this model, and the results were expressed as a vector $M = [l_1, l_2, l_3, \dots \dots \dots, l_i]^F$, representing 68 landmarks identified on the operator's face in each frame via Dlib (King, 2009). In this case, l represents a detected facial landmark at position (x_i, y_i) in any frame F, F is the number of any frame, and i is the index of detected landmarks at any frame, with values ranging from one to 68. Then, Eq. 1 was used to compute the Euclidean distance between any two desirable points. Eventually, this Euclidean distance was used to compute the geometric measurements of eleven facial features investigated in this study (Mehmood et al., 2022). The proposed eleven facial features were retrieved separately from each individual frame and are described in Table 3.10 and presented in Figure 3.16.

Table 3.10: Description of extracted facial features.

Facial Feature	Description and Computation
Eye Area Average (EAA)	The average area of a closed polygon formed by joining the external landmarks on the eyes. $$EAA = \sqrt{S_a[S_a - d(l_{37}, l_{38})][S_a - d(l_{37}, l_{42})][S_a - d(l_{38}, l_{42})]} + [(l_{38}, l_{39} + l_{42}, l_{41})/2][(l_{38}, l_{42} + l_{39}, l_{41})/2] + \sqrt{S_b[S_b - d(l_{39}, l_{41})][S_b - d(l_{39}, l_{40})][S_b - d(l_{41}, l_{40})]}$$ $$\therefore S_a = {[d(l_{37}, l_{38}) + d(l_{38}, l_{42}) + d(l_{37}, l_{42})]}/{2}$$ $$\therefore S_b = {[d(l_{39}, l_{41}) + d(l_{39}, l_{40}) + d(l_{41}, l_{40})]}/{2}$$
Eye Distance Sum (SED)	The distance between the anchor and eye landmarks summed together. $$SED = \lVert l_{31} - l_{43}\rVert + \lVert l_{31} - l_{44}\rVert + \lVert l_{31} - l_{45}\rVert + \lVert l_{31} - l_{46}\rVert + \lVert l_{31} - l_{47}\rVert + \lVert l_{31} - l_{48}\rVert$$
Head Motion (HMO)	The computation of total distance between the anchor point and external landmarks of the face, per frame. $$HMO = \frac{1}{Q}\sum_{i=1}^{F}\lvert l_{F1} - l_{F2}\rvert$$
Eyebrow Sum (SEB)	The total distance between the anchor and eyebrow landmarks, computed as the sum of the Euclidean distances between corresponding points. $$SEB = \lVert l_{31} - l_{18}\rVert + \lVert l_{31} - l_{19}\rVert + \lVert l_{31} - l_{20}\rVert + \lVert l_{31} - l_{21}\rVert + \lVert l_{31} - l_{22}\rVert$$
Nose to Chin Ratio (NTC)	The distance from the anchor landmark to the chin. $$NTC = \frac{2\lVert l_9 - l_{31}\rVert}{\lVert l_8 - l_{22}\rVert - \lVert l_{10} - l_{23}\rVert}$$
Face Area (FAA)	The facial area enclosed by connecting the outermost landmarks on the face to form a closed polygon. $$FAA = \frac{1}{Q}\sum_{i=1}^{N=27}\left(S\big(S - d(l_{31}, l_{12})\big)^2\big(S - d(l_{31}, l_{13})\big)^2\big(S - d(l_{12}, l_{13})\big)^2\right),$$ $$\therefore S = \frac{d(l_{31}, l_{12}) + d(l_{31}, l_{13}) + d(l_{12}, l_{13})}{2}$$
Eye Aspect Ratio (EAR)	The ratio of the height to the width of an operators' eye. $$EAR = \frac{\lVert l_{44} - l_{48}\rVert + \lVert l_{45} - l_{47}\rVert}{2\lVert l_{43} - l_{46}\rVert}$$
Mouth Corner (MCR)	The sum of distance between the anchor and mouth corner landmarks. $$MCR = (\lVert l_{31} - l_{49}\rVert + \lVert l_{31} - l_{55}\rVert)$$
Mouth Outer (MOR)	The total distance between the anchor landmark and the external landmarks, located around the mouth. $$MOR = (\lVert l_{31} - l_{50}\rVert + \lVert l_{31} - l_{51}\rVert + \lVert l_{31} - l_{52}\rVert + \lVert l_{31} - l_{53}\rVert + \lVert l_{31} - l_{54}\rVert + \lVert l_{31} - l_{55}\rVert + \lVert l_{31} - l_{56}\rVert + \lVert l_{31} - l_{57}\rVert + \lVert l_{31} - l_{58}\rVert + \lVert l_{31} - l_{59}\rVert + \lVert l_{31} - l_{60}\rVert + \lVert l_{31} - l_{49}\rVert)$$
Mouth Aspect Ratio (MAR)	The ratio of the height to the width of an operators' mouth. $$MAR = \frac{\lVert l_{64} - l_{66}\rVert + \lVert l_{62} - l_{68}\rVert + \lVert l_{63} - l_{67}\rVert}{3\lVert l_{49} - l_{55}\rVert}$$
Nose to Jaw Ratio (NTJ)	The distance from the anchor landmark to the jaws. $$NTJ = \frac{\lVert l_3 - l_{31}\rVert}{\lVert l_3 - l_{15}\rVert}$$

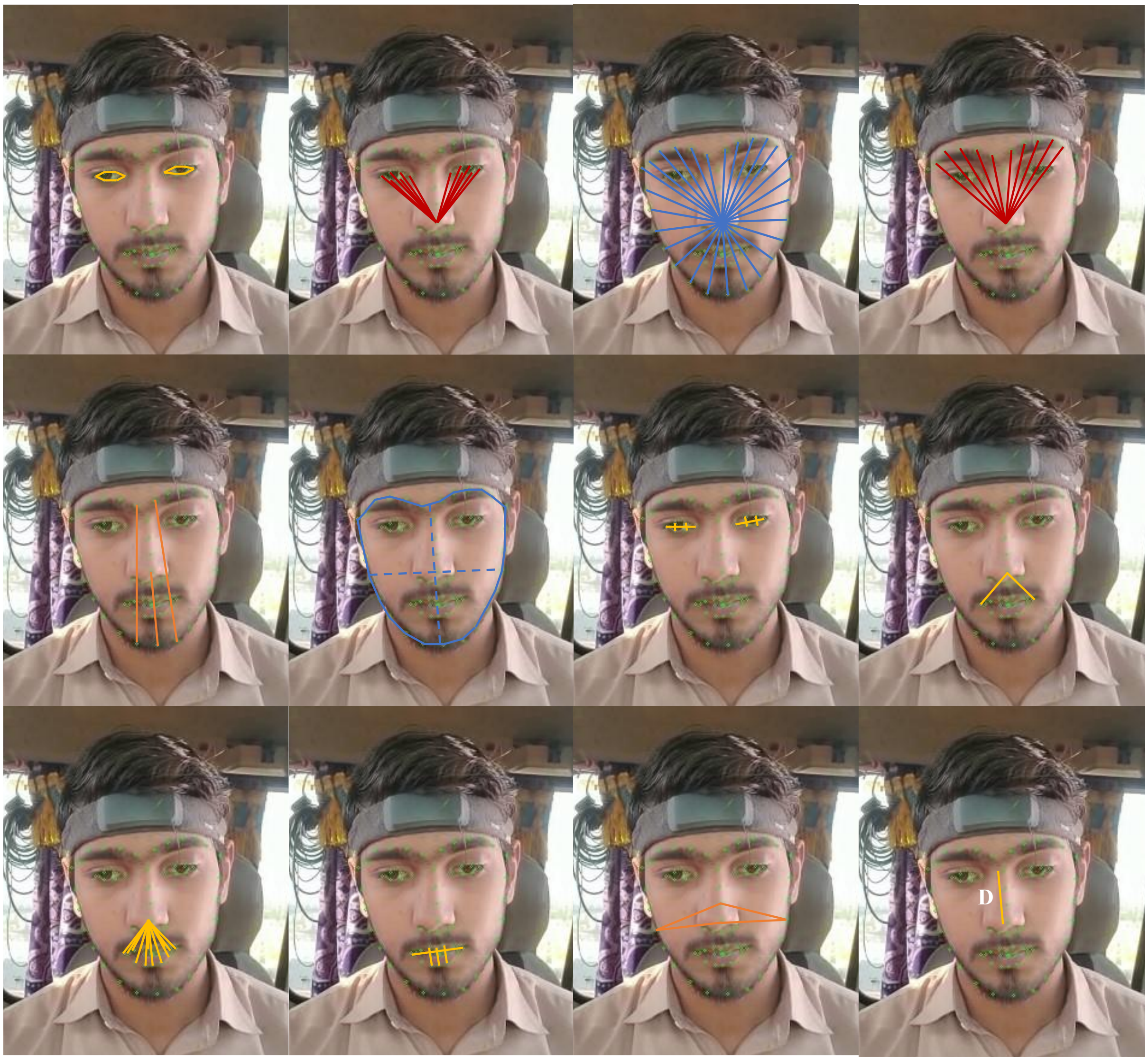


Figure 3.16: Extraction of facial features; (a) eye area, (b) eye distance, (c) head motion, (d) eyebrow, (e) nose-to-chin ratio, (f) face area, (h) eye aspect ratio, (i) mouth corner (j) mouth outer, (k) mouth aspect ratio, (l) nose-to-jaw ratio, and (m) 68 landmarks

3.5.4.3. EDA features

Initially, the EDA was separated into two components: tonic (EDL) and phasic (EDR). The former signifies the differences in sympathetic arousal among individuals, while the latter represents the dynamic component of EDA that reflects rapid changes in response to external stimuli (Greco et al., 2015, Braithwaite, 2013). In our research, we utilized the electrodermal response as a reliable indicator of mental fatigue. According to Poh et al. (2010), attention-demanding tasks can elicit an electrodermal response. Moreover, Collet et al. (2014) found that the electrodermal response is a useful tool for detecting mental fatigue. Subsequently, five distinct features were extracted from the phasic component of the electrodermal activity of each construction equipment operator. These features are mean (μ), standard deviation (σ), coefficient of variance (CV), variance (σ^2) and kurtosis (β_2). Kurtosis is a statistical measure that describes the shape, or peakedness, of a probability distribution. It is usually measured using the standardized fourth moment of a distribution, which is the fourth central moment divided by the variance of the distribution. Similarly, variance is a statistical measure used to quantify the degree of variability or dispersion in a data sample, such as the electrodermal response of operators in our case.

3.5.5. Removal of Artifacts

Artifacts, unwanted fluctuations in data owing to external sources, are present in experimental data (Sweeney et al., 2012). Due to their potential for misinterpretation and skewness in analysis, these artifacts need to be cleaned from the data (Jebelli et al., 2018b, Hwang et al., 2018). In the construction industry, excavator operators are subject to persistent and strenuous movements while conducting excavation operations. These movements are caused by the vibrations of the equipment and the operator's movements as they track the bucket to excavate and deposit the material (Mehmood et al., 2022). Unfortunately, these movements generate artifacts that must be eliminated from the collected data.

The study employed a Muse headband to acquire EEG data from construction equipment operators. This device has its own on-board noise cancellation mechanism, which is based on statistical properties of the data, such as amplitude, variance, and kurtosis, to filter out the noise. If the statistical properties of an EEG signal exceed a predetermined threshold, the signal is deemed noisy and discarded, whereas if they fall below the threshold, the signal is considered clean (Cannard et al., 2021, Arsalan et al., 2019). Considering the constant movement of operators during the excavation operations, the third-order one-dimensional median filter and the Savitzky-Golay (SG) filter (Orfanidis, 1995, Krauss et al., 1994) were further applied to the acquired EEG data for artifact removal. The principle of least squares polynomial approximation is the foundation of the SG filter, making it a good choice for data smoothing (Savitzky and Golay, 1964). In the construction industry, Mehmood et al. (2023), Mehmood et al. (2022) and Aryal et al. (2017) have used this noise cancellation method to smooth the data while preserving the quality of the EEG data.

In the current study, a freely available MATLAB-based software, Ledalab, was used to obtain cleaned, scaled, and meaningful EDA data. EDA recording is susceptible to various forms of noise, such as electrode noise and operator movement. To minimize the most common artifacts in EDA signals, a low-pass filter was applied (Taylor et al., 2015). A high-pass filter with a cut frequency of 0.5 Hz was also used to smooth the EDA signals (Braithwaite, 2013). However, large-magnitude artifacts such as excessive electrode pressure and body motion were not adequately filtered by these methods (Taylor et al., 2015). To address this, a rolling filter was applied to the EDA signals with a rolling filter of 500 data points (Posada-Quintero and Chon, 2020), and EDA was estimated every 500 ms in Micro Siemens.

The facial features data of construction equipment operators was carefully analyzed to eliminate artifacts. The process involved identifying stable facial regions during the extraction of features from every frame. Geometric measurements of facial features were then divided by the Euclidean distance of these stable regions to remove any artifacts. Previous study in construction industry by Mehmood et al. (2022) revealed that the length of the nose line, formed by connecting nose landmarks represented by vector $D = [||l_{32} - l_{28}||]^F$, was effective in eliminating artifacts, as demonstrated in Figure. Specifically, the landmarks indicated by vector D were used to calculate the Euclidean distance of the nose line, as stated by equation $d(l_{32}, l_{28}) = \sqrt{(x_{32} - x_{28})^2 + (y_{32} - y_{28})^2}$. After that, all facial features were normalized by dividing them by D, resulting in normalized facial features from each frame.

3.5.6. Machine learning models

In our study, multiple sensor data points were integrated to classify mental fatigue in construction equipment operators using machine learning. The study utilized three types of input data: EEG, EDA, and geometric measurements of facial features. A wearable Muse headband with 256 Hz per second provided the EEG data, and a wearable E4 watch with 4 Hz per second acquired the EDA data. Similarly, geometric measurements of facial features were extracted from video recordings of equipment operators with a frequency of 30 frames per second. A sliding window approach was utilized with a window size segmentation of 16 s to split the multimodal data, while overlapping of consecutive windows was then employed to ensure that no relevant data was missed. A 50% overlap of adjacent data segment lengths was used in our study (Liu et al., 2021c). As a result, a dataset of 3,600 samples for 16 construction equipment operators was generated. Also, this dataset was split into two parts, with 70% (2520 samples) designated for training and 30% (1080 samples) designated for testing. Subsequently, to accurately classify mental fatigue using data acquired from multiple sensors, we utilized three supervised machine learning classifiers: k-nearest neighbor (KNN), decision tree (DT), and artificial neural network (ANN). While we cannot provide an in-depth introduction to these algorithms due to the length of this paper, the relevant machine learning literature can be consulted for more information (Umer et al., 2020, Aryal et al., 2017, Murphy, 2012, Witten and Frank, 2002). We chose these algorithms because prior research has demonstrated their efficacy in classifying mental fatigue. For instance, Ding et al. (2020) and Hu and Min (2018) compared various machine learning classifiers, including decision tree, k-nearest neighbor, support vector machine, and artificial neural network, for detecting fatigue in drivers. Considering these studies, we trained three supervised machine learning algorithms on our specific multimodal sensor data to classify mental fatigue in construction equipment operators.

3.5.7. Training and performance evaluation of machine learning models

To evaluate the accuracy of our models, we employed k-fold cross-validation, which involves dividing the original training set into k subsets. In our case, we set k to 10 and each subset was of approximately equal size. The models were trained using k-1 subsets and validated on the remaining subset. By repeating this process for each subset, each sample was used to train and validate the models, allowing for a comprehensive assessment of the models' performance. This method ensured that the models were tested on a diverse range of data and minimized the risk of overfitting (Antwi-Afari et al., 2023, Özdemir and Barshan, 2014). To evaluate the performance of the three machine learning models, we used accuracy, precision, recall, specificity, and the F1-score (Attal et al., 2015). Table 3.11 provides a detailed breakdown of each metric. Accuracy is the most commonly used metric to assess classification performance across all classes. It is calculated as the ratio of instances that were correctly labeled to the total number of instances. Precision measures the rate at which positive cases are correctly identified, which is the ratio of positive instances that were correctly labeled to the total number of positive instances classified. Recall (sensitivity) is a measure of how accurately positive examples were identified and is defined as the percentage of all positive instances that were correctly classified.

Specificity, on the other hand, measures the rate at which negative examples are correctly identified as negative, and is calculated as the ratio of correctly identified false-negatives to the total number of false-negatives. Precision and recall are combined into the F1-score, which is used to evaluate the classification model's effectiveness without introducing any systematic bias (Antwi-Afari et al., 2023). Additionally, we plotted the confusion matrix to evaluate each model's performance in specific classes, and the accuracy and loss curves were used to determine the best-performing model. The confusion matrix displays the differences between the true labels of the data and the model-generated labels. The elements on the diagonal represent correctly classified fatigue states, while those off the diagonal represent incorrectly classified fatigue states.

Table 3.11: Performance assessment metrics for machine learning models

Performance metric	Equation
Accuracy	$\left({(TN+TP)}/{(TN+TP+FN+FP)}\right) X\ 100$
Precision	$\left({(TP)}/{(FP+TP)}\right) X\ 100$
Recall	$\left({(TP)}/{(FN+TP)}\right) X\ 100$
Specificity	$\left({(TN)}/{(FP+TN)}\right) X\ 100$
F1-Score	$\left(2 \text{ x } \frac{Recall \text{ x } Precision}{Recall + Precision}\right)$

In this study, the authors utilized the orange data mining tool, which is a Python-based open-source software (Version 3.33.0, Bioinformatics Lab, the University of Ljubljana, Slovenia) to compare and assess various classification algorithms (Demšar et al., 2013). The Orange software's canvas interface enables users to design data analysis workflows by dragging and dropping widgets, which perform various functions such as reading data, displaying tables, selecting features, training predictors, contrasting learning methods, and visualizing data items. Additionally, users can interact with the program to examine visuals and transfer them to other widgets (Kukasvadiya and Divecha, 2017, Naik and Samant, 2016).

3.6. Summary

The chapter provided the detailed methodology for each research objective. It was explained how the experiment was conducted to acquire construction equipment operators' data. The details about each data modality were described. For instance, acquiring video recordings of operators, EEG, and EDA sensor data for each construction equipment operator. Furthermore, it was explained the feature extraction from each data modality as well as the data processing techniques to clean the acquired data. Lastly, data analysis techniques for each research objective were described.

Chapter 4

Experimental Results[7]

4.1. Introduction

This chapter provides research findings for each research objective. In first research objective, it provides changes in geometric measurements of facial features because of mental fatigue. In the second objective it describes ecological validity of proposed methods for construction equipment operators. In the third objective, it describes the feasibility of deep learning techniques to classify mental fatigue using electroencephalography data. Lastly, in the third objective, this chapter provides the feasibility of multimodal data integration to classify mental fatigue in construction equipment operators.

4.2. Objective 1: To study non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features.

4.2.1. Analysis of ground truth mental fatigue data

Table 4.1 provides the descriptive statistics and statistical analysis of ground truth mental fatigue data. For NASA-TLX score the mean score for low mental fatigue (LMF) was 15.76 whereas the same for high mental fatigue (HMF) was 63.41. Similarly, EDA values for LMF were noted to be at a mean value of 0.312 and 2.217 for HMF. Statistical analysis revealed that NASA-TLX scores and EDA were statistically larger for HMF as compared to LMF with p values less than 0.05 for both with effect sizes (η^2) of 0.985 and 0.733, respectively. Besides, correlation analysis between NASA-TLX score and EDA

---

[7] This chapter is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" *Advanced Engineering Informatics*, Volume 54, 101777
Imran Mehmood, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". *Advanced Engineering Informatics*, Volume 56, 101978
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". *Developments in the Built Environment,* Volume 15, 100198
Imran Mehmood, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". *Journal of Safety Research*, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291

values for LMF resulted in a Pearson Correlation Coefficient of 0.869 as shown in Figure 4.1. Similarly, the same for HMF was found to be 0.899.

Table 4.1: Analysis of ground truth for mental fatigue

Metrics	Mean (*SD*)	Median	Mean Range (Max-Min)	*WSR*-Test (Sig.) *	η^2
NASA-TLX Score					
Low Mental Fatigue	15.76 (1.75)	16.00	6.00 (19.00-13.00)	*Z=-3.626, p<0.05*	0.985
High Mental Fatigue	63.41 (5.95)	65.00	19.00 (72.00-53.00)		
Electrodermal Activity					
Low Mental Fatigue	0.31 (0.11)	0.27	0.30 (0.48-0.18)	*Z=-3.621, p<0.05*	0.733
High Mental Fatigue	2.21 (1.19)	2.56	3.12 (3.64-0.52)		

**WSR indicates Wilcoxon signed-rank test for median, η^2 indicates eta partial squared*

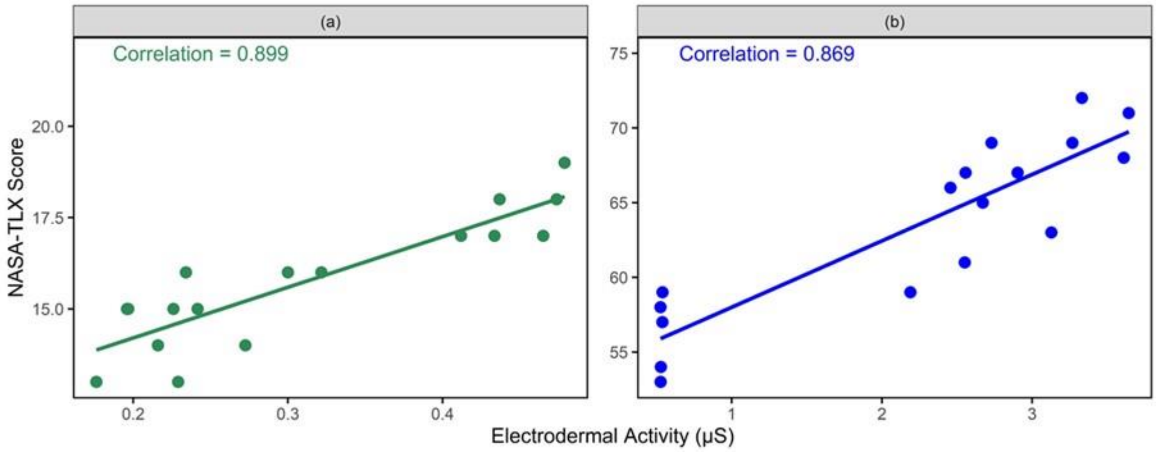


Figure 4.1: Correlation between the ground truth (a) LMF and (b) HMF

4.2.2. Mental fatigue-related facial features assessment

4.2.2.1. Eye-related facial features

Table 4.2 provides the descriptive statistics and statistical analysis of eye-related facial features for mental fatigue. The results indicate that for the eye area, the mean value for LMF was 0.2949 pixels^2 whereas the same for HMF was 0.4302 pixels^2. Similarly, eyebrow features for LMF were noted to be at a mean value of 6.0605 pixels and 6.3379 pixels for HMF. Comparing the mean values for both facial features, the increase from LMF to HMF for the eye area and eyebrow was 45.88% and 4.58%, respectively. Statistical analysis revealed that the median values of eye area and eyebrow activity were statistically larger for HMF as compared to LMF with p values less than 0.01 and 0.05 for both metrics, with an effect size (η^2) of 0.801 and 0.299, respectively. Besides, Figures 4.2(a) and 4.2(b) report box plots of the data statistics of LMF and HMF results for each eye-related facial feature.

Table 4.2: Analysis of eye related facial features for mental fatigue

Metrics	Mean (*SD*)	Median	Mean Range (Max-Min)	*WSR*-Test (Sig.) *	η^2
Eye Area Feature (*pixels*^{2})					
Low Mental Fatigue	0.29 (0.07)	0.29	0.34 (0.48-0.14)	$Z=-3.621, p<0.01$	0.801
High Mental Fatigue	0.43 (0.09)	0.42	0.55 (0.70-0.15)		
Eyebrow Feature (*pixels*)					
Low Mental Fatigue	6.06 (0.54)	6.01	2.30 (7.30-5.00)	$Z=-2.343, p<0.05$	0.299
High Mental Fatigue	6.34 (0.84)	6.34	2.86 (7.81-4.95)		

**WSR indicates Wilcoxon signed-rank test for median,* η^2 *indicates eta partial squared*

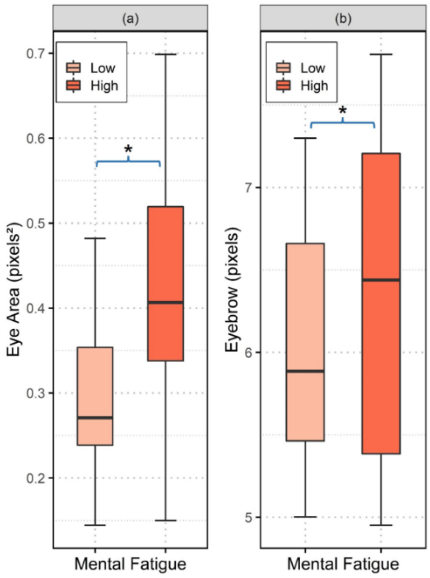


Figure 4.2: Boxplots for features (a) eye area (b) eyebrow, **indicates Wilcoxon signed rank test.*

4.2.2.2. Mouth-related facial features.

Statistical analysis and descriptive statistics on mouth-related face features for mental exhaustion are provided in Table 4.3. According to the findings, for mouth outer facial features, LMF's mean value was 3.4250 pixels while HMF's mean value was 3.8770 pixels. Similarly, for LMF, the mouth corner facial feature had a mean value of 1.4355 pixels and for HMF, the mean value was 1.6642 pixels. Comparing the means of both face features, the increase from LMF to HMF was 13.20 percent for the mouth outer and 15.93 percent for the mouth corner, respectively. Statistical analysis found that the medians for mouth outer and mouth corner were statistically greater for HMF than for LMF, with p values less than 0.01 for both facial features. The overall effect sizes (η^2) of 0.734 and 0.752 was observed between low and high mental fatigue for both the features. Besides, box plots of the data statistics for LMF and HMF results for each mouth-related face feature are shown in Figures 4.3(a) and 4.3(b).

Table 4.3: Analysis of mouth related facial features for mental fatigue.

Metrics	Mean (*SD*)	Median	Mean Range (Max-Min)	*WSR*-Test (Sig.) *	η^2

Mouth Outer Feature (*pixels*)					
Low Mental Fatigue	3.43 (0.41)	3.40	1.90 (4.44-2.53)	*Z=-3.574, p<0.01*	0.734
High Mental Fatigue	3.88 (0.52)	3.89	3.07 (5.37-2.31)		
Mouth Corner Feature (*pixels*)					
Low Mental Fatigue	1.44 (0.21)	1.46	1.00 (1.96-0.96)	*Z=-3.621, p<0.01*	0.752
High Mental Fatigue	1.66 (0.29)	1.68	1.56 (2.43-0.87)		

**WSR indicates Wilcoxon signed-rank test for median, η^2 indicates eta partial squared*

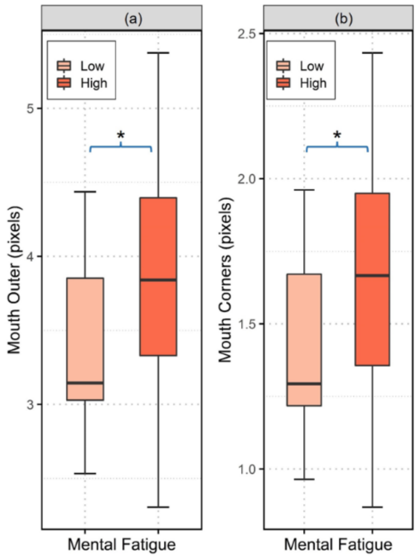


Figure 4.3: Boxplots for features (a) mouth outer (b) mouth corners, **indicates Wilcoxon signed-rank test.*

4.2.2.3. Head-related facial features.

Descriptive statistics and statistical analysis of head-related facial features for mental fatigue are described in Table 4.4. The results indicate that for the face area, the mean value for LMF and HMF was 9.2141 pixels^2 and 11.6928 pixels^2, respectively. Similarly, the mean value of head motion for LMF was 5.8202 pixels per frame whereas the same HMF was 6.1813 pixels per frame. Comparing the means values for both the facial features, the increase from LMF to HMF for the face area and the head motion was 26.9% and 6.20%, respectively. Statistical analysis revealed that the median values of the face area and head motion were statistically larger for HMF as compared to LMF with p values less than 0.01 for both with effect sizes (η^2) of 0.726 and 0.682, respectively. Additionally, Figures 4.4(a) and 4.4(b) illustrate box plots of the LMF and HMF results for each head-related face feature.

Table 4.4: Analysis of head-related facial features for mental fatigue

Metrics	**Mean (*SD*)**	**Median**	**Mean Range (Max-Min)**	***WSR*-Test (Sig.) ***	η^2
Face Area Feature (*pixels*^{2})					
Low Mental Fatigue	9.21 (0.81)	9.17	5.86 (11.27-7.38)	*Z=-3.621, p<0.01*	0.726
High Mental Fatigue	11.70 (1.85)	11.56	10.49 (16.90-6.41)		

Head Motion Feature (*pixels per frame*)					
Low Mental Fatigue	5.82 (0.18)	5.82	1.21 (6.44-5.23)	*Z=-3.479, p<0.01*	0.682
High Mental Fatigue	6.18 (0.32)	6.17	1.73 (7.04-5.31)		

**WSR indicates Wilcoxon signed-rank test for median, η^2 indicates eta partial squared*

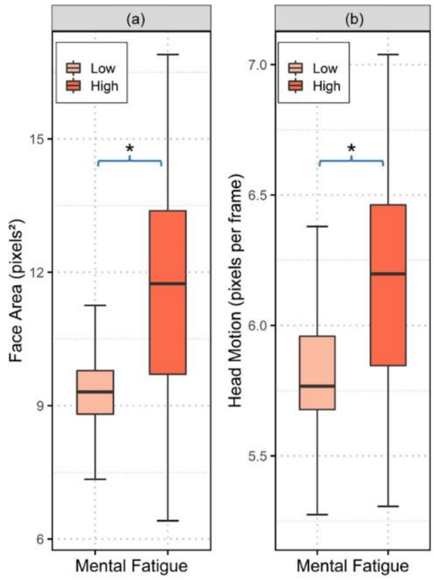


Figure 4.4: Boxplots for features (a) face area (b) head motion, **indicates Wilcoxon signed-rank test*

4.2.3. Correlations between facial features' geometric measurements and subjective mental fatigue scores

Table 4.5 demonstrates the correlations between geometric measurements of facial features and subjective mental fatigue levels. The mean eye area was substantially associated with subjective mental fatigue levels in the LMF (r = 0.6482) and HMF (r = 0.5352) groups. Similarly, geometric measurements of brow distance and facial characteristics were found to be strongly related to subjective mental fatigue levels in both mental fatigue groups, LMF (r = 0.5801) and HMF (r = 0.5941). Furthermore, the mean values of mouth outer and mouth corner face features in both mental fatigue groups were not significantly related to the corresponding subjective scores. Likewise, the geometric measurements and subjective scores for the mean values of face area indicated a significant association in LMF (r = 0.7822) and HMF (r = 0.6829). Finally, with r = 0.6928 and r = 0.7654, subjective scores were determined to have significant correlations to head motion face features in LMF and HMF, respectively.

