PREDICTION OF ACUTE ORAL MUCOSITIS AND OVERALL SURVIVAL FOR NASOPHARYNGEAL CARCINOMA PATIENTS WITH RADIATION THERAPY USING RADIOMICS

YANJING DONG

PhD

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

Prediction of Acute Oral Mucositis and Overall Survival for Nasopharyngeal Carcinoma Patients with Radiation Therapy Using Radiomics

DONG Yanjing

A thesis submitted in partial fulfilment of the requirements for

the degree of Doctor of Philosophy

July 2024

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)

DONG Yanjing (Name of Student)

Abstract

Nasopharyngeal carcinoma (NPC) is a malignant neoplasm that arises from the mucosal epithelium of the nasopharynx. In the field of clinical practice, radiation therapy (RT) remains the primary treatment modality for NPC patients. However, with the increasing emphasis on personalized treatment approaches and advancements in artificial intelligence (AI), researchers have started utilizing AI-based cancer imaging analysis, which combines clinical images with AI techniques, to enhance various clinical tasks related to NPC patients. Predicting outcomes, such as overall survival (OS) and incidence of toxicity after RT, for NPC patients can assist clinicians in assessing the risk profile based on tumor characteristics. This information permits identification of patients with poor prognosis, who may benefit from escalated therapy or inclusion in clinical trials. Conversely, patients with a favorable prognosis, if identified in advance, could receive de-escalated therapy, minimizing the physiological and financial burdens associated with cancer treatment. However, toxicity predictions, specifically for the incidence of acute oral mucositis (AOM), are mainly using single sources of data and has limited predictive capability of models. Similarly, studies related to survival prediction models often lack external validation, especially international validation, calling for the development of more generalizable models.

The aims in our studies are focused mainly on two parts, to investigate the impact of radiomics, dosiomics, extracted from multi-regions and multi-sources, and clinical data on the prediction of severe acute oral mucositis in patients undergoing radiotherapy for NPC, and to develop CT-based generalizable prognostic model with perturbation in an international dataset for the prediction