Table 4.5: Correlations between facial features' geometric measurements and mental fatigue

Facial Feature	NASA-TLX Score		
	Mental Fatigue Labeling	**Low Mental Fatigue**	**High Mental Fatigue**
Eye Area (Pixels^2)	Low Mental Fatigue	0.6482**	
	High Mental Fatigue		0.5352*

Eyebrow Distance (Pixels)	Low Mental Fatigue	0.5807*	
	High Mental Fatigue		0.5941*
Mouth Outer (Pixels)	Low Mental Fatigue	-0.2566	
	High Mental Fatigue		-0.0018
Mouth Corners (Pixels)	Low Mental Fatigue	-0.2952	
	High Mental Fatigue		0.0532
Face Area (Pixels^2)	Low Mental Fatigue	0.7822**	
	High Mental Fatigue		0.6829**
Head Motion (Pixels/Frame)	Low Mental Fatigue	0.6928**	
	High Mental Fatigue		0.7654**

**Correlation is significant at 0.05; **Correlation is significant at 0.01*

4.3. Objective 2: To investigate the validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators.

In the study, all 16 construction equipment operators successfully completed the experiment. Therefore, data from all operators was used for analysis.

4.3.1. Analysis of ground truth data

The NASA-TLX score was used as a ground truth for mental fatigue detection. Statistical analysis and descriptive statistics of the ground truth assessment are shown in Table 4.6. The NASA-TLX demonstrated a substantial rise in subjective mental fatigue, from 11.25 (SD = 2.77) at baseline (T-1) to 65.25 (SD = 4.85) at the end of the last experiment phase (T-4). Table 4.6 shows that as the experiment progressed, operators reported increasing levels of mental fatigue.

Table 4.6: Means and standard deviations of mental fatigue metrics in different time phases.

Metrics	**Time**			
	Baseline (T1)	**20 mins (T2)**	**40 mins (T3)**	**60 mins (T4)**
Subjective Assessment				
NASA-TLX Score (0-100)	11.25 (2.77)	30.81 (2.99)	45.00 (4.27)	65.25 (4.85)
Facial Features				
Eye Aspect Ratio	0.517 (0.116)	0.465 (0.086)	0.380 (0.103)	0.306 (0.024)
Eye Distance (pixels)	2.251 (0.523)	2.317 (0.532)	2.613 (0.783)	3.114 (0.681)
Eyebrow (pixels)	5.976 (0.582)	6.071 (0.595)	6.276 (0.778)	6.448 (0.777)
Mouth Aspect Ratio	0.301 (0.013)	0.314 (0.014)	0.318 (0.013)	0.329 (0.016)
Nose to Jaw Ratio	3.272 (0.166)	3.235 (0.153)	3.249 (0.255)	3.163 (0.277)
Nose to Chin Ratio	2.119 (0.604)	2.058 (0.576)	1.897 (0.569)	1.841 (0.478)
Face Area (pixels^2)	8.653 (0.809)	9.077 (0.857)	10.461 (1.606)	11.705 (2.128)
Head Motion (pixels per frame)	5.659 (0.166)	5.807 (0.161)	6.006 (0.295)	6.149 (0.322)

4.3.2. Mental fatigue related facial metrics.

4.3.2.1. Eye aspect ratio and eye distance:

The descriptive statistics and statistical analysis of eye aspect ratio and eye distance-related facial features are provided in Table 4.6 and Figure 4.5(a) and 4.5(b). The recorded results revealed a decrease in eye aspect ratio from experiment phase T-1 (ratio = 0.517), T-2 (ratio = 0.465), T-3 (ratio = 0.380) to T4 (ratio = 0.306), whereas an increase in eye distance feature was found from experiment phase T-1 (2.251 pixels), T-2 (2.317 pixels), T-3 (2.613 pixels) to T-4 (3.114 pixels). In general, the construction equipment operators showed a significantly decreasing eye aspect ratio due to mental fatigue (GLM: F (3, 45) = 25.597, p < 0.05, partial η_p^2 = 0.631). Furthermore, significant differences in pairwise comparisons was found for eye aspect ratio, between the experiment phases i.e., T1-T2 (t_{Stat} = 4.040, p = 0.001), T2-T3 (t_{Stat} = 2.785, p = 0.014), T3-T4 (t_{Stat} = 2.917, p = 0.011), T1-T3 (t_{Stat} = 3.821, p = 0.002), T1-T4 (t_{Stat} = 8.007, p < 0.001), and T2-T3 (t_{Stat} = 8.611, p < 0.001) using Benjamini-Hochberg corrections, shown in Table 4.7. Nevertheless, the pattern was increasing (F(3, 45) = 12.919, p < 0.05, partial η_p^2 = 0.463) for eye distance feature, Likewise, using Benjamini-Hochberg multi-comparison corrections, significant differences for ED were also found in pairwise comparisons between the experiment phases, i.e., T1-T4 (t_{Stat} = -11.635, p < 0.001), and T2-T4 (t_{Stat} = -8.247, p < 0.001), shown in Table 4.7. However, through paired comparisons in the rest of the experiment phases for eye distance, it was discovered that the differences were not significant. The boxplots of the data statistics for both eye aspect ratio and eye distance are shown in Figures 4.6(a) and 4.6(b), respectively. Attributable to low R^2 values, the variations in these features are due to the mental fatigue of operators, as reflected by the regression analysis displayed in Figure 4.7 of these two facial traits with other features.

Table 4.7: Significance of facial feature with respect to various timepoints

Metrics	**ANOVA**		η^2	**Multi-Comparison Corrections using Benjamini-Hochberg**					
	F	***P***		T1 vs T2	T1 vs T3	T1 vs T4	T2 vs T3	T2 vs T4	T3 vs T4
EAR	25.597	≤ 0.05	0.631	4.040*	3.821*	8.007*	2.785*	8.611*	2.917*
ED	12.919	≤ 0.05	0.463	-0.841	-2.101	-11.635*	-1.359	-8.247*	-2.348
EB	17.636	≤ 0.05	0.540	-4.268*	-4.463*	-5.771*	-3.105*	-5.184*	-1.810
MAR	31.390	≤ 0.05	0.677	-6.584*	-9.511*	-7.026*	-2.516*	-4.524*	-2.686*
NJR	1.067	≥ 0.05	0.066	-	-	-	-	-	-
NCR	12.627	≤ 0.05	0.457	3.037*	4.836*	4.041*	3.949*	3.431*	0.957

FA	24.444	≤ 0.05	0.620	-5.238	-5.192*	-7.215	-3.911*	-5.924*	-2.208*
HM	32.546	≤ 0.05	0.685	-6.657*	-6.635*	-9.328*	-4.423*	-5.684*	-1.919

*EAR is Eye Aspect Ratio; ED is Eye Distance; EB is Eyebrow; MAR is Mouth Aspect Ratio; NJR is Nose to Jaw Ratio; NCR is Nose to Chin Ratio; FA is Face Area; HM is Head Motion; η^2 is effect size Partial eta-squared; *The t_{Stat} is significant at the $p < 0.05$*

4.3.2.2. Eyebrows

Table 4.6 and Figure 4.5(c) provide the descriptive statistics and statistical analysis of eyebrow-related facial features. This feature is a sum of the Euclidean distance between the anchor landmark on the nose and the corresponding landmarks on the eyebrows. The results indicate that the average value of the eyebrow feature increased from experiment phase T-1 (5.976 pixels), T-2 (6.071 pixels), T-3 (6.276 pixels) to T4 (6.448 pixels). There were also significant main effects of time-on-task on eyebrow features (GLM: F (3, 45) = 17.636, $p < 0.05$, partial $\eta_p^2 = 0.540$). Besides, the pairwise comparisons of eyebrow features with Benjamini-Hochberg showed significant differences for Eyebrow between the experiment phases, i.e., T1-T2 (t_{Stat} = -4.268, p = 0.001), T1-T3 (t_{Stat} = -4.463, $p < 0.001$), T1-T4 (t_{Stat} = -5.771, $p < 0.001$), T2-T3 (t_{Stat} = -3.105, p = 0.007), and T2-T4 (t_{Stat} = -5.184, $p < 0.001$), shown in Table 4.7. However, the corrections for the rest of the comparisons were not significant. Besides, figure 4.5(c) indicates that the average Euclidean distance for eyebrow characteristics rose from experiment phase T-1 at baseline to experiment phase T-4. Figure 4.6(c) depicts the boxplots of the data statistics for the eyebrow feature for all experiment phases.

4.3.2.3. Mouth Aspect Ratio

Table 4.6 and Figure 4.5(f) provide the descriptive statistics and statistical analysis of mouth aspect ratio related facial features. The results indicate that there was an increase in mouth aspect ratio from experiment phase T-1 (ratio = 0.301), T-2 (ratio = 0.314), T-3 (ratio = 0.318) to T4 (ratio = 0.329). Considerable main effects of time on task were also found on mouth aspect ratio (GLM: F (3, 45) = 31.390, $p < 0.05$, partial $\eta_p^2 = 0.677$). Subsequent pairwise comparisons with Benjamini-Hochberg corrections showed notable differences in mouth aspect ratio for each of the experiment phases i.e., T1-T2 (t_{Stat} = -6.584, $p < 0.001$), T1-T3 (t_{Stat} = -9.511, $p < 0.001$), T1-T4 (t_{Stat} = -7.026, $p < 0.001$), T2-T3 (t_{Stat} = -2.516, p = 0.024), T2-T4 (t_{Stat} = -4.524, $p < 0.001$), and T3-T4 (t_{Stat} = -2.686, p = 0.017), shown in Table 4.7. However, the rest of the pairwise comparisons were not statistically significant. The pairwise comparison also indicated that the mean value of the mouth aspect ratio at baseline was significantly shorter than at rest of the experiment phases. As shown in Figure 4.5(f), all other pairwise comparisons were not statistically significant. Moreover, figure 4.6(d) depicts boxplots of the mouth aspect ratio data statistics for each experiment phase. Attributable to low R^2 values, it can be concluded that the variation in mouth aspect ratio is due to the mental fatigue of operators, as depicted by the regression analysis displayed in Figure 4.7 of this trait with other features.

4.3.2.4. Nose to Jaw Ratio and Nose to Chin Ratio

Table 4.6, Figures 4.5(d) and 4.5(e) provide the descriptive statistics and statistical analysis of nose-to-jaw ratio and nose-to-chin ratio related facial features. The results indicate that the variation in nose-to-jaw ratio was not monotonous during the experiment phases; T-1 (ratio = 3.272), T-2 (ratio = 3.235), T-3 (ratio = 3.249) to T4 (ratio = 3.163), whereas a decrease pattern was found in the mean value of nose-to-chin ratio during the experiment phases; T-1 (ratio = 2.119), T-2 (ratio = 2.058), T-3 (ratio = 1.897) to T-4 (ratio = 1.841). Considerable main effects of time on task on the nose-to-jaw ratio (GLM: F (3, 45) = 1.067, $p > 0.05$, partial $\eta_p^2 = 0.066$) was not found. Nevertheless, the construction equipment operators showed a significantly decreasing nose to chin ratio due to mental fatigue (GLM: $F(3, 45) = 12.627$, $p < 0.05$, partial $\eta_p^2 = 0.457$) with significant differences in pairwise comparisons was found using Benjamini-Hochberg corrections, between the experiment settings i.e., T1-T2 ($t_{Stat} = 3.037$, $p = 0.008$), T1-T3 ($t_{Stat} = 4.836$, $p < 0.001$), T1-T4 ($t_{Stat} = 4.041$, $p = 0.001$), and T2-T3 ($t_{Stat} = 3.949$, $p = 0.001$), T2-T4 ($t_{Stat} = 3.431$, $p = 0.004$), shown in Table 4.7. However, the pairwise comparisons for NTC were not statistically significant between the last two experiment phases, i.e., T3 and T4. Furthermore, Figures 4.6(e) and 4.6(f) show boxplots of data statistics for nose to chin ratio and nose to jaw ratio across all experiment phases.

4.3.2.5. Face Area and Head Motion

Table 4.6, Figures 4.5(g) and 4.5(h) provides the descriptive statistics and statistical analysis of face area and head motion related facial features. The results indicate that there was an increase in the mean values of face area (FA) feature from experiment phase T-1 (8.653 pixels^2), T-2 (9.077 pixels^2), T-3 (10.461 pixels^2) to T4 (11.705 pixels^2). Besides, an increase in the mean value of head motion (HM) feature was also recorded from experiment phase T-1 (5.659 pixels/frame), T-2 (5.807 pixels/frame), T-3 (6.006 pixels/frame) to T-4 (6.149 pixels/frame). During the excavation operation, a significantly increasing pattern was found in the geometrical measurements of both the facial features i.e., face area (GLM: F (3, 45) = 24.444, $p < 0.05$, partial $\eta_p^2 = 0.620$) and head motion (GLM: F (3, 45) = 32.546, p < 0.05, partial $\eta_p^2 = 0.685$). Subsequently, pairwise comparisons with Benjamini-Hochberg corrections showed significant difference in the mean values of FA for all the experiment settings i.e., T1-T2 (t_{Stat} = -5.238, $p < 0.001$), T1-T3 (t_{Stat} = -5.192, $p < 0.001$), T1-T4 (t_{Stat} = -7.215, $p < 0.001$), T2-T3 (t_{Stat} = -3.911, $p = 0.001$), T2-T4 (t_{Stat} = -5.924, $p < 0.001$), and T3-T4 (t_{Stat} = -2.208, $p = 0.043$), shown in Table 4.7. Similarly, using Benjamini-Hochberg multi-comparison corrections, significant differences in pairwise comparisons were found for head motions between the experiment phases, i.e., T1-T2 (t_{Stat} = -6.657, $p < 0.001$), T1-T3 (t_{Stat} = -6.635, $p < 0.001$), T1-T4 (t_{Stat} = -9.328, $p < 0.001$), T2-T3 (t_{Stat} = -4.423, and $p < 0.001$), and T2-T4 (t_{Stat} = -5.684, $p < 0.001$). However, the rest of pairwise comparisons for both the facial features were not significant. The boxplots of the data statistics for face area and head motion during all phases of the experiment are shown in Figures 4.6(g) and 4.6(h). Due to low R^2 values, it can be concluded that the changes in these traits are due to mental fatigue of operators, as demonstrated by the regression analysis showed in Figure 4.7.

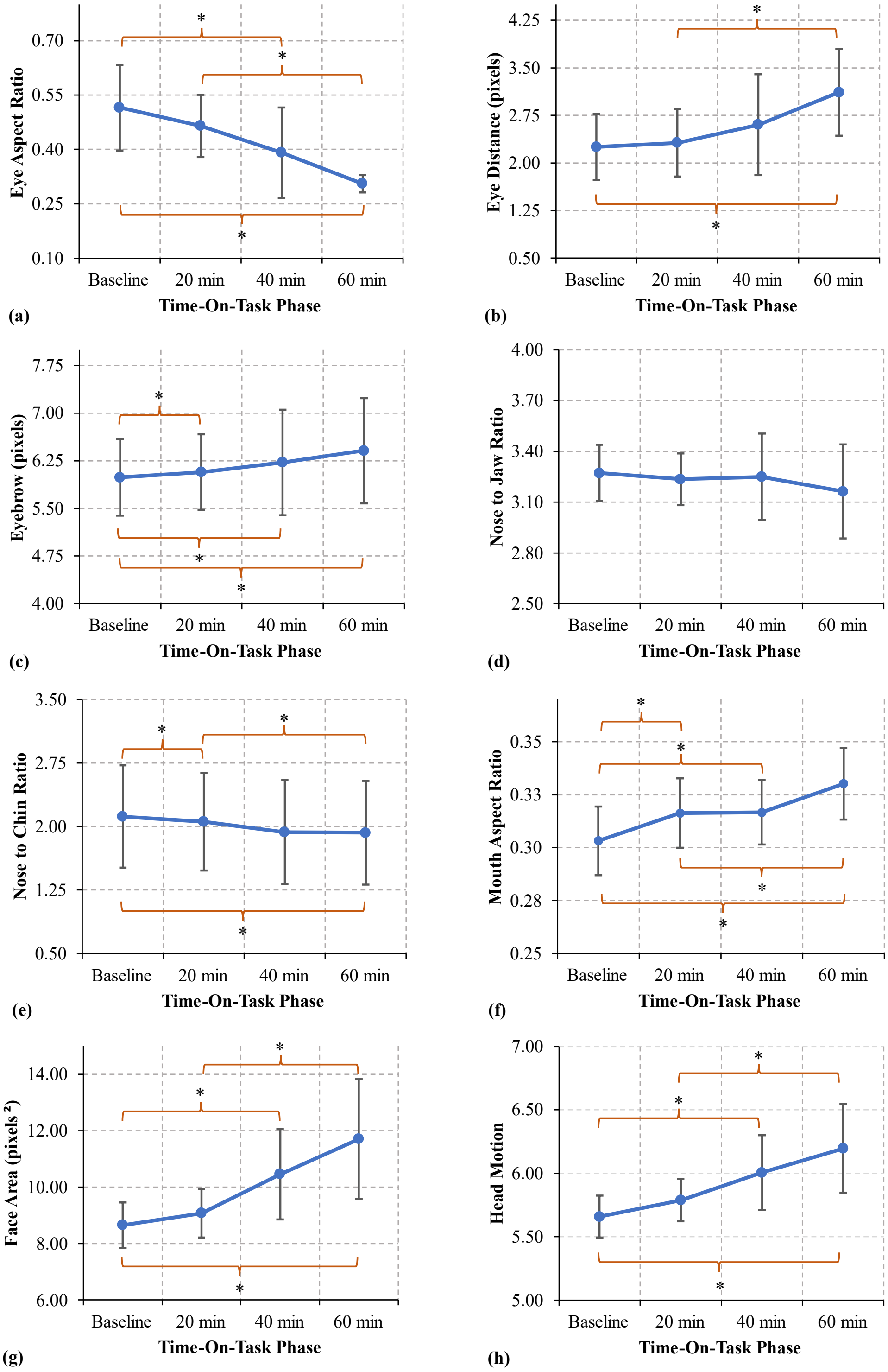


Figure 4.5: Variation in facial features due to mental fatigue with increasing Time-On-Task phases, (a) eye aspect ratio; (b) eye distance; (c) eyebrow; (d) nose to jaw ratio; (e) nose to chin ratio; (f) mouth aspect ratio; (g) face area; (h) head motion

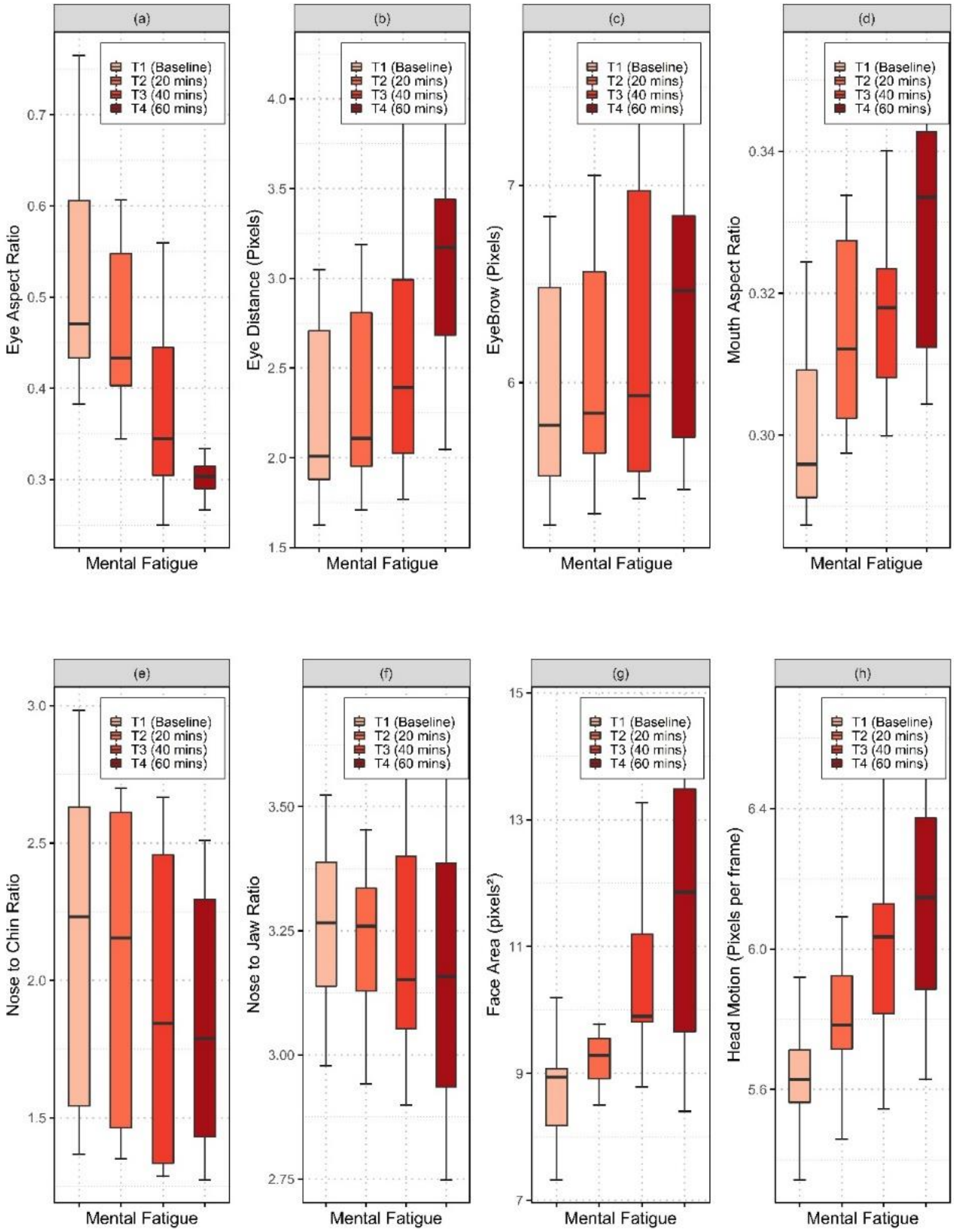


Figure 4.6: Boxplots for facial features (a) eye aspect ratio (b) eye distance (c) eyebrow (d) mouth aspect ratio (e) nose to chin ratio (f) nose to jaw ratio (g) face area and (h) head motion

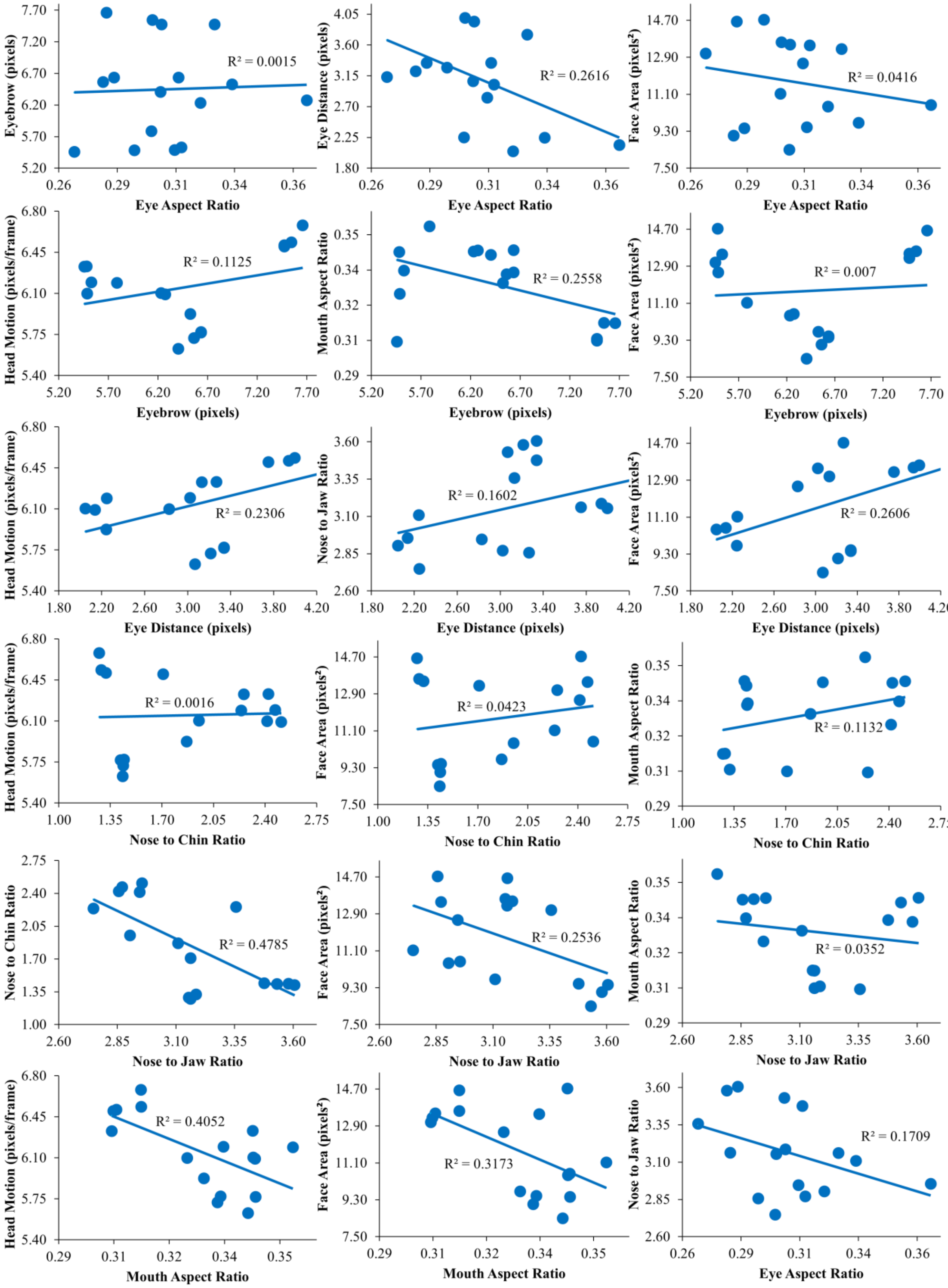


Figure 4.7: Regression statistics between individual facial features at the end of experiment

4.3.3. Analysis of physiological data

Analysis of the physiological signals EEG was performed by applying the paired t-test on the absolute power for each frequency band of the EEG signal obtained from all the channels of the MUSE headband during the four experimental phases: baseline, at 20 min, 40 min, and 60 min. A null hypothesis and p-value were used to determine the t-test decision. The difference between the groups was considered significantly different if the p-value was less than 0.05 and the null hypothesis was 1. Table 4.8 shows a statistically significant difference according to the results of *p*- value for EEG power spectral density in different brain regions. For example, the *t*-test applied to EEG signals revealed that the alpha band was found to be statistically significant at right frontal channel AF8 (between all experiment phases at baseline and 20 mins; 20 mins and 40 mins) and at left frontal channel AF7, it was statistically significant between experiment phases 20 mins and 40 mins only. Likewise, the beta band was found to be statistically significant at left frontal channel AF7 (between experiment phases at 40 mins and 60 mins only) and frontal channel AF8 between all experiment phases. The Delta and gamma bands were found to be statistically significant in the left and right temporal regions. The beta band, on the other hand, showed differences that were statistically significant in both the frontal and temporal parts of the brain. The statistical analysis for all the bands in the respective channels is demonstrated in Table 4.8. Figure 4.8 shows the brain activity visualization obtained using the power spectral density of the EEG data of the construction equipment operators during the four phases of the experiment. On the brain maps, the red color shows strong cortical activity, while the orange color shows little brain activity. It can be observed from the brain maps that the alpha and beta bands of AF7 and AF8 frontal channels have visually distinct brain activity at baseline, 20 min, 40 min, and 60 min of the experiment.

Table 4.8: p-value for EEG power spectral densities in different brain regions

Time	**Channels**	**EEG Frequency Bands (*p* values by *t*-test)**				
		Delta	**Theta**	**Alpha**	**Beta**	**Gamma**
T1 – T2 (0 & 20 min)	AF7	7.011E-09*	0.00071*	0.06148	0.62845	0.09649
	AF8	2.924E-09*	1.438E-09*	0.04877*	1.345E-05*	2.519E-05*
	TP9	3.425E-05*	0.45987	0.56974	0.00568*	1.671E-17*
	TP10	0.00167*	2.883E-07*	1.446E-10*	1.959E-12*	7.304E-13*
T2 – T3 (20 & 40 min)	AF7	4.214E-05*	0.55471	0.00023*	0.76902	0.08094
	AF8	0.60858	0.00053*	0.00016*	3.219E-06*	0.13631
	TP9	0.02326*	0.52230	0.20485	1.716E-06*	0.18105
	TP10	0.01776*	0.98454	0.19671	0.12579	1.678E-11*
T3 – T4	AF7	0.13977	0.71663	0.97207	0.00155*	0.00023*

(40 & 60 min)	AF8	0.00480*	0.00295*	0.00241*	0.00026*	0.00024*
	TP9	0.00882*	0.00046*	0.01284*	0.00357*	0.00627*
	TP10	0.01746*	5.106E-05*	0.17877	0.00441*	0.00289*

**The mean difference is significant at the 0.05 level*

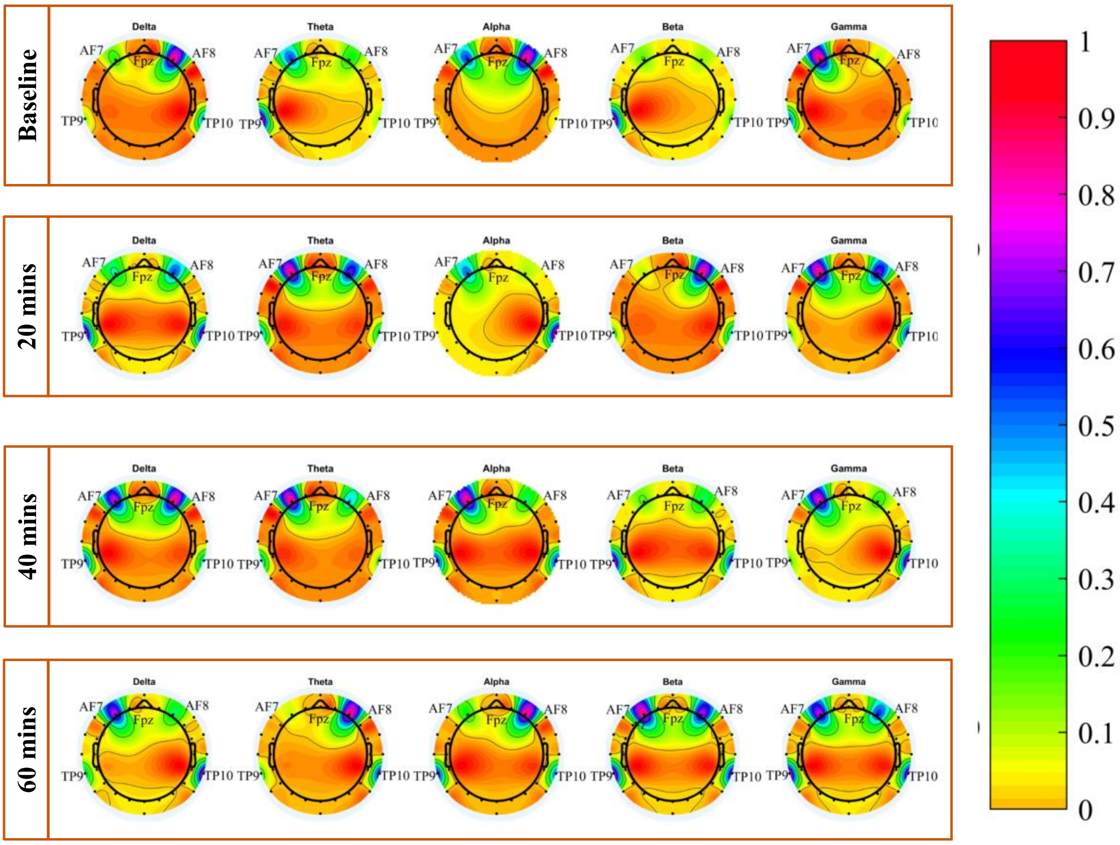


Figure 4.8: Brain activity visualization of different EEG bands for the four experiment phases

4.3.4. Validity of the facial features' geometric measurements

4.3.4.1. Correlations between facial features' geometric measurements and subjective mental fatigue scores

In Table 4.9, correlations between geometric measurements of facial features and subjective mental fatigue scores are shown. The eye aspect ratio at T-1 (r = -0.5202), T-3 (r = -0.6730), and T-4 (r = -0.5760) minutes of the experiment was significantly correlated with the corresponding subjective mental fatigue scores. Similarly, geometric measurements of eye distance facial features were significantly associated with subjective mental fatigue scores during all the experiment phases; T-1 (r = 0.7164), T-2 (r = 0.5029), T-3 (r = 0.6866) and T-4 (r = 0.9264). Furthermore, across all experiment phases, the head motion face feature was substantially linked with the corresponding subjective scores.

However, mouth aspect ratio was only correlated at T-4 (r = -0.5872). Also, at experiment phases T3 (r = 0.5884) and T-4 (r = 0.5078), face area feature was related. However, there was no association between the remaining facial features (e.g., eyebrows, nose to chin ratio, and nose to jaw ratio) and subjective mental fatigue.

Table 4.9: Correlations between facial features and subjective scores

Parameters	NASA-TLX Score				
	Time	**Baseline**	**20 min**	**40 min**	**60 min**
Eye Aspect Ratio	Baseline	-0.5202*			
	20 min		-0.4635		
	40 min			-0.6730**	
	60 min				-0.5760*
Eye Distance (Pixels)	Baseline	0.7164**			
	20 min		0.5029*		
	40 min			0.6866**	
	60 min				0.9264**
Eyebrow (Pixels)	Baseline	0.6318**			
	20 min		0.7327**		
	40 min			0.5695*	
	60 min				0.5967*
Mouth Aspect Ratio	Baseline	0.0075			
	20 min		0.0762		
	40 min			0.2226	
	60 min				-0.5872*
Nose to Jaw Ratio	Baseline	0.1448			
	20 min		0.0241		
	40 min			0.4504	
	60 min				0.2912
Nose to Chin Ratio	Baseline	-0.6134*			
	20 min		-0.5954*		
	40 min			-0.5288*	
	60 min				-0.6011*
Face Area ($Pixels^2$)	Baseline	0.1313			
	20 min		0.1382		
	40 min			0.5884*	
	60 min				0.5078*
Head Motion (Pixels per frame)	Baseline	0.5209*			
	20 min		0.6910**		
	40 min			0.5003*	
	60 min				0.5413*

**Correlation is significant at 0.05; **Correlation is significant at 0.01*

4.3.4.2. Correlations between facial features' geometric measurements and EEG metric

The correlations between facial features and electroencephalography metric [$(\theta + \alpha) / (\alpha + \beta)$] for mental fatigue are shown in Table 4.10. The eye aspect ratio was significantly correlated with EEG during all the experiment phases, i.e., at baseline (r = 0.6849), 20 min (r = 0.5008), 40 min (r = 0.5510), and 60 min (r = -0.5760) of the experiment. Similarly, geometric measurements of head motion facial features during experiment phases; at baseline (r = -0.5042), 20 min (r = -0.6234), 40 min (r = -0.5374), and 60 min (r = -0.4985) were significantly associated with the EEG metric. Furthermore, at baseline, 20 minutes, and 60 minutes of the experiment, the eye distance facial feature was found to be significantly linked with the EEG metric. The findings also revealed that eye aspect ratio was positively associated, whereas the eye distance and head motion facial features were negatively corelated with the EEG metric. However, the correlation of rest of the facial features with EEG metric was not monotonous during all the experiment phases as shown in Table 4.10.

Table 4.10: Correlation between EEG metric and facial features

Facial Features	EEG Metric [$(\theta + \alpha) / (\alpha + \beta)$]			
	Baseline	**20 mins**	**40 mins**	**60 mins**
Eye Aspect Ratio	0.6849*	0.5008*	0.5510*	0.6505*
Eye Distance	-0.6701*	-0.3608	-0.5497*	-0.7155*
Eyebrow	-0.5698*	-0.6034*	-0.4507	-0.2246
Nose to Jaw Ratio	-0.4007	-0.3472	-0.3618	-0.1323
Nose to Chin Ratio	0.5861*	0.4915	0.4717	0.2269
Mouth Aspect Ratio	0.1600	0.4466	0.1282	0.3830
Face Area	-0.1311	-0.3872	-0.5566*	-0.5881*
Head Motion	-0.5042*	-0.6234*	-0.5374*	-0.4985*

**Correlation is significant at the 0.05 level*

4.4. Objective 3: To explore the use of deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data.

In this section, we describe the findings of our investigations and the data we acquired from the operators. All fifteen construction equipment operators successfully completed the experiment. Therefore, data from all operators was used for analysis.

4.4.1. Analysis of ground truth data

The NASA-TLX score was utilized as a ground truth for recognizing mental fatigue states. Accordingly, Table 4.11 displays descriptive and analytical statistics derived from the ground truth evaluation. Subjective mental fatigue was significantly higher at the end of the NASA-TLX than at the start, increasing from 11.067 (SD = 2.764) to 64.733 (SD = 4.543). According to Table 4.11, operators reported increasing mental fatigue as the excavation operation progressed.

Table 4.11: Subjective assessment as a ground truth of mental fatigue

		Mental Fatigue States		
	Baseline	**Alert State**	**Mild Fatigue State**	**Fatigue State**
Subjective Assessment				
NASA-TLX Score (0-100)	11.25 (2.77)	30.81 (2.99)	45.00 (4.27)	65.25 (4.85)

**The scores for each state are mentioned as mean score (standard deviation)*

4.4.2. Analysis of physiological data

The absolute power for each frequency band of the EEG data acquired from all the channels of the MUSE headband was analyzed using the paired t-test across the three experimental phases (alert, mild fatigue, and fatigue) to make inferences about the underlying physiological processes. T-test results were interpreted in light of a null hypothesis and associated p-value. If the p-value for rejecting the null hypothesis was less than 0.05, then there was a statistically significant difference among the studied fatigue states. Table 4.12 displays the t-Stat for the power spectral density of EEG recordings made from various regions of the brain depicted in Figure 4.9. These findings indicate a statistically significant difference. For example, the t-test applied to EEG signals revealed that the alpha band was not found to be statistically significant at right frontal channel i.e., AS-MFS (t_{Stat} = 4.991, p < 0.05) and MFS-FS (t_{Stat} = -3.641, p < 0.05), whereas at the left frontal channel the alpha band was statistically significant only for comparison at fatigue states; AS-MFS (t_{Stat} = -4.816, p < 0.05). However, there was an increase in alpha activity as the experiment progressed from AS to FS, as demonstrated in Figure 4.9. Likewise, a similar trend was also shown for beta band at right frontal channel i.e., AS-MFS (t_{Stat} = 7.172, p < 0.05) and MFS-FS (t_{Stat} = -4.741, p < 0.05). However, this trend was inverse at left frontal channel for beta band. The beta band showed differences that were statistically significant in both the frontal and temporal parts of the brain. Overall, the theta band showed an increasing trend with an increase in mental fatigue from AS to FS in the frontal region of the brain, as demonstrated in Figure 4.9. The Delta band was found to be statistically significant in the left and right temporal regions. However, it was not significant for AS-MFS and MFS-FS at AF8 (t_{Stat} = 0.523) and AF7 (t_{Stat} = -1.559), respectively. Furthermore, the gamma band was found to be statistically significant between MFS-FS with $p < 0.05$ at TP9 (t_{Stat} = -3.175), AF7 (t_{Stat} = 4.814), AF8 (t_{Stat} = -4.791) and TP10 (t_{Stat} = -3.553). Table 4.12 depicts the statistical assessment of all the channels' bands. Previous studies have reported similar findings to our own, such as the study by Zhao et al. (2012) , which found significant theta and beta activity in the frontal regions of the brain. Other studies, including those by Nguyen et al. (2017), Käthner et al. (2014), and Dasari et al. (2010), also reported increased alpha and beta activity in the parietal region of the brain with an increase in mental fatigue. Additionally, Ma et al. (2018) reported increased alpha activity due to mental fatigue. Our investigations into mental fatigue show an increasing trend of frontal theta activity, which is consistent with previous studies by Trejo et al. (2015), Roy et al. (2013), and Dasari et al. (2010). Furthermore, the theta, alpha, and beta bands are the most commonly investigated EEG metrics for measuring mental fatigue. The $(\theta+\alpha)/\beta$ ratio is the most widely used EEG metric for mental fatigue assessment. Higher mental fatigue is associated with an increase in this metric, according to findings by Jap et al. (2009). In our research, we found that time-on-task had a significant increasing effect on the EEG metric $(\theta+\alpha)/\beta$ [$F = 15.011$, $p < 0.05$, $\eta^2 = 0.517$]. The value of the EEG metric $(\theta+\alpha)/\beta$ in the alert state, mild fatigue state, and fatigue state was 1.015, 1.482, and 1.739, respectively, indicating an increase in mental fatigue. These findings are consistent with previous investigations by Ma et al. (2018) and Li et al. (2017a). Construction equipment operators' brain activity was visualized by calculating the power spectral density from their EEG data when they were in the alert state, mild fatigue state and fatigue state, as shown in Figure 4.9. The red color on the brain maps represents high levels of cortical activity, whereas the orange tint represents low levels. Brain activity in the alpha and beta bands of the frontal AF7 and AF8 channels can be seen to change graphically from the alert state to fatigue state on the brain maps.

Table 4.12: t-Stat for EEG power spectral densities at different brain regions

Mental Fatigue States	Channels	EEG Bands (*t-Stat*)				
		Delta	Theta	Alpha	Beta	Gamma
AS and MFS	TP9	2.526*	0.655	-1.325	7.560*	1.403
	AF7	-5.699*	-0.604	-4.816*	0.299	-1.871
	AF8	0.523	4.389*	4.991*	7.172*	1.574
	TP10	-2.662*	0.020	-1.351	-1.621	-17.812*
MFS and FS	TP9	-3009*	-4.464*	-2.823*	-3.450*	-3.175*
	AF7	-1.559	-0.370	-0.036	3.856*	4.814*
	AF8	-3.306*	-3.543*	-3.641*	-4.741*	-4.791*
	TP10	-2.670*	-5.596*	-1.411	-3.348*	-3.553*

**The t-Stat values are significant at the 0.05 level*

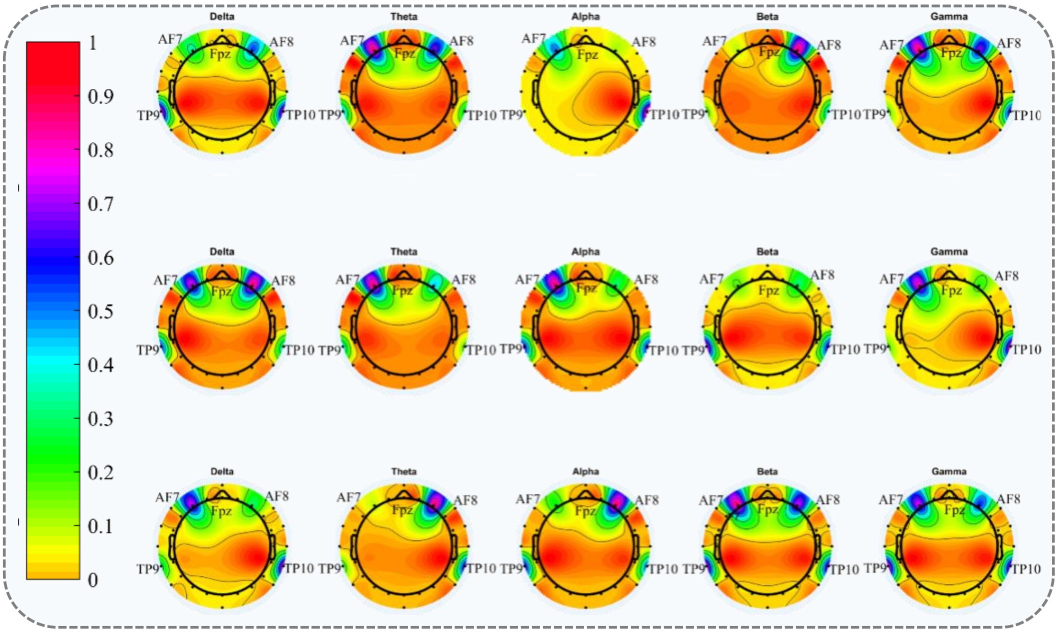


Figure 4.9: Brain visualization using EEG power spectral densities from different brain regions.

4.4.3. Deep learning-based classification results

Three deep learning models, LSTM, Bi-LSTM, and 1D-CNN were used to classify mental fatigue in construction equipment operators into alert, mild fatigue, and fatigue states. The implementation of LSTM, Bi-LSTM, and 1D-CNN employed cutting-edge parameter values, as demonstrated in Table 3.6. To limit experimental error, these models were run on the same system. The classification accuracies of the Bi-LSTM and LSTM deep learning models were both greater than 99%. The 1D-CNN deep learning model, on the other hand, only attained a classification accuracy marginally higher than 69%. However, upon evaluating the performance of the three deep learning models in terms of training time, it was found that the average duration for LSTM, Bi-LSTM, and 1D-CNN models was 68 minutes and 21 seconds, 163 minutes and 56 seconds, and 16 minutes and 57 seconds, respectively, as presented in Table 4.13. The findings show that, when trained on data reflecting operators' brain activity patterns over three increasingly demanding phases of work, the Bi-LSTM model outperformed the other deep learning models investigated in this study in terms of accuracy. Furthermore, LSTM also achieved accuracy slightly lower than Bi-LSTM when trained on the EEG sensor data.

Table 4.13: Classification accuracy and training time

Deep Learning Models	Accuracy (%)	Training Time
Long short-term memory (LSTM)	99.7063	68 mins 21 seconds
Bidirectional long short-term memory (Bi-LSTM)	99.9410	163 mins 56 seconds
One-dimensional convolutional network (1D-CNN)	69.4726	16 mins 57 seconds

4.4.3.1. Long short-term memory

Table 4.14 and Figure 4.10 illustrate the evaluation metrics and confusion matrix for the LSTM model. In general, the evaluation metrics demonstrated a good level of performance of the LSTM model on EEG-based brain activity data for identifying different mental fatigue levels in construction equipment operators. However, the performance of this model was slightly lower than that of Bi-LSTM. The LSTM model attained classification performance values ranging from 99.556% to 99.963% in terms of precision. FS represented 99.963% of instances of correctly identified fatigue levels. In addition, AS and MFS states exhibited the same effect on the LSTM model compared to FS, i.e., 99.556% and 99.589%, respectively. However, their effects were less than FS. Furthermore, higher recall and precision indicated that the model yielded fewer false negatives and false positives, respectively. Likewise, specificity and F1-score measures have values ranging between 99.761% and 99.818% and 99.681% and 99.718%, respectively. High specificity indicates the true negative rate, i.e., that a person identified as being in a fatigued state was in fact in that fatigued state. Besides, the confusion matrix was utilized to determine whether classes were misclassified or confused with others. As illustrated in Figure 4.10, each column depicts the actual mental fatigue states, while each row represents the predicted mental fatigue states. The diagonal cells indicate the correct instances for a more comprehensive evaluation of the classification performance at the end of the 30th epoch. The diagonal members of this matrix represent the cases in the dataset for which classification was accurate.

Incorrectly classified instances include nondiagonal elements. The high values of the diagonal elements imply that the model correctly distinguishes between the three classifications of mental fatigue. The other cells indicate the incidents that were incorrectly classified. It is also evident that alert and mild fatigue states were misclassified more often than fatigue state. In spite of this, the misclassification rate remains remarkably low when compared to their number of classified instances. Furthermore, AS was confused with MFS and FS in 1299 and 1949 instances, respectively.

Table 4.14: Performance evaluation metrics for LSTM model

Indicator	Testing	Alert State	Mild Fatigue State	Fatigue State
Accuracy	99.7063%			
Precision		99.5569%	99.5898%	99.9634%
Recall		99.8807%	99.7735%	99.4693%
Specificity		99.7613%	99.8185%	99.9808%
F1-score		99.7186%	99.6816%	99.7157%

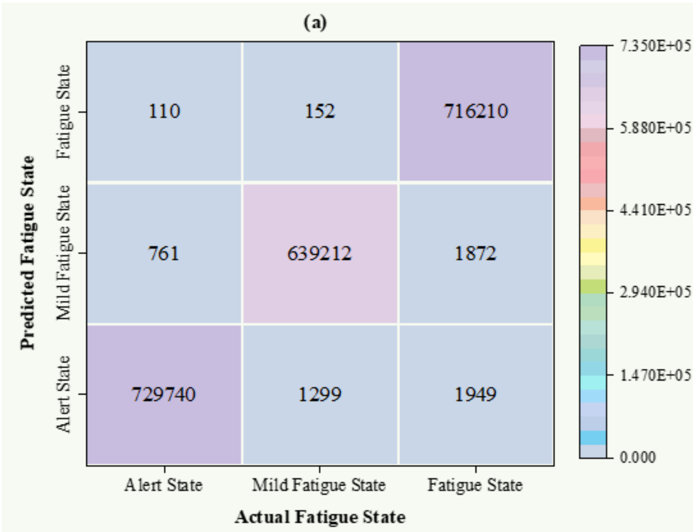


Figure 4.10: Confusion matrix for LSTM Model

4.4.3.2. Bidirectional long short-term memory

The evaluation matrix and confusion matrix of the bidirectional LSTM model are presented in Table 4.15 and Figure 4.11, respectively. Bi-LSTM evaluation measures indicated the highest performance on EEG-based brain activity data for identifying distinct mental fatigue levels in construction equipment operators. This shows that Bi-LSTM is most effective in our construction equipment operation-related task. Results for accuracy-related classification performance for the Bi-LSTM model ranged from

99.840% to 99.995%. The MFS and FS indicated approximately comparable instances of correctly identified fatigue levels with a precision slightly above 99.995%; however, the AS exhibited a little less of an effect on the Bi-LSTM model with a precision of 99.840%. In addition, greater recall and precision indicated that the model produced fewer false negatives and, consequently, false positives. Similarly, specificity measures have values ranging from 99.914% to 99.997%, while the F1-score has values ranging from 99.917% to 99.993%. High specificity demonstrates the true negative rate, i.e., a person identified with any fatigue state was indeed experiencing that fatigue level. According to the confusion matrix in Figure 4.11, it can be observed that MFS and FS are the most recognized classes, with 640609 and 718872 positive instances, respectively. Furthermore, it is notable that AS was misclassified more frequently than MFS and FS. However, the misclassification rate was exceptionally low in comparison to the number of instances that were correctly identified. The confusion matrix further indicates that the AS was 1141 times confused with the FS. However, the confusion among the remaining states was modest.

Table 4.15: Performance evaluation metrics for Bi-LSTM model

	Testing			
Indicator		**Alert State**	**Mild Fatigue State**	**Fatigue State**
Accuracy	99.9410%			
Precision		99.8409%	99.9945%	99.9952%
Recall		99.9972%	99.9915%	99.8390%
Specificity		99.9144%	99.9975%	99.9975%
F1-score		99.9190%	99.9930%	99.9170%

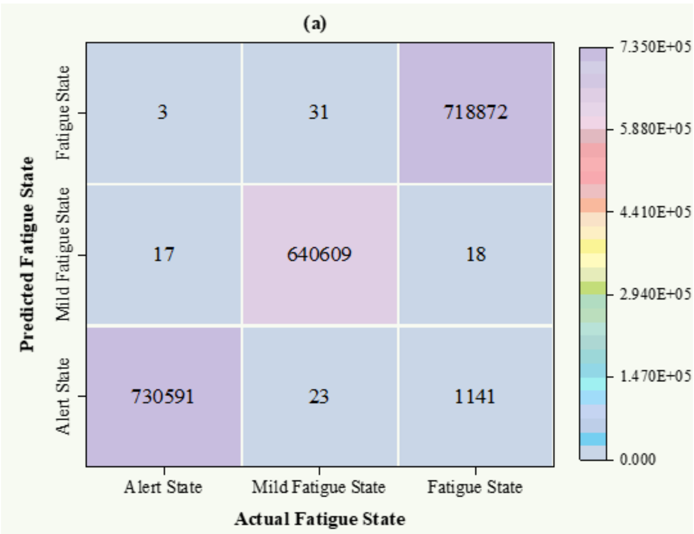


Figure 4.11: Confusion matrix for Bi-LSTM Model

4.4.3.3. One-dimensional convolutional network

Table 4.16 and Figure 4.12 exhibit the evaluation matrix and confusion matrix of the 1-dimensional convolutional network (1DCN) model, with correct classes provided in the diagonal cells for a more detailed evaluation of classification performances at the end of the 30th epoch. When compared to the LSTM and Bi-LSTM models, the evaluation metrics of the 1DCN model achieved the lowest performance. In terms of precision, the 1-dimensional convolutional model produced classification performance values ranging from 54.600% to 84.241%. FS had the highest percentage of accurately classified instances, i.e., 72.545%. Furthermore, AS had the lowest accurately categorized instances, i.e., 65.387%. Moreover, for MFS, the model produced a high number of false negatives and false positives in this state as compared to other fatigue stages, i.e., 227,581 times with AS and 148,916 times with FS. Similarly, specificity measurements range from 74.046% to 92.874%, while the F1-score ranges from 61.607% to 77.957%. These findings reveal that the 1-dimensional convolutional model underperformed the LSTM or Bi-LSTM models based on EEG data in classifying mental fatigue in construction equipment operators. Furthermore, the confusion matrix in Figure 4.12 indicates that FS was the most recognized class, with 52,2348 affirmative instances.

Table 4.16: Performance evaluation metrics for 1D-CNN model

	Testing			
Indicator		**Alert State**	**Mild Fatigue State**	**Fatigue State**
Accuracy	69.4726%			
Precision		74.4194%	54.6009%	84.2415%
Recall		65.3874%	70.6780%	72.5452%
Specificity		87.9317%	74.0461%	92.8743%
F1-score		69.6116%	61.6078%	77.9570%

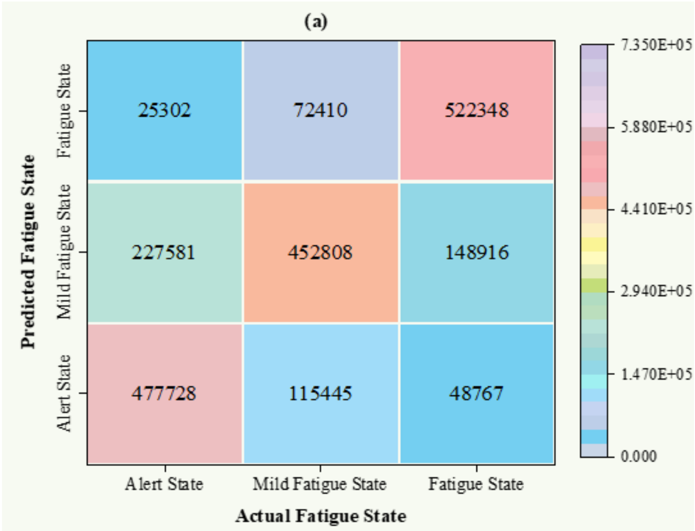


Figure 4.12: Confusion matrix for 1D-CNN Model

4.4.3.4. Train/test accuracy and loss

The accuracy and loss over iterations curves of the three deep learning models investigated in this study are shown in Figure 4.13, respectively. The training and validation results for a bidirectional LSTM model show higher accuracy and lower loss, as shown in Figure 4.13(b). Specifically, the bidirectional LSTM model exhibited the maximum accuracy during training and validation, while the associated loss value was the lowest at the 30^{th} epoch. As a result, the Bi-LSTM model was effectively trained without overfitting the EEG-based brain activity data of construction equipment operators, as demonstrated by the smallest difference between training accuracy and validation accuracy or training loss and validation loss.

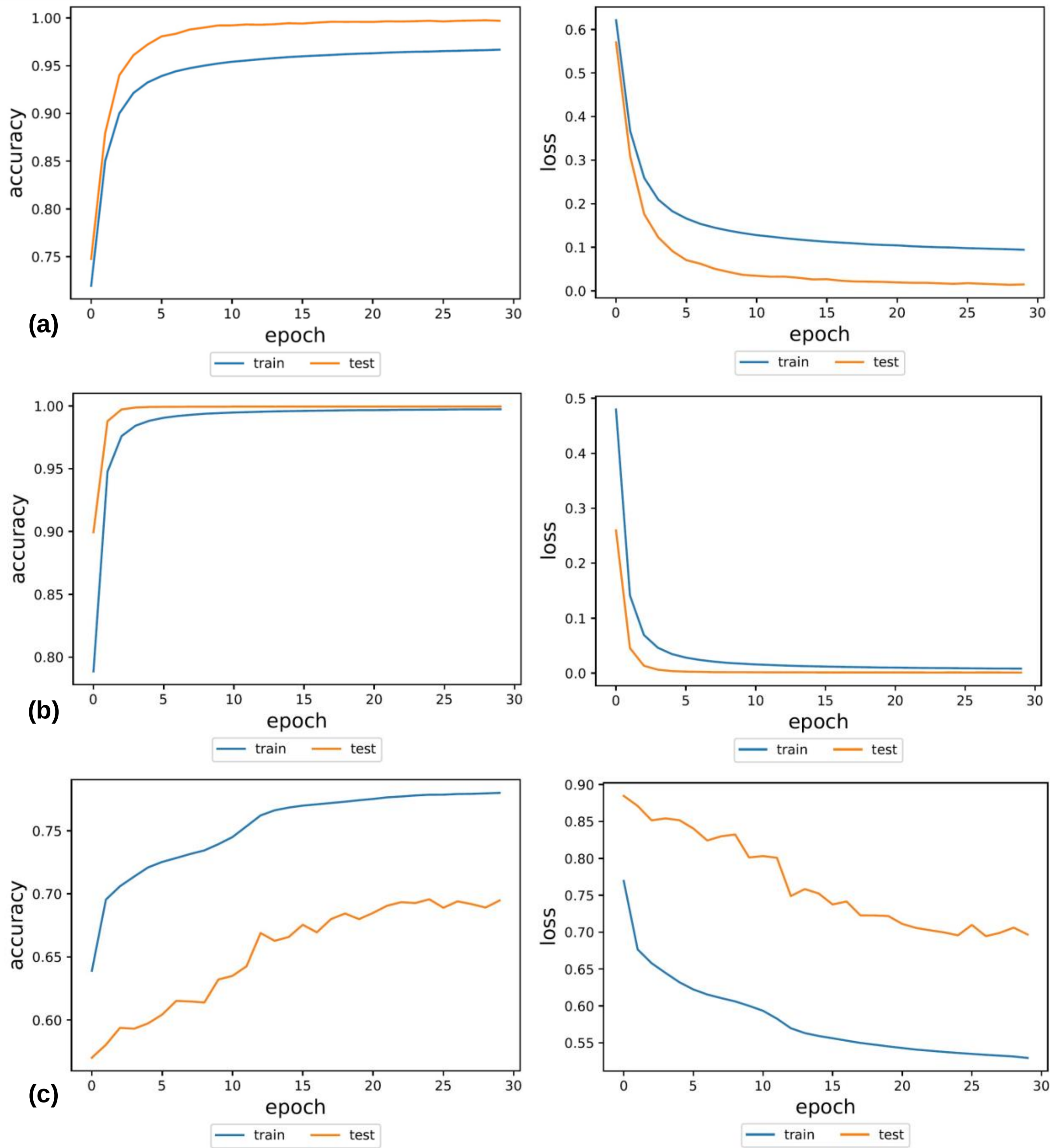


Figure 4.13: Accuracy and loss over iteration curves with the tuned hyperparameters of (a) LSTM model, (b) Bi-LSTM model, and (c) 1D-CNN model

4.4.3.5. Comparison of p-values for deep learning models

The p-values of the Mann-Whitney test computed on the results given by the bidirectional long short-term memory (Bi-LSTM) findings are presented in Table 4.17. The results demonstrate that the bidirectional LSTM model's accuracy was considerably higher to that of the other two models, i.e., the LSTM and the 1D-CNN, for 10-fold cross-validation.

Table 4.17: Mann-Whitney test-based comparison of p-values for Bi-LSTM model

Model	Validation method	LSTM		1D-CNN	
		***p*-value**	**Significance**	***p*-value**	**Significance**
Bi-LSTM	10-fold cross validation	0.007994	Sig.	7.22×10^{-30}	Highly Sig.
LSTM	10-fold cross validation			1.4×10^{-28}	Highly Sig.

** Significant at p < 0.01*

4.5. Objective 4: To study the multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals.

4.5.1. Analysis of ground truth data

In this study, the NASA-TLX score was employed as a reliable measure for identifying mental fatigue states. The findings presented in Table 4.18 demonstrate the descriptive statistics derived from the ground truth analysis. Notably, subjective mental fatigue was found to be significantly higher at the end of the experiment than at the beginning, exhibiting an increase from 11.25 (SD = 2.77) to 65.25 (SD = 4.85). Additionally, the results displayed in Table 4.18 indicate that the operators experienced progressively higher levels of mental fatigue as the excavation operation continued.

Table 4.18: Ground truth of mental fatigue

		Mental Fatigue States		
	Baseline	**Alert State**	**Mild Fatigue State**	**Fatigue State**
Subjective Assessment				
NASA-TLX Score (0-100)	11.25 (2.77)	30.81 (2.99)	45.00 (4.27)	65.25 (4.85)

4.5.2. Machine Learning-based classification results for multimodal data

This study utilized a novel approach to identify and classify mental fatigue states in construction equipment operators by integrating input data from multiple sensors and employing machine learning techniques. Three machine learning models (ANN, k-NN, and DT) were used to classify mental fatigue into alert, mild fatigue, and fatigue states. In addition to the EEG data, input data also included electrodermal activity (EDA) and geometric measurements of facial features. These data were fused together as input for the machine learning models used in the study. Furthermore, input data from multiple sensors was fused in various combinations, including (a) EEG and EDA, (b) EEG and FF, (c)

EDA and FF, and (d) EEG, EDA, and FF. The results, as demonstrated in Table 4.19, Table 4.20, and Table 4.21, indicated that the machine learning models achieved classification accuracies ranging from 56.5% to 97.1%. However, the decision tree models achieved the highest accuracies for all input data combinations, ranging from 85.0% to 97.1%. The findings of the study indicate that the decision models outperformed the other machine learning models investigated in terms of accuracy when trained on input data from multiple sensors of operators over three increasingly demanding phases of work.

4.5.2.1. Neural Network (NN)

The evaluation metrics and confusion matrix presented in Table 4.19 and Figure 4.14 indicate the performance of an Artificial Neural Network (ANN) model for identifying different levels of mental fatigue in construction equipment operators. Overall, the evaluation metrics demonstrated good performance of the model on different input data fusions. However, the performance of the model was slightly lower than that of Decision Tree (DT) models. The ANN model achieved an accuracy ranging from 73.5% to 96.6% for all input data combinations, with the highest accuracy of 96.6% achieved using FF and EDA as input data. The model's classification performance ranged from 93.96% to 98.347% in terms of precision, with FS and MDS representing the highest values of correctly identified fatigue levels. Additionally, higher recall and precision indicated that the model yielded fewer false negatives and false positives, respectively. Specificity and F1-score measures ranged between 96.783% and 99.168% and 95.979% and 98.892%, respectively. Besides, the confusion matrix was utilized to determine whether classes were misclassified or confused with others. As demonstrated in Figure 4.14, the high values of the diagonal elements imply that the model correctly distinguished between the three classifications of mental fatigue. The other cells indicate the incidents that were incorrectly classified. Alert states were misclassified more often than mild fatigue and fatigue states. It was confused with MFS in 23 instances, as demonstrated in Figure 4.14(a). However, the misclassification rate remains remarkably low compared to the number of classified instances. In addition, the results were similar for the combination of all three sensors' data (EEG, EDA, and FF), with an overall classification accuracy of 94.7%. MFS was the most misclassified state, being confused with AS and FS in 14 and 10 instances, respectively. Moreover, the combination of EEG and FF exhibited slightly less accuracy compared to the above-mentioned two combinations, with an overall accuracy of 87.8%. The fourth combination, with EEG and EDA as input data, attained the lowest overall accuracy among the four combinations, with an accuracy of 73.8%. This combination also exhibited the highest number of misclassified classes among all combinations, with MFS being confused with AS and FS in 53 and 62 instances, respectively.

Table 4.19: Performance assessment metrics for NN models

Indicator	Testing	Alert State	Mild Fatigue State	Fatigue State
FF-EDA				
Accuracy	96.6	97.222	96.667	99.259
Precision		93.963	97.619	98.347
Recall		98.082	92.134	99.443
Specificity		96.783	98.895	99.168
F1-score		95.979	94.798	98.892
EEG-EDA				
Accuracy	73.8	85.093	78.519	83.981
Precision		76.289	67.514	77.515
Recall		81.096	67.135	72.981
Specificity		87.133	84.116	89.459
F1-score		78.619	67.324	75.179
EEG-FF				
Accuracy	87.8	93.981	88.426	93.148
Precision		91.667	84.894	86.632
Recall		90.411	78.932	93.871
Specificity		95.804	93.093	92.788
F1-score		91.034	81.805	90.107
EEG-EDA-FF				
Accuracy	94.7	96.481	95.093	97.870
Precision		94.550	93.162	96.409
Recall		95.068	91.854	97.214
Specificity		97.203	96.685	98.197
F1-score		94.809	92.504	96.810

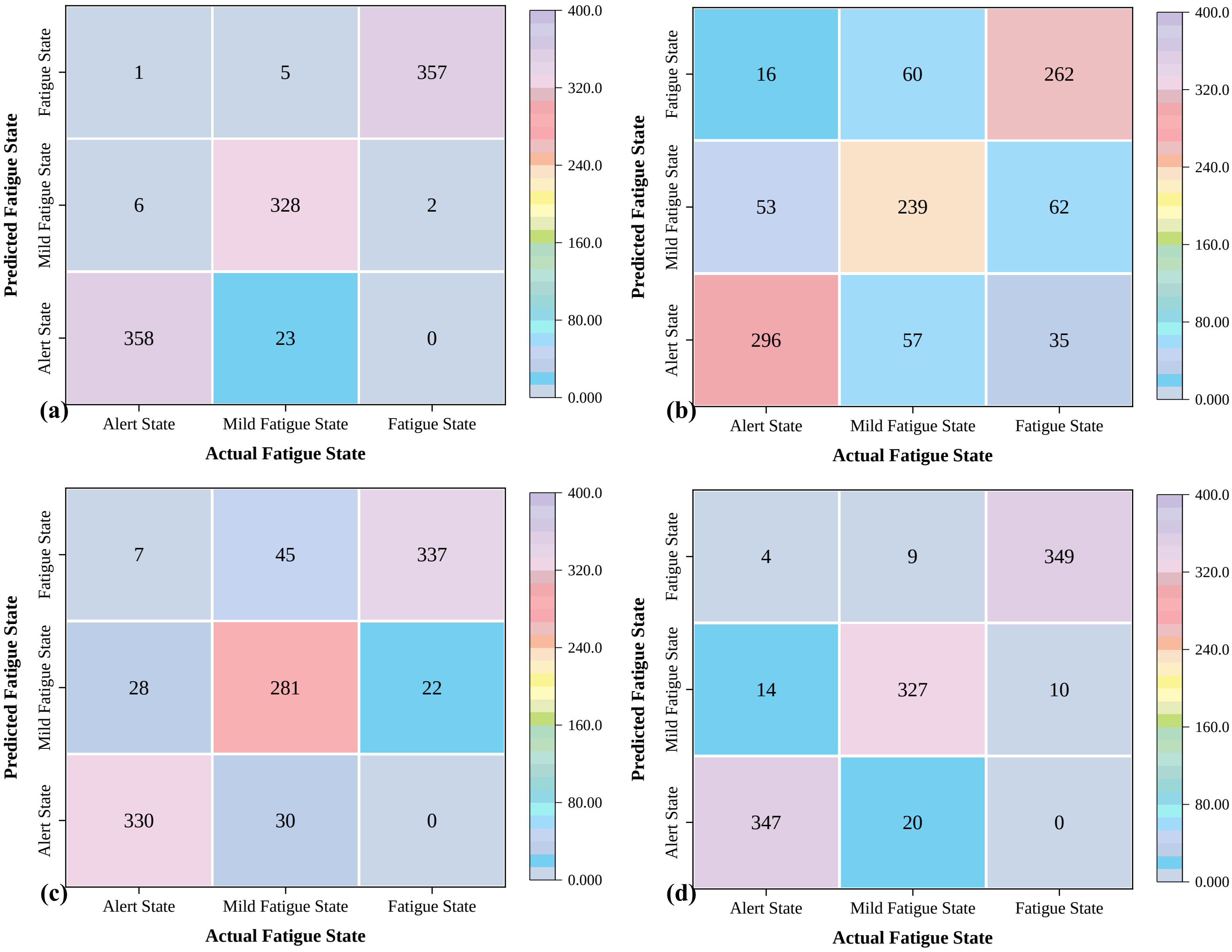


Figure 4.14: Confusion matrix for NN (a) FF-EDA, (b) EEG-EDA, (c) EEG-FF, and (d) EEG-EDA-FF

4.5.2.2. K-Nearest Neighbors (kNN)

Table 4.20 presents the evaluation matrix, and Figure 4.15 shows the confusion matrix of the k-nearest neighbors model. When used on all possible combinations of input data, k-NN performed inferior to ANN and DT at figuring out the different levels of mental fatigue in construction equipment operators. Nonetheless, the overall accuracy, except for one combination of input data, was above 80%. The k-NN model attained overall performance accuracy values ranging from 56.5% to 94.4% for all combinations of input data. The model attained an overall accuracy of 94.4% while employing FF and EDA as input data. Consequently, the MFS indicated higher instances of correctly identified fatigue levels with a precision slightly above 96.5%. However, the AS and FS exhibited a little less effect on the k-NN model with a precision of 92.802% and 94.086%, respectively, as demonstrated in Table 4.20. The model attained the highest values of precision and recall for the aforementioned combination, indicating that it yielded fewer false positives and negatives. Similarly, specificity and F1-score measures have values ranging between 96.084% and 98.481% and 91.259% and 95.759%, respectively, indicating that an operator identified as being in a particular fatigue state was, in fact, in that fatigue state. Furthermore, using EEG and FF as input data, the model attained an overall accuracy of 87.5%, with classification precision values ranging between 86.553% and 88.950%. Interestingly, the model attained higher specificity values, ranging between 92.308% and 94.452%, indicating that the operator who identified any fatigue state was actually experiencing that state. Similarly, a comparable overall accuracy of 85.8% was achieved while using input data from all three sensors combined. Consequently, the classification performance in terms of precision remained between 84.048% and 88.430%. According to the confusion matrix in Figure 4.15, it can be observed that the confusion among the mental fatigue states was modest except for the combination of EEG and EDA. The misclassification rate for this combination was exceptionally high, as demonstrated in Figure 4.15(b). It is noteworthy that the AS and FS were the most recognized states, as shown in Figures 4.15(a), 4.15(c), and 4.15(d). AS was recognized with 361 (FF and EDA), 354 (EEG and FF), and 353 (EEG, EDA, and FF) positive instances. Furthermore, when we see the confusion matrix demonstrated in Figure 4.15(b), AS was 136 and 101 times confused with MFS and FS, respectively.

Table 4.20: Performance assessment metrics for k-NN models

	Testing			
Indicator		**Alert State**	**Mild Fatigue State**	**Fatigue State**
FF-EDA				
Accuracy	94.4	97.037	94.537	97.130
Precision		92.802	96.552	94.086
Recall		98.904	86.517	97.493
Specificity		96.084	98.481	96.949
F1-score		95.756	91.259	95.759
EEG-EDA				
Accuracy	56.5	67.778	70.000	75.185
Precision		51.731	56.061	64.000
Recall		69.589	41.573	57.939
Specificity		66.853	83.978	83.773
F1-score		59.346	47.742	60.819
EEG-FF				
Accuracy	87.5	93.889	88.241	92.873

Precision		86.553	87.055	88.950
Recall		96.986	75.562	89.694
Specificity		92.308	94.475	94.452
F1-score		91.473	80.902	89.320
EEG-EDA-FF				
Accuracy	85.8	92.685	86.389	92.593
Precision		84.048	85.185	88.430
Recall		96.712	71.067	89.415
Specificity		90.629	93.923	94.175
F1-score		89.936	77.489	88.920

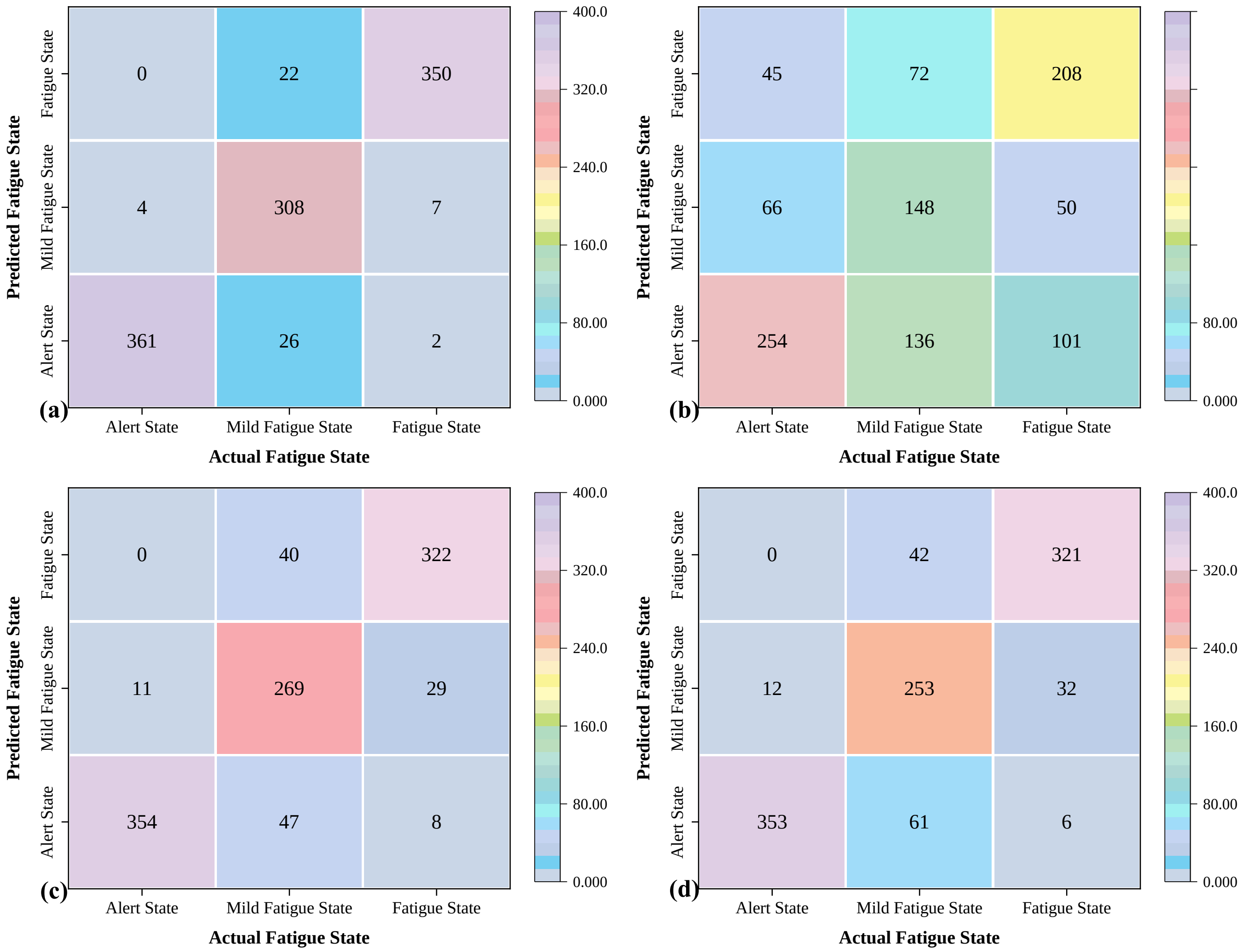


Figure 4.15: Confusion matrix for k-NN (a) FF-EDA, (b) EEG-EDA, (c) EEG-FF, and (d) EEG-EDA-FF

4.5.2.3. *Decision Tree (DT)*

Table 4.21 and Figure 4.16 present the evaluation metrics and confusion matrix for the decision tree (DT) model, which includes correct classifications displayed in the diagonal cells for a more detailed evaluation of classification performance. Compared to the ANN and k-NN models, the DT model achieved the highest overall accuracy, with a range between 85.0% to 97.1% for all input data combinations. It is important to note that using EEG and EDA as input data resulted in an accuracy of 85.0%, while all other input data combinations achieved an accuracy above 96.0%. When using data from all sensors as input, AS had the highest accurately classified instances at 97.568%. In contrast, FS had the lowest percentage of accurately classified instances compared to AS and MFS, at 94.370%. Additionally, the model produced a high number of false negatives and false positives for FS compared to other fatigue stages, with 21 times FS being confused with MFS. However, this confusion numbers are modest compared to the other combinations of input data. The specificity and F1-score measures ranged between 97.087% and 98.741% and 94.084% and 98.231%, respectively, indicating that an operator identified as being in a particular fatigue state was indeed in that fatigue state. Moreover, the FF and EDA, and EEG and FF input data combinations also resulted in higher instances of correctly identified fatigue levels, with modest confusion among mental fatigue states. On the other hand, using EEG and EDA as input data resulted in higher confusion among the mental fatigue states. Nonetheless, the confusion among the states was still modest compared to the findings indicated by the ANN and k-NN models as demonstrated in Tables 4.19 and 4.20, respectively. Figure 4.16(a-d) demonstrates that AS and FS were recognized with 364 and 352 (FF and EDA), 331 and 294 (EEG and FF), 363 and 347 (EEG and FF), and 361 and 352 (EEG, EDA, and FF) positive instances, respectively.

Table 4.21: Performance assessment metrics for DT models

Indicator	Testing	Alert State	Mild Fatigue State	Fatigue State
FF-EDA				
Accuracy	96.9	99.538	96.944	97.407
Precision		98.913	97.640	94.370
Recall		99.726	92.978	98.050
Specificity		99.441	98.895	97.087
F1-score		99.318	95.252	96.175
EEG-EDA				
Accuracy	85.0	92.037	88.426	89.537
Precision		86.423	82.535	85.965
Recall		90.685	82.303	81.894
Specificity		92.727	91.436	93.343

F1-score		88.503	82.419	83.880
EEG-FF				
Accuracy	97.1	98.796	97.778	97.685
Precision		97.059	97.977	96.389
Recall		99.452	95.225	96.657
Specificity		98.462	99.033	98.197
F1-score		98.241	96.581	96.523
EEG-EDA-FF				
Accuracy	96.2	98.796	96.204	97.407
Precision		97.568	96.736	94.370
Recall		98.904	91.573	98.050
Specificity		98.741	98.481	97.087
F1-score		98.231	94.084	96.175

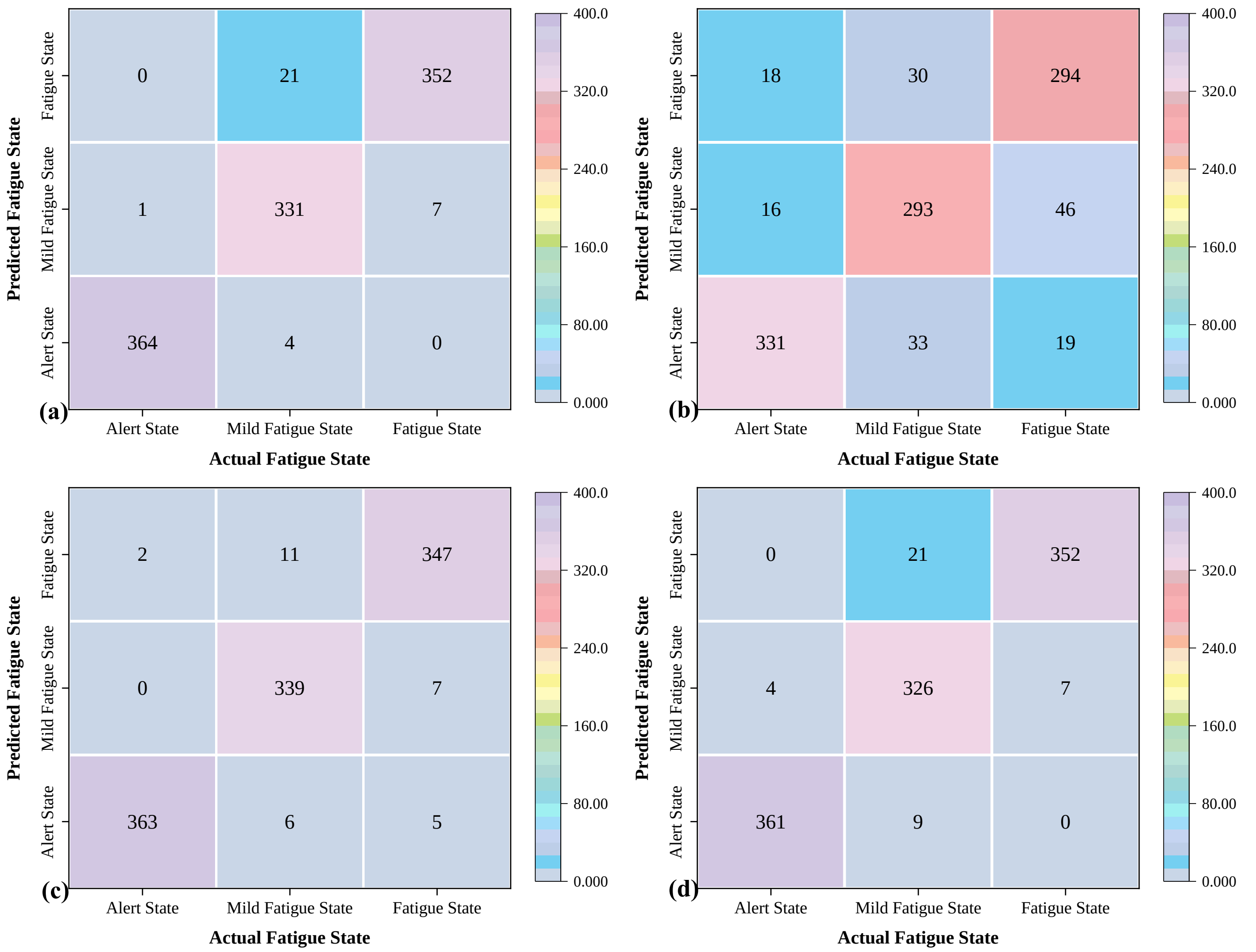


Figure 4.16: Confusion matrix for DT (a) FF-EDA, (b) EEG-EDA, (c) EEG-FF, and (d) EEG-EDA-FF

4.6. Summary

This chapter described the research findings for research objective. The findings indicated a statistically significant difference in the geometric measurements of facial features because of mental fatigue. Furthermore, the findings also established the ecological validity of proposed method for construction equipment operators. Furthermore, the research showed that the bi-LSTM model exhibited highest accuracy for classifying mental fatigue. Lastly, the results provided the feasibility of multimodal data integration from multiple data sources to classify mental fatigue. DT models exhibited highest accuracy.

Chapter 5

Discussion[8]

5.1 Introduction

Construction managers are always concerned with occupational safety management. Globally, the high incidence of mental fatigue among construction equipment operators impedes occupational safety and productivity (Masullo et al., 2021, Das et al., 2020). This is because the construction equipment operations are cognitively demanding and necessitate the operators' undivided attention. Such protracted attention induces mental fatigue in construction equipment operators, which is one of the leading causes of construction-site equipment-related accidents. Therefore, detecting and assessing mental workload is critical to ensuring operator safety, minimizing fatigue-related errors, reliable construction equipment operations (Han et al., 2020, Das et al., 2020), and to reduce equipment-related incidents and make construction sites safe for workers. As a result, it is imperative that the mental fatigue of construction equipment operators be monitored non-invasively. Hence, this chapter discusses the findings of current research. It also compares the findings with the studies conducted in other industries for mental fatigue monitoring and assessment. Lastly, this chapter also provides an overview of limitations and future research.

5.2 Discussion related to non-invasive detection of mental fatigue in operators.

Construction managers are always concerned with occupational safety management. Globally, the high incidence of mental fatigue among construction equipment operators impedes occupational safety and

[8] This chapter is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" *Advanced Engineering Informatics*, Volume 54, 101777
Imran Mehmood, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". *Advanced Engineering Informatics*, Volume 56, 101978
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". *Developments in the Built Environment,* Volume 15, 100198
Imran Mehmood, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". *Journal of Safety Research*, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291

productivity (Masullo et al., 2021, Das et al., 2020). Therefore, detecting and assessing mental workload is critical to ensuring operator safety, minimizing fatigue-related errors, and reliable construction equipment operations (Han et al., 2020, Das et al., 2020). In the construction industry, the current study's non-invasive methodology is unprecedented. The results are statistically significant and provide support for the notion that geometric measurement of facial features could be utilized to detect mental fatigue.

5.2.1 Changes in geometric measurement of eye-related features

While there have been similar studies in other domains, as per the author's knowledge, there is no study in the literature that has utilized geometric measurement of facial features for mental fatigue evaluation in the construction industry generally and construction equipment operators specifically. The findings of this research are in line with those of similar investigations conducted in non-construction domain that have utilized eye-related features for mental stress-led fatigue detection. For example, Giannakakis et al. (2017) found an increase in mean eye aperture in stressful conditions related to watching videos and images. Similarly, an increase in orbicularis oculi (a muscle associated with eyelid movement) electromyography activity was reported in a non-neutral emotional state as compared to a neutral state (Ravaja et al., 2006). Likewise, Bevilacqua et al. (2018) reported a change in eye metrics when people were subjected to stressful scenarios of a game. Besides, the increase in the eyebrow metrics (i.e., mean value) in the current study indicates that the eyebrows are more raised in high mental fatigue and tend to be closer to the nose in low mental fatigue as shown in Table 4.2. These results are in agreement with the previous studies for example, Hazlett (2006) reported more frequent corrugator (a face muscle associated with eyebrows) activity in a non-neutral emotional state than in a neutral state. Similarly, Kimmelman et al. (2020) also found that the eyebrow position is affected by emotions.

5.2.2 Changes in geometric measurement of mouth-related features

Mouth-related features also seem to be a good indicator of mental fatigue among construction equipment operators. For instance, the current study found an increase in the mean values of mouth because of high mental fatigue. Likewise, high mental fatigue also led to an increase in the distance between the two mouth corners and anchor landmark as shown in Figure 4.3(b) and Table 4.3. These results are also in accordance with previously conducted non-construction studies. For example, Tijs et al. (2008) and Ravaja et al. (2006) reported an increase in zygomatic (a face muscle associated with mouth) activity in an emotional state. Similarly, Dinges et al. (2005) and Metaxas et al. (2004) found that lip movement is influenced by stress and anxiety levels. Furthermore, Giannakakis et al. (2017) also reported an increase in variance and median of the maximum magnitude of mouth activity, which indicate faster mouth movements during stressful conditions.

5.2.3 Changes in geometric measurement of head-related features

Face area and head motion metrics were also found to be affected by mental fatigue. Both the head-related features are related to the dynamic body movement i.e., head movement and corporal posture (Bevilacqua et al., 2018) of construction equipment operators during ongoing site tasks. Therefore, an increase in the mean values of both the features might indicate more corporal movement during high mental fatigue. The mean values of the face area are directly connected with the operators' movement towards and away from the camera. An increase in the mean values of the face area indicates that the operators were closer to the camera during high mental fatigue, trying to increase their concentration (Bevilacqua et al., 2018). Similarly, the head motion feature is associated with the vertical, horizontal, and rotational movements of operators' heads during ongoing construction tasks. This feature is affected by head rotations. The increase in the mean value of head motion indicate that the equipment operators were less still in high mental fatigue than in low mental fatigue. It also means more rotation of operators' heads during high mental fatigue. The presence of statistical significance shows that the change in the mean values is associated with equipment operators' high mental fatigue. The findings of this study are also in accordance with the investigations in the domains other than the construction industry, that have utilized head-related features for mental stress-led fatigue detection. For example, Giannakakis et al. (2018) reported an increased head motion, when the participants were watching stressful videos and images. Furthermore, an increase in amplitude of head motion and velocity was also reported in a non-neutral emotional state compared to the neutral state among the participants (Giannakakis et al., 2017). Kusano et al. (2020) also reported more positive values of head motion in stressful conditions. Likewise, Dinges et al. (2005) and Liao et al. (2005a) also reported an increase in head movements during stressful moments. However, the results of this study are contrary to the findings by Bevilacqua et al. (2018), that did not show a statistical significance for head movements between non-neutral and neutral states.

5.2.4 Analysis of the correlation between geometric measurements of face features and subjective scores

There were significant connections between geometric measurements of facial characteristics and subjective mental fatigue scores during the excavation operation experiment. Some variables were associated with subjective scores in both mental fatigue groups, while rest was not associated with either. As indicated in Table 4.5, the eye area and brow facial features were significantly related to subjective scores in the two mental fatigue groups, LMF and HMF. Previous research has discovered that fatigue assessments are strongly linked to eye-related indicators (Sundelin et al., 2013). Tran and Yan (2022) found that increasing the time spent on the activity increased subjective mental fatigue and pupil diameter, with a commensurate reduction in cognitive performance. Similarly, Dziuda et al. (2021) discovered that changes in the drivers' percentage closure of eye time levels were connected with their responses to the fatigue symptoms scale questionnaire before and after the simulator task. Likewise, Zheng et al. (2012) also reported a correlation between eye-related parameters and subjective assessments.

5.2.5 Comparison of geometric measurements of facial features approach according to the published literature.

Appendix E summarizes the comparison between various studies on variations in facial features for mental fatigue assessment. However, it is challenging to compare the variations in facial features for construction equipment operators in this study and previous similar studies in other domains. This is due to several reasons such as the differences in experiment protocols, nature of construction tasks performed on construction sites. According to Liu et al. (2021a), operators of construction equipment work in fundamentally different ways. For example, during equipment operations, excavator operators continuously move their heads to track the excavator’s bucket. Moreover, the studies in other domain were conducted in laboratories or simulators under controlled environments. The data in these studies was collected from the individuals watching stressful videos, playing video games, and undergoing some social exposure or a lane change test. However, these studies cannot capture the dynamics and complexity of a real construction site, where construction workers are always vulnerable to frequently occurring accidents. All these data collection techniques are significantly different from the construction equipment operations. Hence, it is challenging to compare their results with our study in a true sense. We believe that the context of our work is substantially different from the current work in construction domain for mental fatigue detection of construction equipment operators.

5.3 Discussion related to the ecological validity of facial features' geometric measurements.

As far as the authors know, no study has compared the proposed method to invasive methods like electroencephalography that are used to monitor mental fatigue in construction equipment operators. The findings described in Table 4.10 shows significant association between geometric measurements of facial features and brain activity of operators using electroencephalography technology. This serves as ecological validity of proposed method to assess mental fatigue in construction equipment operators at construction sites.

5.3.1 Variations in the facial features' geometric measurements

The findings of this research are in line with those conducted in non-construction domains that have utilized facial features for mental fatigue detection. The current study used geometric measurements of eight facial features: eye aspect ratio, eye distance, eyebrow, nose to chin ratio, nose to jaw ratio, mouth aspect ratio, and head motion. Comparable studies in non-construction domains have used eye-related variables for mental fatigue detection with similar findings. There was a statistically significant difference in eye aspect ratio and eye distance. From baseline until the end of the experiment, they demonstrated a rise in the mean values of eye distance and a decrease in the mean values of eye aspect ratio. The variation in mean values reveals that landmarks were moved closer together as mental fatigue increased among equipment operators. Therefore, such a variation pattern is suggestive of increased blinking and eye closure due to increased mental fatigue. Hence, the construction equipment operators' cognitive effort increased. Likewise, the study found an increase in the eyebrow. However, the increase was not statistically significant. The results are aligned with the previous studies that showed an increase in the blinking of eyes during fatigue states. For example, Giannakakis et al. (2017) and Norzali et al. (2014) reported an increase in the blink rate under stressful situations and concluded that blink rate and mental stress are highly correlated with each other. Nevertheless, Wenhui et al. (2005) reported that the eye blinks decreased with an increase in cognitive effort. A change in eye metrics was also found by

Bevilacqua et al. (2018) in a study where subjects were subjected to stressful scenarios of a game. Likewise, Ravaja et al. (2006) also stated an increase in orbicularis oculi (a muscle associated with eyelid movement) electromyography activity in non-neutral emotional states. Our study found no statistically significant differences in eyebrow activity among operators, although the variation is consistent with earlier research. For example, a study by Kimmelman et al. (2020) stated that eyebrow positions are affected by emotional states.

Mouth-related features of construction equipment operators appear to be indicators of mental fatigue. This study demonstrated an increase in the mean mouth aspect ratio from baseline to the last experiment phase. The increase was statistically significant. The increase in mean values indicates that the position of mouth landmarks strayed away from each other due to increased mental fatigue. Similarly, such a change may be indicative of frequent mouth movements with an increase in mental fatigue. For example, a study by Giannakakis et al. (2017) reported that an increased variation and median of the highest magnitude of mouth activity imply faster mouth movements during stressful conditions. Similarly, as studied by Tang et al. (2016), the mouth remains closed in a normal state while it opens when a subject is in fatigued state. Likewise, Tijs et al. (2008) reported that in emotional states, the zygomatic (a face muscle that is linked to the mouth) is more active.

Mental fatigue also affects facial traits linked to construction equipment operators' dynamic body motions, such as head motion, face area, nose to chin ratio, and nose to jaw ratio. Bevilacqua et al. (2018) stated these dynamic body movements as head movement and physical posture. The operator's head moves vertically, horizontally, and rotationally while operating. Thus, the increase in the mean value of this feature demonstrates that as the experiment progressed, the operators' head motion increased due to mental fatigue. Table 4.6 shows the change, which is statistically significant throughout all experiment stages, indicating greater mental fatigue. Similarly, the current study analyzed nose to jaw and nose to chin ratios. The preceding was to represent the face's shift to the right or left. The latter feature reflected the operator's face tilting upward or downward. The mean nose to chin ratio decreased from the baseline to the completion of the excavation experiment. It is because the operators were advancing towards the camera, but their faces were tilted upwards, indicating they were attempting to keep their focus on the task despite fatigue. However, the differences between the phases were not statistically significant. The present study's findings accord with past research in other sectors. For example, Liao et al. (2005b) and Dinges et al. (2005) found an increase in head movements during non-neutral states. Furthermore, studies by Kusano et al. (2020), Giannakakis et al. (2018), and Giannakakis et al. (2017) also reported an increased head motion under stressful situations such as watching videos. Nevertheless, results from the current study are contrary to the findings by Bevilacqua et al. (2018), where no statistical significance was reported between boring and stressful states.

Additionally, the current study also studied the face area feature which was associated with the movement of equipment operators towards and away from the camera. The current study demonstrated an increase in face area, indicating the movement of operators towards the camera. The increase between the subsequent experiment phases from baseline was 4.90%, 15.24%, and 11.89%, respectively. The findings are consistent with the previous study by Bevilacqua et al. (2018) where there was an increase in the face area of subjects during a stressful state.

5.3.2 Relationship of facial features’ geometric measurements with subjective and objective assessment

During the excavation operation experiment, there were strong relationships between geometric measurements of facial features and subjective mental fatigue scores. Some variables correlated with subjective scores throughout the entire experiment, while others only correlated at one or two stages. For example, face area features were substantially linked with subjective scores during the final two experiment phases, i.e., at 40 and 60 minutes, shown in Table 4.9. Previous studies have found that fatigue assessments are substantially connected to eye-related cues (Sundelin et al., 2013). Likewise, a study by Hopstaken et al. (2015) also reported an increase in subjective mental fatigue and a decrease in baseline pupil diameter as a result of increasing time spent on the activity, with a corresponding decrease in cognitive performance. Similarly, a study by Dziuda et al. (2021) also found that the drivers' responses to the fatigue symptoms scale questionnaire before and after the simulator task were found to be correlated with changes in their percentage closure of eye time levels.

This study found a difference between EEG bands (baseline, 20 min, 40 min, and 60 min) in the evolution of mental fatigue. After an hour of continuous operation of construction equipment, we found alterations in spontaneous brain activity. Five EEG patterns were evaluated in four brain areas: AF7, AF8, TP9, and TP10. Figure 4.8 shows the brain maps using the power spectral density of EEG data from construction equipment operators at the outset, 20 minutes, 40 minutes, and 60 minutes of the experiment. The beta band's power covers the entire brain. The temporal delta and gamma bands revealed a consistent trend. The frontal alpha band rhythm was not monotonous. Figure 4.8 depicts the frontal and temporal lobes of the brain becoming fatigued as the experiment progressed. In some areas, the theta band colors are redder and bluer. The p-values for statistical significance are also monotonous. The findings are consistent with previous research on fatigue (Li et al., 2020b, Eoh et al., 2005). Theta waves, which are linked to brain fatigue, appear early in the sleep cycle, making them sensitive to mental fatigue (Lal and Craig, 2005, Åkerstedt and Gillberg, 1990). Alpha rhythm indicates the condition of relaxation and wakefulness (Li et al., 2020b). In the third and fourth experiment phases of the study, alpha activity was observed in the frontal channels shown, in Figure 4.8, which is in line with previous research. For example, studies by Eoh et al. (2005) and Lal and Craig (2002) reported that the potency of the alpha pattern increases with an increase in mental fatigue. Similarly, another study by Sun et al. (2014) and Craig et al. (2012) also reported that with an increase in mental fatigue, the power of the alpha band increases. This is why it is considered the most reliable indication of mental fatigue (Lal and Craig, 2005). During the excavation operation experiment, there were strong relationships between geometric measurements of facial features and EEG metric. Some variables corresponded with subjective scores throughout all experiment phases, while others correlated at one or two stages only. For example, eye aspect ratio, eye distance, and head motion were substantially linked to the EEG metric during all the experiment phases. As the construction equipment operators were subjected to mental fatigue, their eye aspect ratio decreased, and their eye distance increased from the baseline. The decrease in eye aspect ratio indicates the closing of eyes, thus indicating theta band activity in the brain topography. Likewise, the increase in face area and head motion indicates that the equipment operators were trying to increase their concentration by moving close to the windscreen of the equipment and camera. However, the association of the rest of the facial features with the EEG metric was not found to be monotonous during each of the experiment phases. Overall, geometric measures of facial features produce statistical conclusions that agree with the visual representations of the brain as a result.

5.4 Discussion related to deep learning networks and EEG data for mental fatigue.

The study assessed a novel approach for recognizing and classifying different types of mental fatigue states in equipment operation using deep learning-based networks and wearable EEG data gathered from equipment operators. Subsequently, the study compared three types of deep learning models for training time-series raw EEG data acquired by a wearable headband: long short-term memory (LSTM), bidirectional long short-term memory (bi-LSTM), and one-dimensional convolutional network (1D-1 CNN). To the best of our knowledge, this study is the first to propose a deep learning-based model for recognizing and classifying alert, mild fatigue, and fatigue states from EEG signals in construction equipment operators under sustained attention. The results indicated that mental fatigue can be accurately classified in construction equipment operators with varying mental fatigue levels, i.e., alert state, mild fatigue state, and fatigue state.

5.4.1 Comparison of three deep learning models for mental fatigue classification

Comparing the deep learning models utilized in this study revealed that the bidirectional LSTM model had the highest accuracy of 99.941%. In addition, the results demonstrate that the bidirectional LSTM model achieved precision, recall, specificity, and F1-score metrics ranging from 99.840% to 99.995%, 99.839% to 99.997%, 99.914% to 99.997%, and 99.917% to 99.993%, respectively, when classifying multiple states of mental fatigue in construction equipment operators. Regarding the confusion matrix, it was concluded that the fatigue state (FS) and mild fatigue state (MFS) had the fewest misclassified instances, i.e., 34 and 35, respectively. While alert state (AS) was the most misclassified class, with 1164 instances of misclassification. In addition, the performance of the Bi-LSTM model increases in accuracy and decreases in loss during both training and testing. These findings indicate that the Bi-LSTM model, which is a deep learning network model, could provide a more accurate classification of the mental fatigue states of operators. This finding might be explained from the perspective of the model. Bidirectional LSTM can remember previous time-series patterns and is more effective at processing time-series data. The Bi-LSTM model is a cyclic neural network comprised of two distinct LSTM networks that can collect information not only from past input but also from future input states. Consequently, the concept of bidirectional LSTM's design is to collect the characteristics of a time series at the current time while also having information about the past and the future, resulting in outstanding classification accuracy. In this study, the bidirectional LSTM model surpassed other deep learning-based models in terms of classification accuracy. Similar to previous research, our findings indicate that the bidirectional LSTM model is better than unidirectional LSTM models (Li et al., 2020e), while bidirectional LSTM performs better with time-series data (Siami-Namini et al., 2019). Similarly, Sarkar et al. (2022) reported a high performance of the LSTM architecture compared to CNN when the models were applied to a testing dataset with varied percentages such as 20%, 30%, and 40%. The study established that when dealing with time-series EEG data, LSTM was found to perform better than a convolutional neural network, stating that CNN is more useful for image data. Similarly, Phutela et al. (2022) also reported that LSTM is a promising option for classifying stress-related brain activity data. Likewise, Rastgoo et al. (2019) also stated that long short-term memory has a strong ability to exploit the temporal dependencies in time-series data. Similarly, the results of the study by Cai et al. (2021) and our findings also demonstrated that the proposed approach, based on the three-layer Bi-LSTM prediction model, has a greater conception of the data exhibited in the forecast models, leading to more reliable forecasts of mental fatigue in construction equipment operators. The three-layer Bi-LSTM model learns from its errors during unsupervised training of the EEG-based brain activity patterns, to increase precision while maintaining the original attributes of the input EEG data. As a result, the forecast model of mental fatigue classification we developed in this study is more robust for its application in the construction industry. Hence, the use of bidirectional LSTM and LSTM models is recommended to classify mental fatigue states in construction equipment operators based on EEG data.

5.4.2 Comparison of current approach according to the published literature.

In Table 5.1, we contrast the performance of our approach with other methods found in the literature that are relevant to construction workers. Based on the comparison, it is evident that our classification method employs bidirectional LSTM and LSTM-based deep learning models, is better in performance. Previously, many studies had been conducted in construction to classify the stress or fatigue of construction workers, and some acceptable accuracy had been achieved. However, all these studies used manually crafted features from EEG data and applied machine learning to classify either stress or fatigue. Our approach is significantly different from previous machine learning approaches, where raw EEG data has been directly used without any manual crafting of input features. Although Jebelli et al. (2019c) have used deep learning neural networks to classify mental stress with an accuracy of 86.62%, such a study is significantly different in several ways. For example, their task was not a prolonged task because the focus of the author was to classify mental stress in construction workers. It was a simple task with a short duration that was performed by workers once standing on the ground (low stress) and then on the top of a ladder (high stress). Furthermore, such experimental settings are not suitable to induce mental fatigue in workers, particularly construction equipment operators. Also, it's worth noting that, because of variations in the experimental setup, the nature of tasks performed by operators, the number of subjects, the subjects themselves, etc., direct comparison with the methods is not possible or will be quiet challenging.

Table 5.1: Comparison of mental fatigue classification accuracies in construction domain

Reference	No. of subjects	No. of electrodes	Stress or Fatigue (Levels)	Stimulus (Type of data collection settings)	Classification Method	Accuracy (%)
(Aryal et al., 2017)	12	Beta 1 channel	Fatigue (4)	Psychomotor Vigilance Task (indoor simulated)	Boosted trees	82.60
(Jebelli et al., 2018a)	11	14	Stress (2)	Working on ladder (construction site)	Fully connected NN	79.26
(Jebelli et al., 2018c)	5	14	Stress (2)	Working on ladder (construction site)	OMTL-VonNeuman	77.60
(Jebelli et al., 2019b)	7	14	Stress (2)	Working on ladder (construction site)	Gaussian support vector machine	80.32
(Jeon and Cai, 2022)	30	16	Hazard (3)	Simulated environment. (Laboratory setting)	CatBoost LightGBM	65.2 63.7
Current study	15	4	Fatigue (3)	Construction Equipment Operation (Site)	Deep learning (Bi-LSTM)	99.9410
Current study	15	4	Fatigue (3)	Construction Equipment Operation (Site)	Deep learning (LSTM)	99.7076
Current study	15	4	Fatigue (3)	Construction Equipment Operation (Site)	Deep learning (1D CNN)	69.4726

5.5 Discussion related to multimodal integration for data-driven classification of mental fatigue.

Previous studies have used a single-modal data approach to detect and classify mental fatigue. However, it is unclear which physiological measure is a better indicator of mental fatigue. Thus, the objective of this study was to evaluate a new approach that uses machine learning and multimodal sensor data collected from equipment operators to recognize and classify different types of mental fatigue states during equipment operation. Three types of data from operators, i.e., electroencephalography, electrodermal activity, and geometric measurement of facial features, were gathered during an onsite operation on actual construction sites. The study then compared the performance of three types of machine learning models, including artificial neural networks (ANN), k-nearest neighbors (k-NN), and decision trees (DT), for training the input data collected from multiple sensors. To the best of the authors' knowledge, this study is the first to propose a machine learning-based approach for recognizing and classifying mental fatigue states, including alert, mild fatigue, and fatigue states, in construction equipment operators under sustained attention, by integration of multiple sensor data. The results show that mental fatigue can be accurately classified in construction equipment operators with varying levels of mental fatigue, i.e., alert state, mild fatigue state, and fatigue state, while integrating the acquired data from multiple sensors.

5.5.1 Multimodal data integration and machine learning-based models

In our study, we compared the performance of three machine learning models and found that the decision tree (DT) model outperformed the other two models, with an overall accuracy ranging from 85.0% to 97.1% when using different combinations of input data. The precision, recall specificity, and F1-score of the DT model ranged from 94.370% to 97.568%, 91.573% to 98.904%, 97.087% to 98.741%, and 94.084% to 98.231%, respectively, when integrating data from all sensors. The other two input combinations, EDA and FF, and EEG and FF, also showed high values of assessment metrics, with overall accuracy of 96.9% and 97.1%, respectively. Based on the analysis of the confusion matrix, it was observed that the alert state (AS), mild fatigue state (MFS), and fatigue state (FS) had a relatively small number of instances that were misclassified. For example, when using a combination of FF and EDA as input, the misclassified instances for AS, MFS, and FS were 4, 8, and 21, respectively. When using EEG and FF as input, the misclassified instances for AS, MFS, and FS were 11, 7, and 13, respectively. Similarly, when using EEG, EDA, and FF as input, the misclassified instances for AS, MFS, and FS were 9, 11, and 21, respectively. However, the confusion matrix revealed that the EEG and EDA combination had a larger number of misclassified instances compared to the other three combinations. Nonetheless, the misclassification rate was still lower than that of the ANN and k-NN models. Our findings indicate that the integration of multiple measures can be utilized to identify and categorize mental fatigue in operators during equipment operators.

5.5.2 Comparison with studies in non-construction domain

In this study, we used machine learning to recognize and categorize mental fatigue in equipment operators by integrating multiple types of data for the first time. Our findings indicate that, like in previous studies in other fields, combining data from various sources can be used to identify mental fatigue. However, our study performed better than the studies in the other domains, in terms of performance metrics. For instance, Ding et al. (2020) achieved an accuracy of 58.5% when using a fusion of ECG and EDA for classifying mental workload with neural networks, but combining all physiological measures as input data increased the accuracy to 78.3%. In another study by Xu et al. (2015), a combination of ECG, GSR, SpO2, electroencephalography, and electromyography was used to differentiate cognitive tasks and achieved an accuracy of 73.0% with support vector machines. Similarly, Hirachan et al. (2022) fused data from four sensors, including ECG and EDA, to distinguish cognitive workloads and achieved an accuracy of 74.0% with DT models. The DT model achieved an accuracy of 68.0% when using single-modal data. Additionally, Majid et al. (2022) found that combining data from multiple physiological modalities, such as electroencephalography, galvanic skin response, and photoplethysmography, increased perceived stress classification accuracy to 95.0% for two stress classes and 77.5% for three classes. Likewise, Jaiswal et al. (2022) utilized a fusion of input data from four sensors, namely EEG, ECG, EDA, and EMG, to detect cognitive fatigue and achieved an accuracy of 77.2% using a random forest model. While our study has achieved higher accuracy than previous studies in other domains, making an exact comparison is challenging due to differences in experimental protocols and the nature of tasks performed. Nevertheless, our findings suggest that this study has significant potential to improve mental fatigue assessment for construction operators and workers, which could help reduce the occurrence of injuries and accidents on construction sites.

5.5.3 Comparison with studies in construction industry

The integration of data from multiple sensors in our study yielded higher classification accuracy compared to previous studies in the construction domain that only used single-modal data. Table 5.2

shows a comparison of our approach with other relevant methods in the literature. Prior studies in construction had focused on classifying stress or fatigue using machine learning with single-modal data, achieving acceptable accuracy. In contrast, our approach is significantly different because we integrated input data from multiple sensors in various combinations for mental fatigue classification. For example, Jeon and Cai (2022) used a two-step ensemble approach to classify hazard recognition and cognitive states using single-modal EEG data and achieved 82.3% accuracy with the LightGBM classifier. Jebelli et al. (2019b) used the OMTL-Von Neumann method for stress recognition in construction workers and achieved 77.61% accuracy, while another study by Jebelli et al. (2018c) used non-linear support vector machines to classify construction worker stress with 71.1% accuracy using single-modal EEG data on a construction site. However, these studies differ from ours because they did not focus on prolonged tasks or mental fatigue in construction equipment operators. Additionally, direct comparison with these studies may be challenging due to variations in experimental setups, task nature, number of subjects, and subject differences.

Table 5.2: Comparison of classification accuracies in construction industry studies

Reference	No. of subjects	Mode(s)	Stress or Fatigue (Levels)	Stimulus (Data collection settings)	Classification Method	Accuracy (%)
Jeon and Cai (2022)	30	EEG	Hazard (3)	Simulated environment. (Laboratory setting)	LightGBM	82.3
Jebelli et al. (2018a)	11	EEG	Stress (2)	Working on ladder (Construction site)	Fully connected NN	79.26
Jebelli et al. (2018c)	8	EEG	Stress (2)	Working on ladder (Construction site)	Non-linear support vector machine	71.1
Aryal et al. (2017)	12	Beta 1 channel	Fatigue (4)	Psychomotor Vigilance Task (indoor simulated)	Boosted trees	82.60
Jebelli et al. (2019b)	5	EEG	Stress (2)	Working on ladder (construction site)	OMTL-VonNeuman	77.61
Current study	16	EEG, EDA and FF	Fatigue (3)	Excavation Operation (Construction Site)	ANN k-NN DT	94.7 85.8 96.2

5.6 Limitations and future research

Although this study has deepened the current knowledge regarding the detection of mental fatigue in construction equipment operators using computer vision-based facial feature detection, deep learning-based EEG sensor data and multimodal data fusion as an input, there remain limitations which should be acknowledged addressed in future studies.

(a) The sample size in this study was modest, and there were three mental fatigue levels. Despite the fact that we determined the sample size based on the sample sizes employed in previous research of a similar kind, findings with such a limited number of operators may limit the application of the proposed approach to the construction industry. To generalize the results to the entire population of operators, future studies should collect large data sets representing a variety of mental fatigue states.

(b) This study did not use single modal as well as multimodal data to establish thresholds for the different levels of mental fatigue. Future research may leverage these thresholds to recognize and classify mental fatigue states, depending on whether they can be established and applied to all construction equipment operators. Future studies may also treat mental fatigue identification as a regression problem and calculate the degree of mental fatigue.

(c) There may be many factors that can affect the changes in facial features at a real-world construction site. Future study directions should include in-depth investigations into how other characteristics, such as age, experience, and so on, affect facial feature-based mental fatigue detection during prolonged equipment operations.

(d) The study used machine-learning models and multimodal sensor data to categorize mental fatigue in equipment operators, but the features were manually crafted from various sensors and then combined for classification purposes. Future research should utilize deep learning techniques or a combination of multiple deep learning techniques and raw multimodal data to identify mental fatigue in operators without the manual crafting of features. Unsupervised learning may also be employed in future studies to learn the features related to operators' mental fatigue on unlabelled multimodal sensor data.

(e) The current study used only three types of deep learning networks to recognize and classify mental fatigue in construction equipment operators. Nevertheless, deep learning models based on bidirectional LSTM are intended primarily to handle sequence and time-series data. They do, however, come at a higher cost. They require more time to train the model. As a result, future research could combine multiple deep learning networks as a fusion model or use multi-deep learning models to classify mental fatigue in equipment operators.

(f) This study evaluated mental fatigue using only three levels. Future studies should assess performance using more classes of mental fatigue for a better understanding of mental fatigue in real time.

(g) The ground truth of mental fatigue was based on operators' subjective assessment, which may have been influenced by personal biases. Although the operators were familiar with the evaluation method of mental fatigue level, subjective assessment can still be considered a limitation owing to its lack of objectivity. However, it is a reliable technique for annotating data, despite its potential shortcomings.

(h) This study focused solely on excavation operators as equipment operators. Subsequent research should replicate these results for operators of different types of construction equipment, such as cranes, dozer, and grader operators. In general, it is crucial to collect a large dataset with sufficient samples from various groups of equipment operators to identify additional mental fatigue states that are essential for training, testing, and constructing a comprehensive model for construction operations.

(i) Lighting fluctuations are believed to have an impact on the geometric measurements of face feature detection (Tran et al., 2019, Lee et al., 2018). To avoid this, we ran the experiments on the construction site at the same time each day for the subsequent days under similar weather conditions. However, the future studies should acquire data at various times throughout the day such as morning and evening, under varying weather conditions, to see how fluctuations in lighting on construction sites affect the results connected to facial feature geometric measures and mental fatigue monitoring.

(j) Privacy concerns are valid and real concerns when implementing a system for mental fatigue recognition among construction workers. While the system proposed in the current study holds the potential to revolutionize occupational health by enabling real-time monitoring and proactive interventions, there are also inherent risks related to device hacking, data breaches, privacy issues, and data mismanagement. However, addressing such concerns is beyond the scope of this study. Future studies should focus on evaluating and mitigating these risks to ensure the successful on-field deployment of such a system. This involves carefully assessing the system architecture and hardware to ensure robustness against privacy and data security concerns. Measures should be implemented to secure the collected data, including encryption during transmission and storage, to prevent unauthorized access. Obtaining informed consent from the construction workers is essential. They should be informed of the purpose of data collection and how it will be used solely for on-site safety management. Transparency in the process helps build trust and confidence among 2 workers regarding the protection and security of their data.

(k) This study doesn't acquire any feedback from the experts and the safety personnel. The reason was that the focus of current study was to propose and study the feasibility of automated and non-invasive method to assess mental fatigue in equipment operators. However, future studies should collect such feedback regarding the validity and usability of the results and the feasibility of implementing the proposed approach in real construction projects.

5.7 Summary

The chapter discussed the research findings of the current study. It was revealed that the changes in geometric measurement of facial features were in line with the results in other industries. Furthermore, the results were compared with the other studies in construction industry. For instance, the performance of bi-directional technique was high in accuracy as compared to the performance of machine learning models using EEG sensor data. Likewise, the findings also indicated the performance of machine learning models using multimodal data fusion was better than previously studied single modal data analysis in the construction industry. Lastly, the current study is also subjected to some limitations which can be studied in future studies for mental fatigue assessment in construction workers. In addition, future research has been suggested.

Chapter 6

Conclusions and Contributions[9]

6.1 Introduction

This section summarizes the findings of this study and highlights their importance and value. In addition, this study suggests potential avenues for future research.

6.2 Summary of research findings

The aim of this research is to explore the possibility of using the geometric measurement of facial features as a non-invasive method for assessing mental fatigue in construction equipment operators. The following are the specific research objectives: (1) to study non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features; (2) to investigate the validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators; (3) to explore the use of deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data; (4) to study the multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals. The subsequent subsections provide a summary of the findings of various studies conducted as part of this research.

[9] This chapter is based on research published and reproduced with permission from Elsevier.
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Muhammad Saad Shakeel, Shahnawaz Anwer (2022) "Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators" *Advanced Engineering Informatics*, Volume 54, 101777
Imran Mehmood, Heng Li, Yazan Qarout, Waleed Umer, Shahnawaz Anwer, Haitao Wu, Mudasir Hussain, Maxwell Fordjour Antwi-Afari (2023) "Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data". *Advanced Engineering Informatics*, Volume 56, 101978
Imran Mehmood, Heng Li, Waleed Umer, Aamir Arsalan, Shahnawaz Answer, Mohammed Aquil Mirza, Jie Ma, Maxwell Fordjour Antwi-Afari (2023) "Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals". *Developments in the Built Environment,* Volume 15, 100198
Imran Mehmood, Heng Li, Waleed Umer, Jie Ma, Muhammad Saad Shakeel, Shahnawaz Anwer, Maxwell Fordjour Antwi-Afari, Salman Tariq, Haitao Wu (2024) "Non-invasive monitoring of mental fatigue in construction equipment operators' using their geometric measurement of facial features". *Journal of Safety Research*, https://doi.org/10.1016/j.jsr.2024.01.013, JSR2291

6.2.1 Non-invasive detection of mental fatigue in construction equipment operators through geometric measurements of facial features

To mitigate the risk of construction equipment accidents resulting from operator inattention, this study introduces a construction site process for detecting mental fatigue in construction equipment operators. Computer vision techniques were employed to analyse the video recordings of operators at real construction sites, focusing on geometric measurements of facial features to monitor their mental fatigue. Seventeen excavator operators participated in the study, and facial videos were collected during the excavation tasks. Six distinct facial features, including Euclidean distances and areas, were calculated using sixty-eight facial landmarks. The results demonstrated statistically significant differences in the mean values of all facial features, such as eye area, eyebrow, face area, head motion, mouth outer, and mouth corners. Notably, the eye and facial area measurements exhibited the most significant variations, with mean differences of 45.88% and 26.12%, respectively. Specifically, the study found that the mean value of eye area was 0.2949 pixels^2 for low mental fatigue and 0.4302 pixels^2 for high mental fatigue. Statistical analysis indicated that the median eye area was significantly greater for high mental fatigue than for low mental fatigue, with p-values below 0.01 and an effect size (η^2) of 0.801. Similarly, the findings revealed that the mean face area values for low and high mental fatigue were 9.2141 and 11.6928 pixels^2, respectively. Statistical analysis confirmed that the median face area was significantly larger for high mental fatigue than for low mental fatigue, with p-values below 0.01 and an effect size (η^2) of 0.726. The key contribution of this study is the demonstration of contactless measurement as a promising approach for detecting and evaluating mental fatigue in construction equipment operators. By enabling proactive identification of operator fatigue, particularly in the complex context of construction equipment operations, contactless measurement methods enhance our understanding of mental fatigue and help mitigate the risk of fatigue-related errors and illnesses.

6.2.2 Validity of facial features' geometric measurements for a real-time assessment of mental fatigue in construction equipment operators

This study developed a construction site procedure for detecting mental fatigue in construction equipment operators in order to reduce the risk of equipment-related accidents. Excavator operators were involved in recording facial videos and EEG sensor data while performing excavation tasks. Eight facial features, namely eye aspect ratio, eye distance, eyebrows, mouth aspect ratio, nose-to-jaw ratio, nose-to-chin ratio, face area, and head motion, were calculated based on the Euclidean distance and areas derived from sixty-eight facial landmarks. The geometric measurements of these facial features and EEG sensor data were compared at different time intervals: baseline, 20 min, 40 min, and 60 min. The results revealed statistically significant differences in the mean values of several facial features, including eye aspect ratio, eye distance, eye distance, mouth aspect ratio, face area, and head motion, between the different phases of the experiment. However, no statistically significant differences were found in the remaining facial features. Furthermore, the brain maps generated from the power spectral density of the recorded EEG data supported the presence of mental fatigue in the operators' brains at the corresponding time frames. The key contribution of this study is its demonstration of the ecological validity of contactless measures for detecting and evaluating mental fatigue in construction equipment operators by establishing their association with wearable EEG sensor data. The proposed method, which is non-invasive and based on video records, eliminates the need for wearable sensing technology for mental fatigue monitoring. Given the dynamic and complex nature of construction site operations, this methodology is deemed more user-friendly, practical, and suitable for mental fatigue monitoring in the construction industry. Its implementation has the potential to reduce equipment-related accidents, injuries, and errors at construction sites by proactively monitoring operators' mental fatigue.

6.2.3 Deep learning-based construction equipment operators’ mental fatigue classification using wearable EEG sensor data.

The objective of this study was to evaluate a novel approach that utilizes deep learning-based networks and EEG sensor data to distinguish and classify different mental fatigue states. Subjective assessments were conducted to label three mental fatigue states: alert, mild fatigue, and fatigue. The brain activity patterns of 15 equipment operators were then captured using a wearable headband EEG sensor while they performed a monotonous and prolonged excavation task at a real construction site. The performance of the three deep learning models (LSTM, bidirectional LSTM, and 1D-CNN) was evaluated using accuracy, precision, recall, specificity, and F1-score as metrics for classification performance. The statistical significance of the results obtained from the three deep learning models was assessed using the Mann-Whitney test. The experimental findings revealed that the bidirectional

LSTM model outperformed the other deep learning models, achieving an accuracy of 99.941% and an F1-score ranging from 99.917% to 99.993%. Both the bidirectional LSTM and LSTM models outperformed the 1D-CNN models, although the difference in their accuracies was less than 1%. These findings support the effectiveness of using the bidirectional LSTM model, which is commonly used for time-series and sequential data classification, to learn sequential brain activity patterns captured by an EEG sensor to distinguish and classify different mental fatigue states during construction operations. Moreover, this approach can contribute to the development of real-time wearable EEG sensor computing by leveraging the brain activity pattern performance and a bidirectional LSTM model for the classification of different mental fatigue states. Additionally, it will enhance safety and health management at construction sites by enabling safety managers to continuously monitor the real-time mental fatigue levels of construction equipment operators.

6.2.4 Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: incorporating electroencephalography, electrodermal activity, and video signals.

This study presented a novel approach for effectively classifying mental fatigue levels in construction equipment operators by leveraging supervised machine learning and the fusion of multimodal sensor data. Sixteen equipment operators participated in an excavation task on a construction site, and their mental fatigue was subjectively assessed using NASA-TLX as a reference. Throughout the experiment, simultaneous EEG and EDA measurements were conducted using wearable Muse headbands and E4 watches, respectively, while video recordings captured geometric measurements of the facial features. The monotonous and prolonged excavation task induced mental fatigue. Following the experiment, the features were extracted and integrated from multiple sensors to create input data. Three supervised machine learning models, namely artificial neural network (ANN), k-nearest neighbours (k-NN), and decision tree (DT), were employed, along with four combinations of multimodal data, to classify the three levels of mental fatigue: alert, mild fatigue, and fatigue. The performance of these models was evaluated using various assessment metrics including accuracy, precision, recall, specificity, and F1-score. The experimental results demonstrate that the DT model outperformed the other models across all combinations of multimodal data, achieving an overall accuracy of 96.9% (FF and EDA), 85.0%

(EEG and EDA), 97.1% (EEG and EDA), and 96.2% (EEG, EDA, and FF). Although the overall accuracy of the ANN and k-NN models in this study was slightly lower than that of the DT model, their performance still surpassed those of previous studies conducted with single-modal data. These findings support the utilization of the DT model and fusion of data from multiple sensors to accurately classify mental fatigue states during construction equipment operations. This approach contributes to the development of a unified real-time system that combines multiple sensors and machine learning for the classification of mental fatigue in operators. Furthermore, the implementation of such a system will enhance safety and health management at construction sites by enabling safety managers to monitor the mental fatigue levels of operators in real-time, thereby reducing the occurrence of injuries and accidents.

6.3 Significance and contributions

The findings of this study make the following significant contributions:

(a) This study demonstrated the feasibility of utilizing changes in the geometric measurements of facial features to detect mental fatigue in excavator operators during extended operations at construction sites. By collecting data from actual construction environments, this study provides valuable insights that can contribute to the implementation of a real-time system for detecting fatigue in operators based on their facial features.

(b) In this study, a nonintrusive method was employed to detect fatigue by utilizing contactless measurements of facial features. This innovative approach deviates from the traditional methods that rely on wearable sensors. By implementing such a system, safety managers can take proactive measures to prevent errors and accidents caused by mental fatigue-induced lapses in operators' attention. This novel approach has great potential to enhance safety and prevent workplace incidents.

(c) The findings of this study provide valuable insights that can aid construction managers in developing a framework for effectively managing worker shifts. Researchers have examined the changes in facial features and brain activity of construction equipment operators across three levels of mental fatigue over a period of one hour. Construction managers can leverage this information by periodically monitoring operators, ideally every 30–45 min. By introducing breaks between shifts, operators have the opportunity to recover from the mental fatigue they have accumulated, thereby promoting their overall well-being and reducing the risk of errors or accidents.

(d) The accumulation of mental fatigue among equipment operators in real construction settings may occur at a faster rate than that in controlled laboratory settings. This study recognizes this difference and utilizes data collected from actual construction sites to enhance the ecological validity of the findings. The results provide strong support for the effectiveness of using geometric measurements of facial features as a contactless approach to manage mental fatigue among construction equipment operators. This opens new possibilities for implementing practical and reliable strategies to address mental fatigue in the construction industry.

(e) The study included eye-related, mouth-related, and head-related facial features in the investigations to assess mental fatigue. Based on the findings, the ranking order of facial features in terms of their relevance to mental fatigue is head-related features, eye-related features and mouth related features. The head-related features were highly related with the ground truth of mental fatigue, followed by eye-related facial features. This information can be used in future studies for further in-depth analysis and well as on-site applications.

(f) The proposed method offers a valuable solution for recognizing and categorizing the mental fatigue states in construction equipment operators. By utilizing EEG data and deep learning networks, it is possible to identify and classify mental fatigue that can lead to attentional lapses among these operators. Detecting mental fatigue is a crucial initial step towards proactively preventing attention failure. Consequently, this approach, based on EEG data and deep learning networks, serves as an effective tool for intervention, enabling tracking and identification of various states of brain fatigue in operators. By doing so, it effectively reduces incidents caused by mental fatigue at construction sites and helps minimize accidents. Furthermore, this method holds promise for supporting workers in various occupations within the construction industry, such as monitoring the mental states of structural design engineers, who frequently encounter demanding tasks requiring sustained attention and multiple redesigns within tight timeframes.

(g) The findings of this study hold great significance for researchers in the construction industry. Mental fatigue is a common occurrence resulting from prolonged equipment operations. The method presented in this study is not limited to excavation operations alone, which involve tasks such as earth excavation and material transfer using buckets and trucks. It can also be applied to other repetitive and protracted equipment operations in the construction industry, including cranes. By recognizing the prevalence of mental fatigue in various equipment-related tasks, this method offers valuable insights and potential applications beyond excavation, opening up possibilities for improved fatigue management in a wide range of construction operations.

(h) The approach presented in this study is promising for advancing the development of real-time wearable EEG sensor computing. Leveraging the performance of brain activity pattern analysis and employing a bidirectional LSTM model enables the classification of different states of mental fatigue in construction equipment operators. This breakthrough has significant implications for workplace safety managers, who can utilize these data to enhance the protection and well-being of their workers. The results of this study highlight the effectiveness of deep learning models, particularly Bi-LSTM and LSTM, in learning and predicting mental fatigue states based on the brain activity patterns of equipment operators. The high accuracy achieved by all three deep learning models in this study demonstrated their potential for reliable mental fatigue classification. However, it should be noted that the misclassification of mild fatigue states was more common compared to other fatigue states, indicating potential challenges in accurately identifying this particular state. Nevertheless, the findings of this study have broader implications than mental fatigue classification. The same approach can be extended to address other cognitive failures such as mental stress, mental workload, hazard identification, and emotions. By leveraging the insights gained from this study, incident management strategies can be improved to better support construction workers facing cognitive issues. This advancement has the potential to enhance the overall safety and well-being of the construction industry by mitigating the risks associated with cognitive impairment.

(i) In contrast to previous studies that relied on single-modal data for detecting mental stress or fatigue, this study presents a significant breakthrough by showing the effectiveness of data fusion from multiple sensors in accurately classifying the mental fatigue levels of construction operators. These findings have substantial implications, indicating that practitioners and researchers can leverage a unified system equipped with multiple sensors to detect and monitor mental fatigue among equipment operators. By integrating data from various sensors such as EEG, facial recognition, and physiological sensors, a comprehensive picture of an operator's mental fatigue can be obtained. This holistic approach enables a more accurate and nuanced assessment of mental fatigue, allowing timely intervention and proactive management of operator well-being. The use of multiple sensors in a single system enhances the precision and reliability of mental fatigue detection, thereby enabling a more comprehensive understanding of an operator's cognitive state. The findings of this study will pave the way for the development and implementation of advanced monitoring systems that utilize data fusion from multiple sensors. Such systems hold immense potential for enhancing workplace safety and productivity by providing real-time insights into the mental fatigue levels of construction operators. By leveraging this integrated approach, practitioners and researchers can make informed decisions, implement appropriate interventions, and optimize work schedules to mitigate the risks associated with mental fatigue in construction operations.

(j) The findings of this study highlight the practicality of utilizing wearable electroencephalography (EEG), electrodermal activity (EDA) sensors, and a mobile camera to gather on-site experimental data for detecting mental fatigue. These insights have significant implications for real-time management of fatigue among construction workers. The findings of this study have significant practical implications. The ability to collect real-time data using wearable devices and mobile cameras opens up new opportunities for fatigue management in construction settings. This information can be used to develop proactive strategies for fatigue prevention and intervention to ensure the well-being and safety of construction workers.

6.4 Framework for applying research outcomes.

This section presents a framework that describes a systematic, step-by-step approach for applying the research outcomes on construction sites to assess mental fatigue in construction equipment operators.

Figure 6.1: Framework to apply research outcomes.

6.5 Chapter summary

In conclusion, this chapter brings together the findings from multiple investigations, offering a comprehensive understanding of how to measure mental fatigue among construction workers and equipment operators. The utilization of wearable sensors and mobile cameras for the real-time monitoring of physiological parameters and facial features provides a promising avenue for accurately assessing mental fatigue. These insights are intended to inspire further research in this field and foster ongoing advancements in fatigue measurement and management practices.

References

ABD RAHMAN, F. & OTHMAN, M. F. Real Time Eye Blink Artifacts Removal in Electroencephalogram Using Savitzky-Golay Referenced Adaptive Filtering. *In:* IBRAHIM, F., USMAN, J., MOHKTAR, M. S. & AHMAD, M. Y., eds. International Conference for Innovation in Biomedical Engineering and Life Sciences, 2016// 2016 Singapore. Springer Singapore, 68-71.

ABDELJABER, O., SASSI, S., AVCI, O., KIRANYAZ, S., IBRAHIM, A. A. & GABBOUJ, M. 2018. Fault detection and severity identification of ball bearings by online condition monitoring. *IEEE Transactions on Industrial Electronics,* 66**,** 8136-8147.

ADANE, M. M., GELAYE, K. A., BEYERA, G. K., SHARMA, H. R. & YALEW, W. W. 2013. Occupational injuries among building construction workers in Gondar City, Ethiopia. *Occupational Medicine & Health Affairs*.

AHN, C. R., LEE, S., SUN, C., JEBELLI, H., YANG, K. & CHOI, B. 2019. Wearable Sensing Technology Applications in Construction Safety and Health. *Journal of Construction Engineering and Management,* 145**,** 03119007.

AHN, S., NGUYEN, T., JANG, H., KIM, J. G. & JUN, S. C. 2016. Exploring Neuro-Physiological Correlates of Drivers' Mental Fatigue Caused by Sleep Deprivation Using Simultaneous EEG, ECG, and fNIRS Data. *Frontiers in Human Neuroscience,* 10.

ÅKERSTEDT, T. & GILLBERG, M. 1990. Subjective and objective sleepiness in the active individual. *International journal of neuroscience,* 52**,** 29-37.

ALBERT, A., PANDIT, B., PATIL, Y. & LOUIS, J. 2020. Does the potential safety risk affect whether particular construction hazards are recognized or not? *Journal of safety research,* 75**,** 241-250.

ALGHADIR, A. & ANWER, S. 2015. Prevalence of musculoskeletal pain in construction workers in Saudi Arabia. *The Scientific World Journal,* 2015.

ALLISON, R. W., HON, C. K. & XIA, B. 2019. Construction accidents in Australia: Evaluating the true costs. *Safety Science,* 120**,** 886-896.

ANSARI, S., NAGHDY, F., DU, H. & PAHNWAR, Y. N. 2022. Driver Mental Fatigue Detection Based on Head Posture Using New Modified reLU-BiLSTM Deep Neural Network. *IEEE Transactions on Intelligent Transportation Systems,* 23**,** 10957-10969.

ANTWI-AFARI, M. F., ANWER, S., UMER, W., MI, H.-Y., YU, Y., MOON, S. & HOSSAIN, M. U. 2023. Machine learning-based identification and classification of physical fatigue levels: A novel method based on a wearable insole device. *International Journal of Industrial Ergonomics,* 93**,** 103404.

ANTWI-AFARI, M. F., QAROUT, Y., HERZALLAH, R., ANWER, S., UMER, W., ZHANG, Y. & MANU, P. 2022. Deep learning-based networks for automated recognition and classification of awkward working postures in construction using wearable insole sensor data. *Automation in Construction,* 136**,** 104181.

ANWER, S., LI, H., ANTWI-AFARI, M. F. & WONG, A. Y. L. 2021. Associations between physical or psychosocial risk factors and work-related musculoskeletal disorders in construction workers based on literature in the last 20 years: A systematic review. *International Journal of Industrial Ergonomics,* 83**,** 103113.

ARAVIND, A., AGARWAL, A., JAISWAL, A., PANJIYARA, A. & SHASTRY, M. 2019. Fatigue detection system based on eye blinks of drivers. *Int. J. Eng. Adv. Technol,* 8**,** 72-75.

ARSALAN, A., MAJID, M., BUTT, A. R. & ANWAR, S. M. 2019. Classification of Perceived Mental Stress Using A Commercially Available EEG Headband. *IEEE Journal of Biomedical and Health Informatics,* 23**,** 2257-2264.

ARYAL, A., GHAHRAMANI, A. & BECERIK-GERBER, B. 2017. Monitoring fatigue in construction workers using physiological measurements. *Automation in Construction,* 82**,** 154-165.

ATTAL, F., MOHAMMED, S., DEDABRISHVILI, M., CHAMROUKHI, F., OUKHELLOU, L. & AMIRAT, Y. 2015. Physical human activity recognition using wearable sensors. *Sensors,* 15**,** 31314-31338.

BACHURINA, V. & ARSALIDOU, M. 2022. Multiple levels of mental attentional demand modulate peak saccade velocity and blink rate. *Heliyon,* 8**,** e08826.

BAI, X.-P. & QIAN, C. 2021. Factor validity and reliability performance analysis of human behavior in green architecture construction engineering. *Ain Shams Engineering Journal,* 12**,** 4291-4296.
BALTRUŠAITIS, T., ROBINSON, P. & MORENCY, L. OpenFace: An open source facial behavior analysis toolkit. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 7-10 March 2016 2016. 1-10.
BEHRENS, M., GUBE, M., CHAABENE, H., PRIESKE, O., ZENON, A., BROSCHEID, K.-C., SCHEGA, L., HUSMANN, F. & WEIPPERT, M. 2023. Fatigue and Human Performance: An Updated Framework. *Sports Medicine,* 53**,** 7-31.
BEVILACQUA, F., BACKLUND, P. & ENGSTROM, H. Variations of Facial Actions While Playing Games with Inducing Boredom and Stress. 2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), 7-9 Sept. 2016 2016. 1-8.
BEVILACQUA, F., ENGSTRÖM, H. & BACKLUND, P. 2018. Automated Analysis of Facial Cues from Videos as a Potential Method for Differentiating Stress and Boredom of Players in Games. *International Journal of Computer Games Technology,* 2018**,** 8734540.
BIRHANE, G. E., YANG, L., GENG, J. & ZHU, J. 2022. Causes of construction injuries: a review. *International Journal of Occupational Safety and Ergonomics,* 28**,** 343-353.
BITKINA, O. V., PARK, J. & KIM, H. K. 2021. The ability of eye-tracking metrics to classify and predict the perceived driving workload. *International Journal of Industrial Ergonomics,* 86**,** 103193.
BOKSEM, M. A. S. & TOPS, M. 2008. Mental fatigue: Costs and benefits. *Brain Research Reviews,* 59**,** 125-139.
BORGHINI, G., VECCHIATO, G., TOPPI, J., ASTOLFI, L., MAGLIONE, A., ISABELLA, R., CALTAGIRONE, C., KONG, W., WEI, D. & ZHOU, Z. Assessment of mental fatigue during car driving by using high resolution EEG activity and neurophysiologic indices. 2012 annual international conference of the IEEE engineering in medicine and biology society, 2012. IEEE, 6442-6445.
BORRAGÁN, G., SLAMA, H., DESTREBECQZ, A. & PEIGNEUX, P. 2016. Cognitive fatigue facilitates procedural sequence learning. *Frontiers in human neuroscience,* 10**,** 86.
BOUCSEIN, W. 2012. *Electrodermal activity*, Springer Science & Business Media.
BRAITHWAITE, J. 2013. A Guide for Analysing Electrodermal Activity & Skin Conductance Responses for Psychological Experiments/J. Jason Braithwaite, Derrick G Watson, Robert Jones, Mickey Rowe.–Selective Attention & Awareness Laboratory Behavioural Brain Sciences Centre, University of Birmingham. UK: Tech. Rep.
BROWN, I. D. 1994. Driver fatigue. *Human factors,* 36**,** 298-314.
BUCSUHÁZY, K., MATUCHOVÁ, E., ZŮVALA, R., MORAVCOVÁ, P., KOSTÍKOVÁ, M. & MIKULEC, R. 2020. Human factors contributing to the road traffic accident occurrence. *Transportation research procedia,* 45**,** 555-561.
BYERS, J. C. 1989. Traditional and raw task load index (TLX) correlations: are paired comparisons necessary? *Advances in Industrial Erfonomics and Safety l: Taylor and Francis.*
CAI, C., TAO, Y., ZHU, T. & DENG, Z. 2021. Short-Term Load Forecasting Based on Deep Learning Bidirectional LSTM Neural Network. *Applied Sciences,* 11**,** 8129.
CANNARD, C., WAHBEH, H. & DELORME, A. Validating the wearable MUSE headset for EEG spectral analysis and Frontal Alpha Asymmetry. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021. IEEE, 3603-3610.
CARBONARI, A., GIRETTI, A. & NATICCHIA, B. 2011. A proactive system for real-time safety management in construction sites. *Automation in construction,* 20**,** 686-698.
CECH, J. & SOUKUPOVA, T. 2016. Real-time eye blink detection using facial landmarks. *Cent. Mach. Perception, Dep. Cybern. Fac. Electr. Eng. Czech Tech. Univ. Prague***,** 1-8.
CESA-BIANCHI, N. & ORABONA, F. 2021. Online Learning Algorithms. *Annual Review of Statistics and Its Application,* 8**,** 165-190.
CHAE, J., HWANG, S., SEO, W. & KANG, Y. 2021. Relationship between rework of engineering drawing tasks and stress level measured from physiological signals. *Automation in Construction,* 124**,** 103560.

CHAERUN NISA, E. & KUAN, Y.-D. 2021. Comparative Assessment to Predict and Forecast Water-Cooled Chiller Power Consumption Using Machine Learning and Deep Learning Algorithms. *Sustainability,* 13**,** 744.
CHAN, M. 2011. Fatigue: The most critical accident risk in oil and gas construction. *Construction Management and Economics,* 29**,** 341-353.
CHARLES, R. L. & NIXON, J. 2019. Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics,* 74**,** 221-232.
CHAVAILLAZ, A., SCHWANINGER, A., MICHEL, S. & SAUER, J. 2019. Work design for airport security officers: Effects of rest break schedules and adaptable automation. *Applied ergonomics,* 79**,** 66-75.
CHEN, C.-F. & HSU, Y.-C. 2020. Taking a closer look at bus driver emotional exhaustion and well-being: evidence from Taiwanese urban bus drivers. *Safety and Health at Work,* 11**,** 353-360.
CHEN, J., TAYLOR, J. E. & COMU, S. 2017a. Assessing task mental workload in construction projects: A novel electroencephalography approach. *Journal of Construction Engineering and Management,* 143**,** 04017053.
CHEN, Y., LU, B., CHEN, Y. & FENG, X. 2015. Breathable and Stretchable Temperature Sensors Inspired by Skin. *Scientific Reports,* 5**,** 11505.
CHEN, Y., MCCABE, B. & HYATT, D. 2017b. Impact of individual resilience and safety climate on safety performance and psychological stress of construction workers: A case study of the Ontario construction industry. *Journal of safety research,* 61**,** 167-176.
CHENG, B., FAN, C., FU, H., HUANG, J., CHEN, H. & LUO, X. 2022. Measuring and Computing Cognitive Statuses of Construction Workers Based on Electroencephalogram: A Critical Review. *IEEE Transactions on Computational Social Systems***,** 1-16.
CHENG, Q., WANG, W., JIANG, X., HOU, S. & QIN, Y. 2019. Assessment of Driver Mental Fatigue Using Facial Landmarks. *IEEE Access,* 7**,** 150423-150434.
CHEW, J. Y., KAWAMOTO, M., OKUMA, T., YOSHIDA, E. & KATO, N. 2021. Multi-modal approach to evaluate adaptive visual stimuli of remote operation system using gaze behavior. *International Journal of Industrial Ergonomics,* 86**,** 103223.
CHOI, B., JEBELLI, H. & LEE, S. 2019. Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers' perceived risk. *Safety Science,* 115**,** 110-120.
CHOI, B. & LEE, S. 2017. Role of social norms and social identifications in safety behavior of construction workers. II: Group analyses for the effects of cultural backgrounds and organizational structures on social influence process. *Journal of Construction Engineering and Management,* 143**,** 04016125.
CHOI, J., GU, B., CHIN, S. & LEE, J.-S. 2020. Machine learning predictive model based on national data for fatal accidents of construction workers. *Automation in Construction,* 110**,** 102974.
CLB. 2020. "China Labour Bulletin - Worker Safety" available at: https://clb.org.hk/content/work-safety (accessed on 14 August 2022).
COLLET, C., SALVIA, E. & PETIT-BOULANGER, C. 2014. Measuring workload with electrodermal activity during common braking actions. *Ergonomics,* 57**,** 886-896.
CRAIG, A., TRAN, Y., WIJESURIYA, N. & NGUYEN, H. 2012. Regional brain wave activity changes associated with fatigue. *Psychophysiology,* 49**,** 574-582.
CRAIK, A., HE, Y. & CONTRERAS-VIDAL, J. L. 2019. Deep learning for electroencephalogram (EEG) classification tasks: a review. *Journal of neural engineering,* 16**,** 031001.
DAS, S., MAITI, J. & KRISHNA, O. B. 2020. Assessing mental workload in virtual reality based EOT crane operations: A multi-measure approach. *International Journal of Industrial Ergonomics,* 80**,** 103017.
DASARI, D., CROWE, C., LING, C., ZHU, M. & DING, L. EEG pattern analysis for physiological indicators of mental fatigue in simulated air traffic control tasks. Proceedings of the human factors and ergonomics society annual meeting, 2010. SAGE Publications Sage CA: Los Angeles, CA, 205-209.
DASARI, D., SHOU, G. & DING, L. 2013. EEG index for time-on-task mental fatigue in real air traffic controllers obtained via independent component analysis.

DAZA, I. G., BERGASA, L. M., BRONTE, S., YEBES, J. J., ALMAZÁN, J. & ARROYO, R. 2014. Fusion of optimized indicators from Advanced Driver Assistance Systems (ADAS) for driver drowsiness detection. *Sensors,* 14**,** 1106-1131.
DEL SAVIO, A., LUNA, A., CÁRDENAS-SALAS, D., VERGARA, M. & URDAY, G. 2022. Dataset of manually classified images obtained from a construction site. *Data in Brief,* 42**,** 108042.
DEMŠAR, J., CURK, T., ERJAVEC, A., GORUP, Č., HOČEVAR, T., MILUTINOVIČ, M., MOŽINA, M., POLAJNAR, M., TOPLAK, M. & STARIČ, A. 2013. Orange: data mining toolbox in Python. *the Journal of machine Learning research,* 14**,** 2349-2353.
DESMOND, P. A. & MATTHEWS, G. 2009. Individual differences in stress and fatigue in two field studies of driving. *Transportation Research Part F: Traffic Psychology and Behaviour,* 12**,** 265-276.
DESMOND, P. A., NEUBAUER, M. C., MATTHEWS, G. & HANCOCK, P. A. 2012. *The handbook of operator fatigue*, Ashgate Publishing, Ltd.
DING, Y., CAO, Y., DUFFY, V. G., WANG, Y. & ZHANG, X. 2020. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. *Ergonomics,* 63**,** 896-908.
DING, Y., MA, J. & LUO, X. 2022. Applications of natural language processing in construction. *Automation in Construction,* 136**,** 104169.
DINGES, D. F., RIDER, R. L., DORRIAN, J., MCGLINCHEY, E. L., ROGERS, N. L., CIZMAN, Z., GOLDENSTEIN, S. K., VOGLER, C., VENKATARAMAN, S. & METAXAS, D. N. 2005. Optical computer recognition of facial expressions associated with stress induced by performance demands. *Aviation, space, and environmental medicine,* 76**,** B172-B182.
DISSANAYAKE, U. C., STEUBER, V. & AMIRABDOLLAHIAN, F. 2022. EEG Spectral Feature Modulations Associated With Fatigue in Robot-Mediated Upper Limb Gross and Fine Motor Interactions. *Frontiers in Neurorobotics,* 15**,** 192.
DOERR, J. M., DITZEN, B., STRAHLER, J., LINNEMANN, A., ZIEMEK, J., SKOLUDA, N., HOPPMANN, C. A. & NATER, U. M. 2015. Reciprocal relationship between acute stress and acute fatigue in everyday life in a sample of university students. *Biological Psychology,* 110**,** 42-49.
DUAN, R., DENG, H., TIAN, M., DENG, Y. & LIN, J. 2022. SODA: A large-scale open site object detection dataset for deep learning in construction. *Automation in Construction,* 142**,** 104499.
DZIUDA, Ł., BARAN, P., ZIELIŃSKI, P., MURAWSKI, K., DZIWOSZ, M., KREJ, M., PIOTROWSKI, M., STABLEWSKI, R., WOJDAS, A., STRUS, W., GASIUL, H., KOSOBUDZKI, M. & BORTKIEWICZ, A. 2021. Evaluation of a Fatigue Detector Using Eye Closure-Associated Indicators Acquired from Truck Drivers in a Simulator Study. *Sensors,* 21**,** 6449.
EL KERDAWY, M., EL HALABY, M., HASSAN, A., MAHER, M., FAYED, H., SHAWKY, D. & BADAWI, A. 2020. The Automatic Detection of Cognition Using EEG and Facial Expressions. *Sensors (Basel, Switzerland),* 20**,** 3516.
EOH, H. J., CHUNG, M. K. & KIM, S.-H. 2005. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *International Journal of Industrial Ergonomics,* 35**,** 307-320.
EREN, L. 2017. Bearing fault detection by one-dimensional convolutional neural networks. *Mathematical Problems in Engineering,* 2017.
EREN, L., INCE, T. & KIRANYAZ, S. 2019. A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *Journal of Signal Processing Systems,* 91**,** 179-189.
EUROSTAT 2020. "Accidents at work statistics" https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics.
FAN, X., ZHOU, Q., LIU, Z. & XIE, F. 2015. Electroencephalogram assessment of mental fatigue in visual search. *Bio-medical materials and engineering,* 26**,** S1455-S1463.
FANG, Q., LI, H., LUO, X., DING, L., LUO, H., ROSE, T. M. & AN, W. 2018a. Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. *Automation in Construction,* 85**,** 1-9.
FANG, W., DING, L., LUO, H. & LOVE, P. E. 2018b. Falls from heights: A computer vision-based approach for safety harness detection. *Automation in Construction,* 91**,** 53-61.

FANG, Y. & CHO, Y. K. 2017. Effectiveness Analysis from a Cognitive Perspective for a Real-Time Safety Assistance System for Mobile Crane Lifting Operations. *Journal of Construction Engineering and Management,* 143**,** 05016025.

FARAG, A., SCOTT, L., PERKHOUNKOVA, Y., SAEIDZADEH, S. & HEIN, M. 2022. A human factors approach to evaluate predicators of acute care nurse occupational fatigue. *Applied Ergonomics,* 100**,** 103647.

FENG, Y., ZHANG, S. & WU, P. 2015. Factors influencing workplace accident costs of building projects. *Safety science,* 72**,** 97-104.

FITZPATRICK, S. & KUO, J. R. 2016. The impact of stimulus arousal level on emotion regulation effectiveness in borderline personality disorder. *Psychiatry Research,* 241**,** 242-248.

FRONE, M. R. & TIDWELL, M.-C. O. 2015. The meaning and measurement of work fatigue: Development and evaluation of the Three-Dimensional Work Fatigue Inventory (3D-WFI). *Journal of occupational health psychology,* 20**,** 273.

FRYDENLUND, A. & RUDZICZ, F. Emotional affect estimation using video and EEG data in deep neural networks. Advances in Artificial Intelligence: 28th Canadian Conference on Artificial Intelligence, Canadian AI 2015, Halifax, Nova Scotia, Canada, June 2-5, 2015, Proceedings 28, 2015. Springer, 273-280.

GHODDOOSIAN, R., GALIB, M. & ATHITSOS, V. A Realistic Dataset and Baseline Temporal Model for Early Drowsiness Detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 16-17 June 2019 2019. 178-187.

GIANNAKAKIS, G., GRIGORIADIS, D., GIANNAKAKI, K., SIMANTIRAKI, O., RONIOTIS, A. & TSIKNAKIS, M. 2019. Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing***,** 1-1.

GIANNAKAKIS, G., MANOUSOS, D., SIMOS, P. & TSIKNAKIS, M. Head movements in context of speech during stress induction. 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018. IEEE, 710-714.

GIANNAKAKIS, G., PEDIADITIS, M., MANOUSOS, D., KAZANTZAKI, E., CHIARUGI, F., SIMOS, P. G., MARIAS, K. & TSIKNAKIS, M. 2017. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control,* 31**,** 89-101.

GOETZ, C., BAVARESCO, R., KUNST, R. & BARBOSA, J. 2022. Industrial intelligence in the care of workers' mental health: A review of status and challenges. *International Journal of Industrial Ergonomics,* 87**,** 103234.

GRECO, A., VALENZA, G., LANATA, A., SCILINGO, E. P. & CITI, L. 2015. cvxEDA: A convex optimization approach to electrodermal activity processing. *IEEE Transactions on Biomedical Engineering,* 63**,** 797-804.

GRIER, R. A. How high is high? A meta-analysis of NASA-TLX global workload scores. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2015. SAGE Publications Sage CA: Los Angeles, CA, 1727-1731.

HALLOWELL, M. R., HINZE, J. W., BAUD, K. C. & WEHLE, A. 2013. Proactive construction safety control: Measuring, monitoring, and responding to safety leading indicators. *Journal of construction engineering and management,* 139**,** 04013010.

HAN, Y., JIN, R., WOOD, H. & YANG, T. 2019. Investigation of Demographic Factors in Construction Employees' Safety Perceptions. *KSCE Journal of Civil Engineering,* 23**,** 2815-2828.

HAN, Y., YIN, Z., ZHANG, J., JIN, R. & YANG, T. 2020. Eye-Tracking Experimental Study Investigating the Influence Factors of Construction Safety Hazard Recognition. *Journal of Construction Engineering and Management,* 146**,** 04020091.

HART, S. G. NASA-task load index (NASA-TLX); 20 years later. Proceedings of the human factors and ergonomics society annual meeting, 2006a. Sage publications Sage CA: Los Angeles, CA, 904-908.

HART, S. G. 2006b. Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 50**,** 904-908.

HASANZADEH, S., ESMAEILI, B. & DODD, M. D. 2018. Examining the Relationship between Construction Workers’ Visual Attention and Situation Awareness under Fall and Tripping Hazard Conditions: Using Mobile Eye Tracking. *Journal of Construction Engineering and Management,* 144**,** 04018060.

HAZLETT, R. L. 2006. Measuring emotional valence during interactive experiences: boys at video game play. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Montréal, Québec, Canada: Association for Computing Machinery.

HENNI, K., MEZGHANI, N., GOUIN-VALLERAND, C., RUER, P., OUAKRIM, Y. & VALLIÈRES, É. Feature selection for driving fatigue characterization and detection using visual-and signal-based sensors. Applied Informatics, 2018. Springer, 1-15.

HINZE, J. W. & TEIZER, J. 2011. Visibility-related fatalities related to construction equipment. *Safety science,* 49**,** 709-718.

HIRACHAN, N., MATHEWS, A., ROMERO, J. & ROJAS, R. F. Measuring Cognitive Workload Using Multimodal Sensors. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 11-15 July 2022 2022. 4921-4924.

HKOSHS 2020. Hong Kong Occupational Safety and Health Statistics, . *https://www.labour.gov.hk/eng/osh/content10.htm,* (Accessed July 15, 2021).

HOCHREITER, S. & SCHMIDHUBER, J. 1997. Long short-term memory. *Neural computation,* 9**,** 1735-1780.

HOLGADO, D., TROYA, E., PERALES, J. C., VADILLO, M. A. & SANABRIA, D. 2020. Does mental fatigue impair physical performance? A replication study. *European Journal of Sport Science***,** 1-9.

HOPSTAKEN, J. F., VAN DER LINDEN, D., BAKKER, A. B. & KOMPIER, M. A. 2015. A multifaceted investigation of the link between mental fatigue and task disengagement. *Psychophysiology,* 52**,** 305-15.

HOPSTAKEN, J. F., VAN DER LINDEN, D., BAKKER, A. B., KOMPIER, M. A. J. & LEUNG, Y. K. 2016. Shifts in attention during mental fatigue: Evidence from subjective, behavioral, physiological, and eye-tracking data. *Journal of Experimental Psychology: Human Perception and Performance,* 42**,** 878-889.

HSE, H. A. S. E. 2020. Construction statistics in Great Britain, Available from:http://www.hse.gov.uk/statistics/industry/construction.pdf, Accessed date: September 30, 2021.

HU, J. & MIN, J. 2018. Automated detection of driver fatigue based on EEG signals using gradient boosting decision tree model. *Cognitive neurodynamics,* 12**,** 431-440.

HU, X. & LODEWIJKS, G. 2020. Detecting fatigue in car drivers and aircraft pilots by using non-invasive measures: The value of differentiation of sleepiness and mental fatigue. *Journal of safety research,* 72**,** 173-187.

HU, Z., CHAN, W. T., HU, H. & XU, F. 2023. Cognitive Factors Underlying Unsafe Behaviors of Construction Workers as a Tool in Safety Management: A Review. *Journal of Construction Engineering and Management,* 149**,** 03123001.

HWANG, S., JEBELLI, H., CHOI, B., CHOI, M. & LEE, S. 2018a. Measuring Workers’ Emotional State during Construction Tasks Using Wearable EEG. *Journal of Construction Engineering and Management,* 144**,** 04018050.

HWANG, S., JEBELLI, H., CHOI, B., CHOI, M. & LEE, S. 2018b. Measuring workers' emotional state during construction tasks using wearable EEG. *Journal of Construction Engineering and Management,* 144**,** 04018050.

ILO 2022. International Labour Organization, World statistic. The enormous burden of poor working conditions. *International Labour Organization.*

IWASAKI, M. & NOGUCHI, Y. 2016. Hiding true emotions: micro-expressions in eyes retrospectively concealed by mouth movements. *Scientific Reports,* 6**,** 22049.

IZMIRLIAN, G. 2020. Strong consistency and asymptotic normality for quantities related to the Benjamini–Hochberg false discovery rate procedure. *Statistics & Probability Letters,* 160**,** 108713.

JAAFAR, M. H., ARIFIN, K., AIYUB, K., RAZMAN, M. R., ISHAK, M. I. S. & SAMSURIJAN, M. S. 2018. Occupational safety and health management in the construction industry: a review. *International Journal of Occupational Safety and Ergonomics,* 24**,** 493-506.

JAISWAL, A., ZADEH, M. Z., HEBRI, A. & MAKEDON, F. 2022. Assessing Fatigue with Multimodal Wearable Sensors and Machine Learning. *arXiv preprint arXiv:2205.00287*.

JAP, B. T., LAL, S., FISCHER, P. & BEKIARIS, E. 2009. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications,* 36**,** 2352-2359.

JEBELLI, H., CHOI, B. & LEE, S. 2019a. Application of Wearable Biosensors to Construction Sites. I: Assessing Workers’ Stress. *Journal of Construction Engineering and Management,* 145**,** 04019079.

JEBELLI, H., HWANG, S. & LEE, S. 2018a. EEG-based workers' stress recognition at construction sites. *Automation in Construction,* 93**,** 315-324.

JEBELLI, H., HWANG, S. & LEE, S. 2018b. EEG signal-processing framework to obtain high-quality brain waves from an off-the-shelf wearable EEG device. *Journal of Computing in Civil Engineering,* 32**,** 04017070.

JEBELLI, H., KHALILI, M. M., HWANG, S. & LEE, S. 2018c. A Supervised Learning-Based Construction Workers’ Stress Recognition Using a Wearable Electroencephalography (EEG) Device. *Construction Research Congress 2018.*

JEBELLI, H., KHALILI, M. M. & LEE, S. 2019b. A Continuously Updated, Computationally Efficient Stress Recognition Framework Using Electroencephalogram (EEG) by Applying Online Multitask Learning Algorithms (OMTL). *IEEE Journal of Biomedical and Health Informatics,* 23**,** 1928-1939.

JEBELLI, H., KHALILI, M. M. & LEE, S. Mobile EEG-Based Workers’ Stress Recognition by Applying Deep Neural Network. *In:* MUTIS, I. & HARTMANN, T., eds. Advances in Informatics and Computing in Civil and Construction Engineering, 2019// 2019c Cham. Springer International Publishing, 173-180.

JEBELLI, H., MAHDI KHALILI, M. & LEE, S. 2019d. A Continuously Updated, Computationally Efficient Stress Recognition Framework Using Electroencephalogram (EEG) by Applying Online Multitask Learning Algorithms (OMTL). *IEEE J Biomed Health Inform,* 23**,** 1928-1939.

JEELANI, I., ALBERT, A., HAN, K. & AZEVEDO, R. 2019. Are visual search patterns predictive of hazard recognition performance? Empirical investigation using eye-tracking technology. *Journal of construction engineering and management,* 145**,** 04018115.

JEON, J. & CAI, H. 2022. Multi-class classification of construction hazards via cognitive states assessment using wearable EEG. *Advanced Engineering Informatics,* 53**,** 101646.

JOHNSTON, B. & CHAZAL, P. D. 2018. A review of image-based automatic facial landmark identification techniques. *EURASIP Journal on Image and Video Processing,* 2018**,** 86.

JOVANOVIC, T., NORRHOLM, S. D., SAKOMAN, A. J., ESTERAJHER, S. & KOZARIĆ-KOVAČIĆ, D. 2009. Altered resting psychophysiology and startle response in Croatian combat veterans with PTSD. *International Journal of Psychophysiology,* 71**,** 264-268.

KADUK, S. I., ROBERTS, A. P. J. & STANTON, N. A. 2021. Driving performance, sleepiness, fatigue, and mental workload throughout the time course of semi-automated driving—Experimental data from the driving simulator. *Human Factors and Ergonomics in Manufacturing & Service Industries,* 31**,** 143-154.

KÄTHNER, I., WRIESSNEGGER, S. C., MÜLLER-PUTZ, G. R., KÜBLER, A. & HALDER, S. 2014. Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain–computer interface. *Biological psychology,* 102**,** 118-129.

KAUR, C., SINGH, P., BISHT, A., JOSHI, G. & AGRAWAL, S. 2022. Recent Developments in Spatio-Temporal EEG Source Reconstruction Techniques. *Wireless Personal Communications,* 122**,** 1531-1558.

KE, J., DU, J. & LUO, X. 2021a. The effect of noise content and level on cognitive performance measured by electroencephalography (EEG). *Automation in Construction,* 130**,** 103836.

KE, J., ZHANG, M., LUO, X. & CHEN, J. 2021b. Monitoring distraction of construction workers caused by noise using a wearable Electroencephalography (EEG) device. *Automation in Construction,* 125**,** 103598.

KIM, K. & CHO, Y. K. 2020. Effective inertial sensor quantity and locations on a body for deep learning-based worker's motion recognition. *Automation in Construction,* 113**,** 103126.

KIMMELMAN, V., IMASHEV, A., MUKUSHEV, M. & SANDYGULOVA, A. 2020. Eyebrow position in grammatical and emotional expressions in Kazakh-Russian Sign Language: A quantitative study. *PLOS ONE,* 15**,** e0233731.

KING, D. E. 2009. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research,* 10**,** 1755-1758.
KINGMA, D. P. & BA, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
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.
KIRANYAZ, S., GASTLI, A., BEN-BRAHIM, L., AL-EMADI, N. & GABBOUJ, M. 2018. Real-time fault detection and identification for MMC using 1-D convolutional neural networks. *IEEE Transactions on Industrial Electronics,* 66**,** 8760-8771.
KOC, K. & GURGUN, A. P. 2022. Scenario-based automated data preprocessing to predict severity of construction accidents. *Automation in Construction,* 140**,** 104351.
KRAUSS, T. P., SHURE, L. & LITTLE, J. 1994. *Signal processing toolbox for use with MATLAB®: user's guide*, The MathWorks.
KUKASVADIYA, M. S. & DIVECHA, N. H. 2017. Analysis of data using data mining tool orange. *International Journal of Engineering Development and Research,* 5**,** 1836-1840.
KUSANO, H., HORIGUCHI, Y., BABA, Y. & KASHIMA, H. Stress Prediction from Head Motion. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), 6-9 Oct. 2020 2020. 488-495.
KUWAHARA, A., NISHIKAWA, K., HIRAKAWA, R., KAWANO, H. & NAKATOH, Y. 2022. Eye fatigue estimation using blink detection based on Eye Aspect Ratio Mapping (EARM). *Cognitive Robotics,* 2**,** 50-59.
LABOR, H. K. D. O. 2022. Summary of Occupational Safety and Health Statistics of 1st Quarter of 2022, https://www.labour.gov.hk/common/osh/pdf/summary_OSH_Statistics_en.pdf (Accessed on: 12 August 2022).
LAITINEN, H. & PÄIVÄRINTA, K. 2010. A new-generation safety contest in the construction industry–a long-term evaluation of a real-life intervention. *Safety science,* 48**,** 680-686.
LAL, S. K. & CRAIG, A. 2002. Driver fatigue: electroencephalography and psychological assessment. *Psychophysiology,* 39**,** 313-321.
LAL, S. K. & CRAIG, A. 2005. Reproducibility of the spectral components of the electroencephalogram during driver fatigue. *International Journal of Psychophysiology,* 55**,** 137-143.
LAZARO, M. J., YUN, M. H. & KIM, S. 2022. Stress-level and attentional functions of experienced and novice young adult drivers in intersection-related hazard situations. *International Journal of Industrial Ergonomics,* 90**,** 103315.
LECUN, Y., BENGIO, Y. & HINTON, G. 2015. Deep learning. *Nature,* 521**,** 436-444.
LEE, B. & CHUNG, W. 2012. Driver Alertness Monitoring Using Fusion of Facial Features and Bio-Signals. *IEEE Sensors Journal,* 12**,** 2416-2422.
LEE, B. G., CHOI, B., JEBELLI, H. & LEE, S. 2021. Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach. *Journal of Building Engineering,* 42**,** 102824.
LEE, G. & LEE, S. 2022. Feasibility of a Mobile Electroencephalogram (EEG) Sensor-Based Stress Type Classification for Construction Workers. *Construction Research Congress 2022.*
LEE, H.-W., PENG, F.-F., LEE, X.-Y., DAI, H.-N. & ZHU, Y. Research on face detection under different lighting. 2018 IEEE International Conference on Applied System Invention (ICASI), 2018. IEEE, 1145-1148.
LEE, H., YANG, K., KIM, N. & AHN, C. R. 2020. Detecting excessive load-carrying tasks using a deep learning network with a Gramian Angular Field. *Automation in Construction,* 120**,** 103390.
LI, G., HUANG, S., XU, W., JIAO, W., JIANG, Y., GAO, Z. & ZHANG, J. 2020a. The impact of mental fatigue on brain activity: A comparative study both in resting state and task state using EEG. *BMC neuroscience,* 21**,** 1-9.
LI, G., HUANG, S., XU, W., JIAO, W., JIANG, Y., GAO, Z. & ZHANG, J. 2020b. The impact of mental fatigue on brain activity: a comparative study both in resting state and task state using EEG. *BMC Neuroscience,* 21**,** 20.

LI, G., LEE, C. H., JUNG, J. J., YOUN, Y. C. & CAMACHO, D. 2020c. Deep learning for EEG data analytics: A survey. *Concurrency and Computation: Practice and Experience,* 32**,** e5199.

LI, H., WANG, D., CHEN, J., LUO, X., LI, J. & XING, X. 2019a. Pre-service fatigue screening for construction workers through wearable EEG-based signal spectral analysis. *Automation in Construction,* 106**,** 102851.

LI, J., LI, H., UMER, W., WANG, H., XING, X., ZHAO, S. & HOU, J. 2020d. Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology. *Automation in Construction,* 109**,** 103000.

LI, J., LI, H., WANG, F., CHENG, A. S. K., YANG, X. & WANG, H. 2021a. Proactive analysis of construction equipment operators' hazard perception error based on cognitive modeling and a dynamic Bayesian network. *Reliability Engineering & System Safety,* 205**,** 107203.

LI, J., LI, H., WANG, H., UMER, W., FU, H. & XING, X. 2019b. Evaluating the impact of mental fatigue on construction equipment operators' ability to detect hazards using wearable eye-tracking technology. *Automation in Construction,* 105**,** 102835.

LI, R., CHEN, Y. V. & ZHANG, L. 2021b. A method for fatigue detection based on Driver's steering wheel grip. *International Journal of Industrial Ergonomics,* 82**,** 103083.

LI, R., SU, W. & LU, Z. 2017a. Physiological signal analysis for fatigue level of experienced and inexperienced drivers. *Traffic Injury Prevention,* 18**,** 139-144.

LI, W., ZHANG, L. & LIANG, W. 2017b. An Accident Causation Analysis and Taxonomy (ACAT) model of complex industrial system from both system safety and control theory perspectives. *Safety science,* 92**,** 94-103.

LI, Y.-H., HARFIYA, L. N., PURWANDARI, K. & LIN, Y.-D. 2020e. Real-Time Cuffless Continuous Blood Pressure Estimation Using Deep Learning Model. *Sensors,* 20**,** 5606.

LIANG, W. C., YUAN, J., SUN, D. C. & LIN, M. H. 2009. Changes in Physiological Parameters Induced by Indoor Simulated Driving: Effect of Lower Body Exercise at Mid-Term Break. *Sensors,* 9**,** 6913-6933.

LIAO, P.-C., ZHOU, X., CHONG, H.-Y., HU, Y. & ZHANG, D. 2022. Exploring construction workers' brain connectivity during hazard recognition: a cognitive psychology perspective. *International Journal of Occupational Safety and Ergonomics***,** 1-9.

LIAO, W., ZHANG, W., ZHU, Z. & JI, Q. A decision theoretic model for stress recognition and user assistance. AAAI, 2005a. 529-534.

LIAO, W., ZHANG, W., ZHU, Z. & JI, Q. A real-time human stress monitoring system using dynamic Bayesian network. 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)-workshops, 2005b. IEEE, 70-70.

LIN, L., HUANG, C., NI, X., WANG, J., ZHANG, H., LI, X. & QIAN, Z. 2015. Driver fatigue detection based on eye state. *Technology and health care,* 23**,** S453-S463.

LIU, P., CHI, H.-L., LI, X. & GUO, J. 2021a. Effects of dataset characteristics on the performance of fatigue detection for crane operators using hybrid deep neural networks. *Automation in Construction,* 132**,** 103901.

LIU, Q., LIU, Y., CHEN, K., WANG, L., LI, Z., AI, Q. & MA, L. 2021b. Research on Channel Selection and Multi-Feature Fusion of EEG Signals for Mental Fatigue Detection. *Entropy,* 23**,** 457.

LIU, W., ZHANG, Z., NIE, J. & FU, B. Research on the Correlation Between the Viewing Screen Layout of In-Vehicle Information Terminal and Crew's Mental Workload. International Conference on Man-Machine-Environment System Engineering, 2016. Springer, 341-346.

LIU, X., LI, G., WANG, S., WAN, F., SUN, Y., WANG, H., BEZERIANOS, A., LI, C. & SUN, Y. 2021c. Toward practical driving fatigue detection using three frontal EEG channels: A proof-of-concept study. *Physiological Measurement,* 42**,** 044003.

LIU, Y., CHEN, X., PENG, H. & WANG, Z. 2017. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion,* 36**,** 191-207.

MA, J., GU, J., JIA, H., YAO, Z. & CHANG, R. 2018. The Relationship Between Drivers' Cognitive Fatigue and Speed Variability During Monotonous Daytime Driving. *Frontiers in Psychology,* 9.

MA, J., LI, X., REN, Y., YANG, R. & ZHAO, Q. 2021a. Landmark-Based Facial Feature Construction and Action Unit Intensity Prediction. *Mathematical Problems in Engineering,* 2021**,** 6623239.

MA, L., GUO, H. & FANG, Y. 2021b. Analysis of Construction Workers’ Safety Behavior Based on Myers-Briggs Type Indicator Personality Test in a Bridge Construction Project. *Journal of Construction Engineering and Management,* 147**,** 04020149.
MAFFEI, A. & ANGRILLI, A. 2019. Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiology & Behavior,* 204**,** 256-263.
MAJID, M., ARSALAN, A. & ANWAR, S. M. 2022. A Multimodal Perceived Stress Classification Framework using Wearable Physiological Sensors. *arXiv preprint arXiv:2206.10846*.
MASULLO, M., TOMA, R., PASCALE, A., RUGGIERO, G. & MAFFEI, L. Research methodology used to investigate the effects of noise on overhead crane operator's performances. International Ergonomics Conference, 2020. Springer, 223-231.
MASULLO, M., TOMA, R. A., PASCALE, A., RUGGIERO, G. & MAFFEI, L. Research Methodology Used to Investigate the Effects of Noise on Overhead Crane Operator's Performances. 2021 Cham. Springer International Publishing, 223-231.
MAT RONI, S., DJAJADIKERTA, H. G., MAT RONI, S. & DJAJADIKERTA, H. G. 2021. Non-Parametric Tests. *Data Analysis with SPSS for Survey-based Research***,** 219-260.
MEHMOOD, I., LI, H., QAROUT, Y., UMER, W., ANWER, S., WU, H., HUSSAIN, M. & FORDJOUR ANTWI-AFARI, M. 2023. Deep learning-based construction equipment operators' mental fatigue classification using wearable EEG sensor data. *Advanced Engineering Informatics,* 56**,** 101978.
MEHMOOD, I., LI, H., UMER, W., ARSALAN, A., SAAD SHAKEEL, M. & ANWER, S. 2022. Validity of facial features' geometric measurements for real-time assessment of mental fatigue in construction equipment operators. *Advanced Engineering Informatics,* 54**,** 101777.
MEHRABIAN, A. 2017. Communication without words. *Communication theory.* Routledge.
MEM 2018. "A report on the safety production situation of the national construction industry" available at: https://www.mem.gov.cn/awhsy_3512/awhbgswj/201807/t20180725_247933.shtml (accessed on 14 August 2022).
METAXAS, D., VENKATARAMAN, S. & VOGLER, C. Image-Based Stress Recognition Using a Model-Based Dynamic Face Tracking System. 2004 Berlin, Heidelberg. Springer Berlin Heidelberg, 813-821.
MILSTEIN, N. & GORDON, I. 2020. Validating Measures of Electrodermal Activity and Heart Rate Variability Derived From the Empatica E4 Utilized in Research Settings That Involve Interactive Dyadic States. *Frontiers in Behavioral Neuroscience,* 14.
MING-HSUAN, Y., KRIEGMAN, D. J. & AHUJA, N. 2002. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24**,** 34-58.
MITROPOULOS, P. & MEMARIAN, B. 2013. Task Demands in Masonry Work: Sources, Performance Implications, and Management Strategies. *Journal of Construction Engineering and Management,* 139**,** 581-590.
MOLAN, G. & MOLAN, M. 2021. Theoretical model for accident prevention based on root cause analysis with graph theory. *Safety and health at work,* 12**,** 42-50.
MOON, S., LEE, G. & CHI, S. 2022. Automated system for construction specification review using natural language processing. *Advanced Engineering Informatics,* 51**,** 101495.
MORALES, J. M., DÍAZ-PIEDRA, C., RIEIRO, H., ROCA-GONZÁLEZ, J., ROMERO, S., CATENA, A., FUENTES, L. J. & DI STASI, L. L. 2017. Monitoring driver fatigue using a single-channel electroencephalographic device: A validation study by gaze-based, driving performance, and subjective data. *Accident Analysis & Prevention,* 109**,** 62-69.
MURPHY, K. P. 2012. *Machine learning: a probabilistic perspective*, MIT press.
NAIK, A. & SAMANT, L. 2016. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science,* 85**,** 662-668.
NAKAGOME, S., CRAIK, A., SUJATHA RAVINDRAN, A., HE, Y., CRUZ-GARZA, J. G. & CONTRERAS-VIDAL, J. L. 2022. Deep learning methods for EEG neural classification. *Handbook of Neuroengineering.* Springer.
NGUYEN, T., AHN, S., JANG, H., JUN, S. C. & KIM, J. G. 2017. Utilization of a combined EEG/NIRS system to predict driver drowsiness. *Scientific reports,* 7**,** 43933.
NIU, Y., LI, Z. & FAN, Y. 2021. Analysis of truck drivers' unsafe driving behaviors using four machine learning methods. *International Journal of Industrial Ergonomics,* 86**,** 103192.

NOGHABAEI, M., HAN, K. & ALBERT, A. 2021. Feasibility Study to Identify Brain Activity and Eye-Tracking Features for Assessing Hazard Recognition Using Consumer-Grade Wearables in an Immersive Virtual Environment. *Journal of Construction Engineering and Management,* 147**,** 04021104.

NORZALI, M., KASHIMA, M., SATO, K. & WATANABE, M. Facial Visual-Infrared Stereo Vision Fusion Measurement as an Alternative for Physiological Measurement. 2014.

OJHA, A., SHAKERIAN, S., HABIBNEZHAD, M. & JEBELLI, H. Feasibility Verification of Multimodal Wearable Sensing System for Holistic Health Monitoring of Construction Workers. *In:* WALBRIDGE, S., NIK-BAKHT, M., NG, K. T. W., SHOME, M., ALAM, M. S., EL DAMATTY, A. & LOVEGROVE, G., eds. Proceedings of the Canadian Society of Civil Engineering Annual Conference 2021, 2023// 2023 Singapore. Springer Nature Singapore, 283-294.

OLAH, C. 2015. Understanding lstm networks.

ORDÓÑEZ, F. J. & ROGGEN, D. 2016. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors,* 16**,** 115.

ORFANIDIS, S. J. 1995. *Introduction to signal processing*, Prentice-Hall, Inc.

OSHA 2019. US Department of Labor, Commonly Used Statistics (Accessed on: 25 March 2022) https://www.osha.gov/data/commonstats.

ÖZDEMIR, A. T. & BARSHAN, B. 2014. Detecting falls with wearable sensors using machine learning techniques. *Sensors,* 14**,** 10691-10708.

PALEJEV, D. & SAVOV, M. 2021. On the Convergence of the Benjamini–Hochberg Procedure. *Mathematics,* 9**,** 2154.

PBS 2015. "Pakistan Bureau of Statistics - Labour Force Statistics (2014-15)", Islamabad, available at: https://www.pbs.gov.pk/sites/default/files/labour_force/publications/lfs2014_15/t33-pak.pdf (accessed on 14 August 2022).

PBS 2018. "Pakistan Bureau of Statistics – Labour Force Statistics (2017-18)", PBS, Islamabad, available at: https://www.pbs.gov.pk/sites/default/files/labour_force/publications/lfs2017_18/Table-30_perc_R.pdf (accessed on 14 August 2022).

PBS 2021. "Pakistan Bureau of Statistics – Labour Force Statistics (2020-21)", PBS, Islamabad, available at: https://www.pbs.gov.pk/sites/default/files/labour_force/publications/lfs2020_21/tables/Table_28.pdf (accessed on 14 August 2022).

PEDROTTI, M., MIRZAEI, M. A., TEDESCO, A., CHARDONNET, J.-R., MÉRIENNE, F., BENEDETTO, S. & BACCINO, T. 2014. Automatic Stress Classification With Pupil Diameter Analysis. *International Journal of Human–Computer Interaction,* 30**,** 220-236.

PHUTELA, N., RELAN, D., GABRANI, G., KUMARAGURU, P. & SAMUEL, M. 2022. Stress Classification Using Brain Signals Based on LSTM Network. *Computational Intelligence and Neuroscience,* 2022**,** 7607592.

POH, M.-Z., SWENSON, N. C. & PICARD, R. W. 2010. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *IEEE transactions on Biomedical engineering,* 57**,** 1243-1252.

POSADA-QUINTERO, H. F. & CHON, K. H. 2020. Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review. *Sensors,* 20**,** 479.

PRABASWARI, A. D., BASUMERDA, C. & UTOMO, B. W. The mental workload analysis of staff in study program of private educational organization. IOP Conference Series: Materials Science and Engineering, 2019. IOP Publishing, 012018.

PUSPAWARDHANI, E. H., SURYOPUTRO, M. R., SARI, A. D., KURNIA, R. D. & PURNOMO, H. 2016. Mental workload analysis using NASA-TLX method between various level of work in plastic injection division of manufacturing company. *Advances in safety management and human factors.* Springer.

RAHEEL, A., MAJID, M., ALNOWAMI, M. & ANWAR, S. M. 2020. Physiological Sensors Based Emotion Recognition While Experiencing Tactile Enhanced Multimedia. *Sensors,* 20**,** 4037.

RAHEEL, A., MAJID, M. & ANWAR, S. M. 2021. DEAR-MULSEMEDIA: Dataset for emotion analysis and recognition in response to multiple sensorial media. *Information Fusion,* 65**,** 37-49.

RAJULA, H. S. R., VERLATO, G., MANCHIA, M., ANTONUCCI, N. & FANOS, V. 2020. Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. *Medicina (Kaunas),* 56.

RASTGOO, M. N., NAKISA, B., MAIRE, F., RAKOTONIRAINY, A. & CHANDRAN, V. 2019. Automatic driver stress level classification using multimodal deep learning. *Expert Systems with Applications,* 138**,** 112793.

RAUFI, B. & LONGO, L. 2022. An Evaluation of the EEG Alpha-to-Theta and Theta-to-Alpha Band Ratios as Indexes of Mental Workload. *Frontiers in Neuroinformatics,* 16.

RAVAJA, N., SAARI, T., SALMINEN, M., LAARNI, J. & KALLINEN, K. 2006. Phasic Emotional Reactions to Video Game Events: A Psychophysiological Investigation. *Media Psychology,* 8**,** 343-367.

ROSEBROCK, A. 2017. Facial landmarks with dlib, OpenCV, and Python [Online], https://www.pyimagesearch.com/2017/04/03/facial-landmarks-dlib-opencv-python/.

ROY, R. N., BONNET, S., CHARBONNIER, S. & CAMPAGNE, A. Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI. 2013 35th annual international conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2013. IEEE, 6607-6610.

ROY, Y., BANVILLE, H., ALBUQUERQUE, I., GRAMFORT, A., FALK, T. H. & FAUBERT, J. 2019. Deep learning-based electroencephalography analysis: a systematic review. *Journal of Neural Engineering,* 16**,** 051001.

SAEDI, S., FINI, A. A. F., KHANZADI, M., WONG, J., SHEIKHKHOSHKAR, M. & BANAEI, M. 2022. Applications of electroencephalography in construction. *Automation in Construction,* 133**,** 103985.

SAGONAS, C., ANTONAKOS, E., TZIMIROPOULOS, G., ZAFEIRIOU, S. & PANTIC, M. 2016. 300 Faces In-The-Wild Challenge: database and results. *Image and Vision Computing,* 47**,** 3-18.

SAMARA, A., GALWAY, L., BOND, R. & WANG, H. Sensing Affective States Using Facial Expression Analysis. *In:* GARCÍA, C. R., CABALLERO-GIL, P., BURMESTER, M. & QUESADA-ARENCIBIA, A., eds. Ubiquitous Computing and Ambient Intelligence, 2016// 2016 Cham. Springer International Publishing, 341-352.

SANEI, S. & CHAMBERS, J. A. 2013. *EEG signal processing*, John Wiley & Sons.

SARKAR, A., SINGH, A. & CHAKRABORTY, R. 2022. A deep learning-based comparative study to track mental depression from EEG data. *Neuroscience Informatics,* 2**,** 100039.

SARKAR, S., PRAMANIK, A., MAITI, J. & RENIERS, G. 2020. Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Safety science,* 125**,** 104616.

SAVITZKY, A. & GOLAY, M. J. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry,* 36**,** 1627-1639.

SEO, J. & LEE, S. 2021. Automated postural ergonomic risk assessment using vision-based posture classification. *Automation in Construction,* 128**,** 103725.

SHAO, B., HU, Z., LIU, Q., CHEN, S. & HE, W. 2019. Fatal accident patterns of building construction activities in China. *Safety science,* 111**,** 253-263.

SHARMA, N. & GEDEON, T. 2014. Modeling observer stress for typical real environments. *Expert Syst. Appl.,* 41**,** 2231–2238.

SHI, S.-Y., TANG, W.-Z. & WANG, Y.-Y. A review on fatigue driving detection. ITM Web of Conferences, 2017. EDP Sciences, 01019.

SIAMI-NAMINI, S., TAVAKOLI, N. & NAMIN, A. S. The Performance of LSTM and BiLSTM in Forecasting Time Series. 2019 IEEE International Conference on Big Data (Big Data), 9-12 Dec. 2019 2019. 3285-3292.

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**,** 1929-1958.

STANCIN, I., FRID, N., CIFREK, M. & JOVIC, A. 2021. EEG signal multichannel frequency-domain ratio indices for drowsiness detection based on multicriteria optimization. *Sensors,* 21**,** 6932.

STEWART, J. 2011. *Calculus*, Cengage Learning.

SUN, Y., LIM, J., KWOK, K. & BEZERIANOS, A. 2014. Functional cortical connectivity analysis of mental fatigue unmasks hemispheric asymmetry and changes in small-world networks. *Brain and cognition,* 85**,** 220-230.

SUNDELIN, T., LEKANDER, M., KECKLUND, G., VAN SOMEREN, E. J. W., OLSSON, A. & AXELSSON, J. 2013. Cues of Fatigue: Effects of Sleep Deprivation on Facial Appearance. *Sleep,* 36**,** 1355-1360.

SWEENEY, K. T., WARD, T. E. & MCLOONE, S. F. 2012. Artifact Removal in Physiological Signals—Practices and Possibilities. *IEEE Transactions on Information Technology in Biomedicine,* 16**,** 488-500.

TANG, X., ZHOU, P. & WANG, P. Real-time image-based driver fatigue detection and monitoring system for monitoring driver vigilance. 2016 35th Chinese Control Conference (CCC), 27-29 July 2016 2016. 4188-4193.

TAO, D., TAN, H., WANG, H., ZHANG, X., QU, X. & ZHANG, T. 2019. A systematic review of physiological measures of mental workload. *International journal of environmental research and public health,* 16**,** 2716.

TAYLOR, S., JAQUES, N., CHEN, W., FEDOR, S., SANO, A. & PICARD, R. Automatic identification of artifacts in electrodermal activity data. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015. IEEE, 1934-1937.

TECHERA, U., HALLOWELL, M., LITTLEJOHN, R. & RAJENDRAN, S. 2018. Measuring and Predicting Fatigue in Construction: Empirical Field Study. *Journal of Construction Engineering and Management,* 144**,** 04018062.

TEHRANI, B. M., WANG, J. & TRUAX, D. 2021. Assessment of mental fatigue using electroencephalography (EEG) and virtual reality (VR) for construction fall hazard prevention. *Engineering, Construction and Architectural Management,* ahead-of-print.

THIFFAULT, P. & BERGERON, J. 2003. Monotony of road environment and driver fatigue: a simulator study. *Accident Analysis & Prevention,* 35**,** 381-391.

THODOROFF, P., PINEAU, J. & LIM, A. Learning robust features using deep learning for automatic seizure detection. Machine learning for healthcare conference, 2016. PMLR, 178-190.

TIJS, T. J. W., BROKKEN, D. & IJSSELSTEIJN, W. A. Dynamic Game Balancing by Recognizing Affect. 2008 Berlin, Heidelberg. Springer Berlin Heidelberg, 88-93.

TRAN, C.-K., TSENG, C.-D., CHANG, L. & LEE, T.-F. 2019. Face recognition under varying lighting conditions: improving the recognition accuracy for local descriptors based on weber-face followed by difference of Gaussians. *Journal of the Chinese Institute of Engineers,* 42**,** 593-601.

TRAN, C. C. & YAN, S. 2022. Development of an Eye Response-Based Mental Workload Evaluation Method: A Study of User interface in a Nuclear Power Plant. *International Journal of Technology and Human Interaction (IJTHI),* 18**,** 1-22.

TREJO, L. J., KUBITZ, K., ROSIPAL, R., KOCHAVI, R. L. & MONTGOMERY, L. D. 2015. EEG-based estimation and classification of mental fatigue. *Psychology,* 6**,** 572.

TÜRK, Ö. & ÖZERDEM, M. S. 2021. The convolutional neural network approach from electroencephalogram signals in emotional detection. *Concurrency and Computation: Practice and Experience,* 33**,** e6356.

TYAS, A. E., WIBAWA, A. D. & PURNOMO, M. H. Theta, Alpha and Beta Activity in the Occipital Based on EEG Signals for Mental Fatigue in High School Students. 2020 International Conference on Smart Technology and Applications (ICoSTA), 20-20 Feb. 2020 2020. 1-7.

UMER, W. 2022. Simultaneous monitoring of physical and mental stress for construction tasks using physiological measures. *Journal of Building Engineering,* 46**,** 103777.

UMER, W., LI, H., LU, W., SZETO, G. P. Y. & WONG, A. Y. 2018. Development of a tool to monitor static balance of construction workers for proactive fall safety management. *Automation in Construction,* 94**,** 438-448.

UMER, W., LI, H., YANTAO, Y., ANTWI-AFARI, M. F., ANWER, S. & LUO, X. 2020. Physical exertion modeling for construction tasks using combined cardiorespiratory and thermoregulatory measures. *Automation in Construction,* 112**,** 103079.

UMER, W., YU, Y. & AFARI, M. F. A. 2022. Quantifying the Effect of Mental Stress on Physical Stress for Construction Tasks. *Journal of Construction Engineering and Management,* 148**,** 04021204.

VAHDATIKHAKI, F., EL AMMARI, K., LANGROODI, A. K., MILLER, S., HAMMAD, A. & DOREE, A. 2019. Beyond data visualization: A context-realistic construction equipment training simulators. *Automation in construction,* 106**,** 102853.

VAN CUTSEM, J., MARCORA, S., DE PAUW, K., BAILEY, S., MEEUSEN, R. & ROELANDS, B. 2017. The Effects of Mental Fatigue on Physical Performance: A Systematic Review. *Sports Medicine,* 47**,** 1569-1588.

VAN DER LINDEN, D., FRESE, M. & SONNENTAG, S. 2003. The Impact of Mental Fatigue on Exploration in a Complex Computer Task: Rigidity and Loss of Systematic Strategies. *Human Factors,* 45**,** 483-494.

VELARDE, G., BRAÑEZ, P., BUENO, A., HEREDIA, R. & LOPEZ-LEDEZMA, M. 2022. An Open Source and Reproducible Implementation of LSTM and GRU Networks for Time Series Forecasting. *Engineering Proceedings,* 18**,** 30.

VENKATAANUSHA*, P., ANURADHA, C., CHANDRA MURTY, D. P. S. R. & CHEBROLU, D. S. K. 2019. Detecting Outliers in High Dimensional Data Sets Using Z-Score Methodology. *International Journal of Innovative Technology and Exploring Engineering,* 9**,** 48-53.

VILLANI, V., GABBI, M. & SABATTINI, L. Promoting operator's wellbeing in Industry 5.0: detecting mental and physical fatigue. 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 9-12 Oct. 2022 2022. 2030-2036.

WAGSTAFF, A. S. & SIGSTAD LIE, J.-A. 2011. Shift and night work and long working hours – a systematic review of safety implications. *Scandinavian Journal of Work, Environment & Health***,** 173-185.

WALAMBE, R., NAYAK, P., BHARDWAJ, A. & KOTECHA, K. 2021. Employing Multimodal Machine Learning for Stress Detection. *Journal of Healthcare Engineering,* 2021**,** 9356452.

WANG, C., GURAGAIN, B., VERMA, A. K., ARCHER, L., MAJUMDER, S., MOHAMUD, A., FLAHERTY-WOODS, E., SHAPIRO, G., ALMASHOR, M. & LENNÉ, M. 2019a. Spectral analysis of EEG during microsleep events annotated via driver monitoring system to characterize drowsiness. *IEEE Transactions on Aerospace and Electronic Systems,* 56**,** 1346-1356.

WANG, D., CHEN, J., ZHAO, D., DAI, F., ZHENG, C. & WU, X. 2017. Monitoring workers' attention and vigilance in construction activities through a wireless and wearable electroencephalography system. *Automation in construction,* 82**,** 122-137.

WANG, D., LI, H. & CHEN, J. 2019b. Detecting and measuring construction workers' vigilance through hybrid kinematic-EEG signals. *Automation in Construction,* 100**,** 11-23.

WANG, J., CHEN, D., ZHU, M. & SUN, Y. 2021. Risk assessment for musculoskeletal disorders based on the characteristics of work posture. *Automation in Construction,* 131**,** 103921.

WANG, M., ZHAO, Y. & LIAO, P.-C. 2022. EEG-based work experience prediction using hazard recognition. *Automation in Construction,* 136**,** 104151.

WANG, Y., ZHAI, G., ZHOU, S., CHEN, S., MIN, X., GAO, Z. & HU, M. 2018. Eye fatigue assessment using unobtrusive eye tracker. *Ieee Access,* 6**,** 55948-55962.

WENG, C.-H., LAI, Y.-H. & LAI, S.-H. Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network. Computer Vision – ACCV 2016 Workshops, 2017 Cham. Springer International Publishing, 117-133.

WENHUI, L., WEIHONG, Z., ZHIWEI, Z. & QIANG, J. A Real-Time Human Stress Monitoring System Using Dynamic Bayesian Network. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops, 21-23 Sept. 2005 2005. 70-70.

WILLIAMSON, A., LOMBARDI, D. A., FOLKARD, S., STUTTS, J., COURTNEY, T. K. & CONNOR, J. L. 2011. The link between fatigue and safety. *Accident Analysis & Prevention,* 43**,** 498-515.

WITTEN, I. H. & FRANK, E. 2002. Data mining: practical machine learning tools and techniques with Java implementations. *Acm Sigmod Record,* 31**,** 76-77.

WU, C., LI, X., GUO, Y., WANG, J., REN, Z., WANG, M. & YANG, Z. 2022. Natural language processing for smart construction: Current status and future directions. *Automation in Construction,* 134**,** 104059.

WU, E. Q., DENG, P. Y., QIU, X. Y., TANG, Z., ZHANG, W. M., ZHU, L. M., REN, H., ZHOU, G. R. & SHENG, R. S. F. 2021a. Detecting Fatigue Status of Pilots Based on Deep Learning Network Using EEG Signals. *IEEE Transactions on Cognitive and Developmental Systems,* 13**,** 575-585.

WU, H., ZHONG, B., LI, H., LOVE, P., PAN, X. & ZHAO, N. 2021b. Combining computer vision with semantic reasoning for on-site safety management in construction. *Journal of Building Engineering,* 42**,** 103036.

WU, Y., MIWA, T. & UCHIDA, M. Heart rate based evaluation of operator fatigue and its effect on performance during pipeline work. Advances in Physical Ergonomics and Human Factors: Proceedings of the AHFE 2017 International Conference on Physical Ergonomics and Human Factors, July 17-21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8, 2018. Springer, 446-454.

XING, X., LI, H., LI, J., ZHONG, B., LUO, H. & SKITMORE, M. 2019. A multicomponent and neurophysiological intervention for the emotional and mental states of high-altitude construction workers. *Automation in Construction,* 105**,** 102836.

XING, X., ZHONG, B., LUO, H., ROSE, T., LI, J. & ANTWI-AFARI, M. F. 2020a. Effects of physical fatigue on the induction of mental fatigue of construction workers: A pilot study based on a neurophysiological approach. *Automation in Construction,* 120.

XING, X., ZHONG, B., LUO, H., ROSE, T., LI, J. & ANTWI-AFARI, M. F. 2020b. Effects of physical fatigue on the induction of mental fatigue of construction workers: A pilot study based on a neurophysiological approach. *Automation in Construction,* 120**,** 103381.

XU, J., MIN, J. & HU, J. 2018. Real-time eye tracking for the assessment of driver fatigue. *Healthcare technology letters,* 5**,** 54-58.

XU, Q., NWE, T. L. & GUAN, C. 2015. Cluster-Based Analysis for Personalized Stress Evaluation Using Physiological Signals. *IEEE Journal of Biomedical and Health Informatics,* 19**,** 275-281.

YANG, J., YE, G., XIANG, Q., KIM, M., LIU, Q. & YUE, H. 2021. Insights into the mechanism of construction workers' unsafe behaviors from an individual perspective. *Safety Science,* 133**,** 105004.

YANG, K., AHN, C. R. & KIM, H. 2020. Deep learning-based classification of work-related physical load levels in construction. *Advanced Engineering Informatics,* 45**,** 101104.

YANG, X., LI, H., YU, Y., LUO, X., HUANG, T. & YANG, X. 2018. Automatic pixel-level crack detection and measurement using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering,* 33**,** 1090-1109.

YEŞILMEN, S. & TATAR, B. 2022. Efficiency of convolutional neural networks (CNN) based image classification for monitoring construction related activities: A case study on aggregate mining for concrete production. *Case Studies in Construction Materials,* 17**,** e01372.

YOUNG, M. S., BROOKHUIS, K. A., WICKENS, C. D. & HANCOCK, P. A. 2015. State of science: mental workload in ergonomics. *Ergonomics,* 58**,** 1-17.

YU, Y., YANG, X., LI, H., LUO, X., GUO, H. & FANG, Q. 2019. Joint-level vision-based ergonomic assessment tool for construction workers. *Journal of construction engineering and management,* 145**,** 04019025.

ZADEH, A., CHONG LIM, Y., BALTRUSAITIS, T. & MORENCY, L.-P. Convolutional experts constrained local model for 3d facial landmark detection. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017. 2519-2528.

ZARGARI MARANDI, R., MADELEINE, P., OMLAND, Ø., VUILLERME, N. & SAMANI, A. 2018. Eye movement characteristics reflected fatigue development in both young and elderly individuals. *Scientific Reports,* 8**,** 13148.

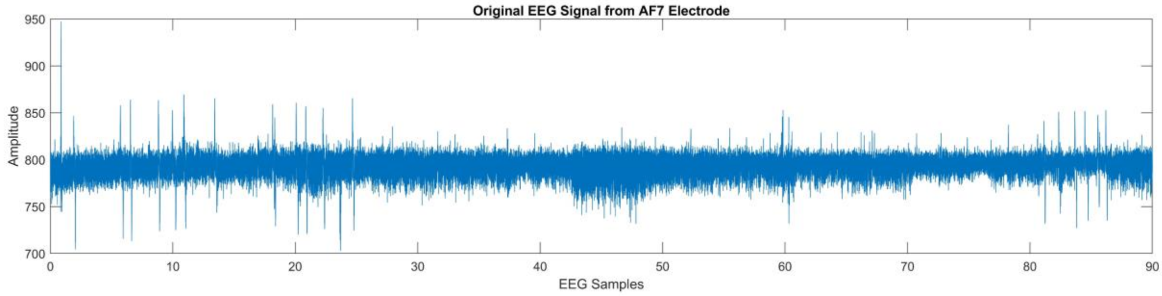
ZENG, H., YANG, C., DAI, G., QIN, F., ZHANG, J. & KONG, W. 2018. EEG classification of driver mental states by deep learning. *Cognitive Neurodynamics,* 12**,** 597-606.

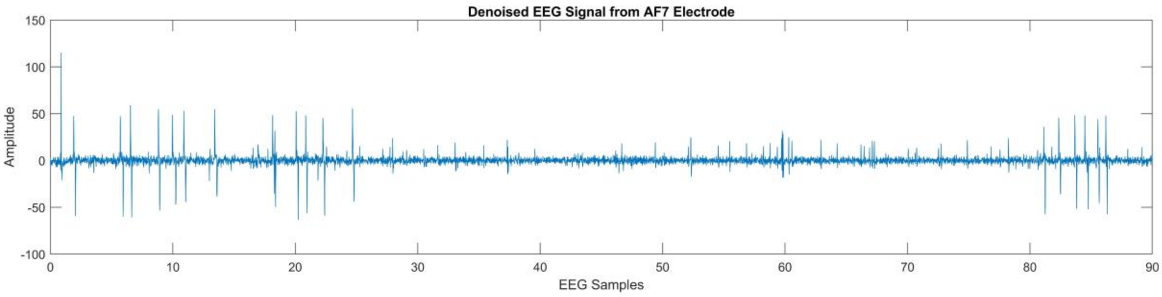
ZHANG, C. & ZHANG, Z. 2010a. A survey of recent advances in face detection.

ZHANG, F., FLEYEH, H., WANG, X. & LU, M. 2019. Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction,* 99**,** 238-248.
ZHANG, H., YAN, X. & LI, H. 2018. Ergonomic posture recognition using 3D view-invariant features from single ordinary camera. *Automation in Construction,* 94**,** 1-10.
ZHANG, Y., MA, J., ZHANG, C. & CHANG, R. 2021. Electrophysiological frequency domain analysis of driver passive fatigue under automated driving conditions. *Scientific Reports,* 11**,** 20348.
ZHANG, Z. & ZHANG, J. 2010b. A new real-time eye tracking based on nonlinear unscented Kalman filter for monitoring driver fatigue. *Journal of Control Theory and Applications,* 8**,** 181-188.
ZHAO, C., ZHAO, M., LIU, J. & ZHENG, C. 2012. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator. *Accident Analysis & Prevention,* 45**,** 83-90.
ZHAO, G., LIU, Y.-J. & SHI, Y. 2018. Real-time assessment of the cross-task mental workload using physiological measures during anomaly detection. *IEEE Transactions on Human-Machine Systems,* 48**,** 149-160.
ZHAO, J. & OBONYO, E. 2020. Convolutional long short-term memory model for recognizing construction workers' postures from wearable inertial measurement units. *Advanced Engineering Informatics,* 46**,** 101177.
ZHAO, J. & OBONYO, E. 2021. Applying incremental Deep Neural Networks-based posture recognition model for ergonomics risk assessment in construction. *Advanced Engineering Informatics,* 50**,** 101374.
ZHAO, L., LI, M., HE, Z., YE, S., QIN, H., ZHU, X. & DAI, Z. 2022a. Data-driven learning fatigue detection system: A multimodal fusion approach of ECG (electrocardiogram) and video signals. *Measurement,* 201**,** 111648.
ZHAO, Y. S., JAAFAR, M. H., MOHAMED, A. S. A., AZRAAI, N. Z. & AMIL, N. 2022b. Ergonomics Risk Assessment for Manual Material Handling of Warehouse Activities Involving High Shelf and Low Shelf Binning Processes: Application of Marker-Based Motion Capture. *Sustainability,* 14**,** 5767.
ZHENG, B., JIANG, X., TIEN, G., MENEGHETTI, A., PANTON, O. N. & ATKINS, M. S. 2012. Workload assessment of surgeons: correlation between NASA TLX and blinks. *Surg Endosc,* 26**,** 2746-50.
ZHENG, Y., LIU, Q., CHEN, E., GE, Y. & ZHAO, J. L. Time series classification using multi-channels deep convolutional neural networks. International conference on web-age information management, 2014. Springer, 298-310.
ZHONG, B., XING, X., LUO, H., ZHOU, Q., LI, H., ROSE, T. & FANG, W. 2020. Deep learning-based extraction of construction procedural constraints from construction regulations. *Advanced Engineering Informatics,* 43**,** 101003.
ZHU, Z., PARK, M.-W., KOCH, C., SOLTANI, M., HAMMAD, A. & DAVARI, K. 2016. Predicting movements of onsite workers and mobile equipment for enhancing construction site safety. *Automation in Construction,* 68**,** 95-101.

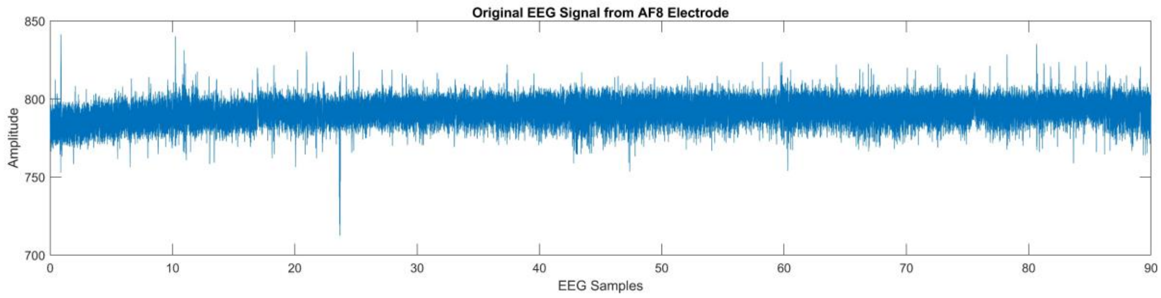
Appendix

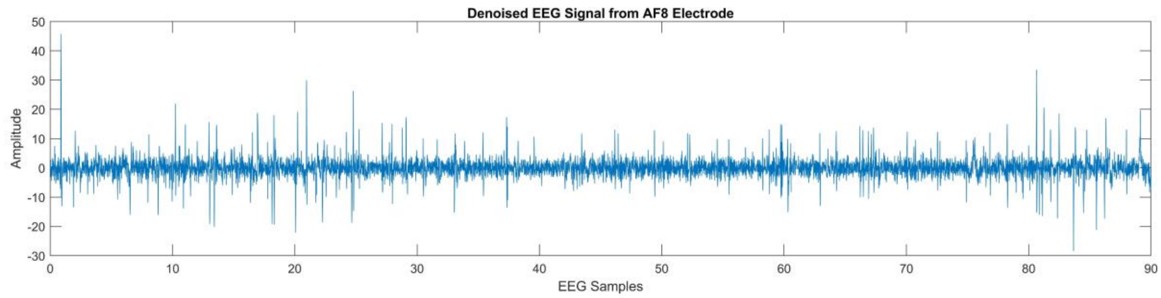
Appendix A. Example of artifact removal from electrode AF7



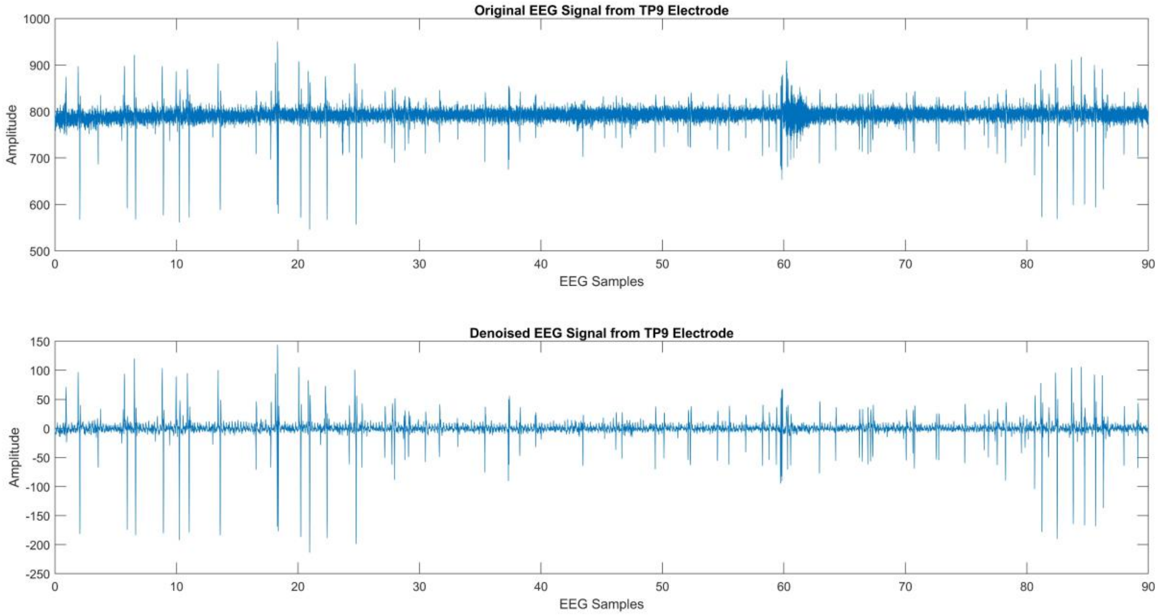


Appendix B. Example of artifact removal from electrode AF8

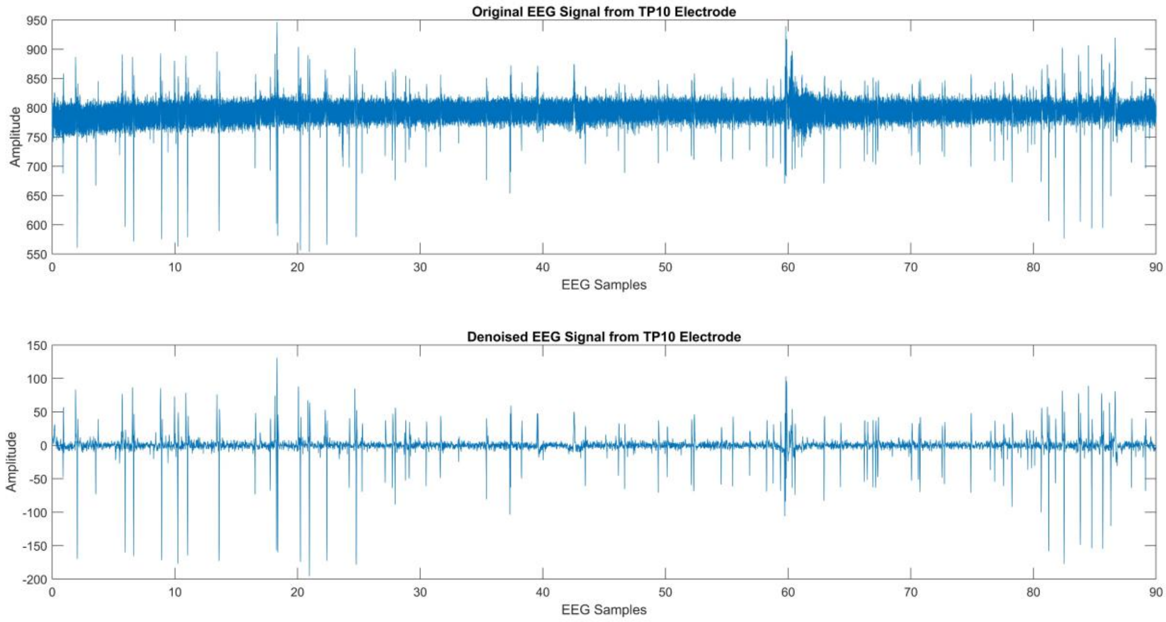




Appendix C. Example of artefact removal from electrode TP9



Appendix D. Example of artefact removal from electrode TP10



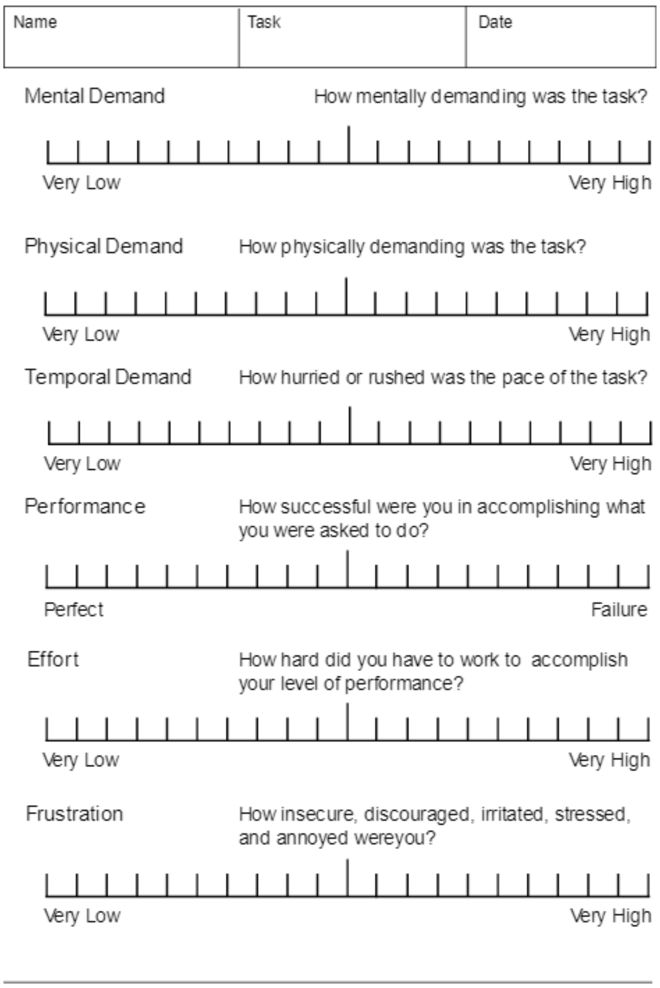
Appendix E. Comparison of facial features based non-invasive assessments.

Citation	Stimuli	Features	Results	*Conclusions and Recommendation*
Pedrotti et al. (2014)	Lane change test (LCT)	Visual Analog Scale (VAS) Pupil Diameter (PD)	*Visual Analog Scale* at t_1, F (1,30) = 0.03 at t_2, F (1,30) = 10.06 at t_3, F (1,30) = 7.79 at t_4, F (1,30) = 5.64 *Average Pupil Diameter* at t_1, F (1,27) = 0.05 at t_2, F (1,27) = 6.93 at t_3, F (1,27) = 5.35 at t_4, F (1,27) = 5.31	The subjective assessment of stress through VAS indicates the stress scores for the experimental group were significantly higher than control group. Among the participants in the control and experiment groups, there was no difference in the average PD for both the groups at the start. However, at the later stage during all the trials, the average PD was significantly larger in the experiment group than in control group during each trial from t_2 to t_4.
Bevilacqua et al. (2016)	Video Games; Mushroom,	Facial Actions (FA) Annotations	*Mushroom Game* FA Annotations at T_0 = 90	The results demonstrate that more annotations of facial actions were made

	Tetris, and Platformer Game	Heart Rate (HR)	FA Annotations at T_1 = 98 HR low at T0 and high *Tetris Game* FA Annotations at T_0 = 110 FA Annotations at T_1 = 159 *Platformer Game* FA Annotations at T_0 = 88 FA Annotations at T_1 = 181 *Hear Rate (HR)* at T_0 = Low HR at T_1 = High HR	during the stressful phase of the games, indicating that participants maintained a neutral expression for a longer period during the boring part. Furthermore, preliminary findings from the collected heart rate data show that participants had a higher heart rate towards the end of the games than at the beginning. The author also states that in the context of the experiment, FA provides an ambiguous foundation for detecting boredom/stressful states when observed at the group level. An individual-level investigation, on the other hand, may yield additional information about the relationship between FA and stress and boredom emotional states.
Bevilacqua et al. (2018)	Video Games; Mushroom, Tetris, and Platformer Game	Mouth Outer (F_1) Mouth Corner (F_2) Eye Area (F_3) Eyebrow Activity (F_4) Face Area (F_5) Face Motion (F_6) Facial COM (F_7)	*Mushroom Game* Percentage of change from H_0 to H_1 is: F_1 = -12.9, F_2 = -15.0, F_3 = -8.9, F_4 = -8.0, F_5 = -11.3, F_6 = 47.2 & F_7 = -12.9 *Platformer Game* Percentage of change from H_0 to H_1 is: F_1 = -7.4, F_2 = -8.2, F_3 = -6.8, F_4 = -4.9, F_5 = -5.9, F_6 = 0.9 & F_7 = -3.6 *Tetris Game* Percentage of change from H_0 to H_1 is: F_1 = -1.5, F_2 = -2.1, F_3 = -2.6, F_4 = -3.3, F_5 = -1.4, F_6 = -11.3 & F_7 = -2.7	The paper proposed a system for automated analysis of facial cues from movies, as well as an empirical evaluation of its potential applicability as a tool for detecting player stress and boredom. The results demonstrate statistically significant variations in the values of the following facial features observed during boring and stressful times of gameplay: mouth outer, mouth corner, eye region, brow activity, and face area. Variations in features face motion and facial COM were not statistically significant. The study's findings support the idea that an automated facial analysis method can be utilized to distinguish between participants' states of boredom and tension.
Giannakakis et al. (2017)	Neutral (N) and Stressful (S) states for: Social Exposure Phase Emotion Recall Phase Stressful Images/Mental Task Phase Stressful Videos Phase	Eye Blink Eye Aperture Mouth Related Features Head Movement Head Velocity Heart Rate	*Social Exposure Phase* Eye blinks: N = 23.9 (14.2), S = 8.8 (5.5); Eye Aperture: N = 453.5 (56.6), S = 513.6 (84.1); Mouth: N = -0.008 (0.004); Head Movement: N = 4.8 (4.2), S = 11.7 (4.4); Head Velocity: N = 0.16 (0.11), S = 0.44 (0.23); Heart Rate: N = 81 (14.7), S = 89.7 (13.3) *Emotion Recall Phase* Eye blinks: N = 18.8 (13.9), S = 25.9 (15.0); Eye Aperture: N = 493.0 (105.2), S = 393.2 (77.0); Mouth: N = -0.011 (0.006), S = -0.008 (0.003); Head Movement: N = 4.7 (3.4), S = 9.6 (8.4); Head Velocity: N = 0.12 (0.06), S = 0.20 (0.12); Heart Rate: N = 70.9 (8.9), S = 74.7 (9.5) *Stressful Images/Mental Task Phase* Eye blinks: N = 26.9 (13.2), S = 26.8 (14.0); Eye Aperture: N = 455.4 (53.1), S = 424.1 (101.1); Head Movement: N = 15.8 (8.4), S = 15.5 (8.1); Head Velocity: N = 0.24 (0.14), S = 0.40 (0.26); Heart Rate: N = 74.7 (9.4), S = 83.0 (10.8) *Stressful Videos Phase*	According to the findings, stress and anxiety increase the blink rate of the eyes. Small, quick movements in the head's amplitude and velocity are also linked to stress. Additionally, the heart rate was found to rise when people were under stress or anxious.

			Eye blinks: N = 24.6 (16.7), S = 23.7 (12.8); Eye Aperture: N = 438.8 (62.4), S = 447.5 (64.8); Mouth: N = -0.009 (0.004), S = -0.014 (0.006); Head Movement: N = 4.5 (2.5), S = 15.0 (13.0); Head Velocity: N = 0.13 (0.06), S = 0.16 (0.09); Heart Rate: N = 72.0 (11.7), S = 74.7 (8.4)	
Current study	Excavation Operations at Construction Site	NASA-TLX Score EDA Values Eye Area (E1) Eyebrow (E2) Mouth Outer (M3) Mouth Corner (M4) Face Area (H5) Head Motion (H6)	*NASA-TLX Score:* LMF = 15.76 (1.75), HMF = 63.41 (5.95) *EDA Values:* LMF = 0.31 (0.11), HMF = 2.21 (1.19) *Eye Area:* LMF = 0.29 (0.07), HMF = 0.43 (0.09) *Eyebrow:* LMF = 6.06 (0.54), HMF = 6.34 (0.84) *Mouth Outer:* LMF = 3.43 (0.41), HMF = 3.88 (0.52) *Mouth Corner:* LMF = 1.44 (0.21), HMF = 1.66 (0.29) *Face Area:* LMF = 9.21 (0.81), HMF = 11.70 (1.85) *Head Motion:* LMF = 5.82 (0.18), HMF = 6.18 (0.32)	The results indicate that there was statistically significant difference in the mean values of all the facial features for low and high mental fatigue. Specifically, the most noteworthy variation was for eye and face area metrics with respective mean differences of 45.88% and 26.9%. The mental fatigue labeling by subjective and objective assessment correlated with each other.

Appendix F. NASA-TLX Questionnaire



Name	Task	Date

Mental Demand — How mentally demanding was the task?
Very Low — Very High

Physical Demand — How physically demanding was the task?
Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?
Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?
Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?
Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?
Very Low — Very High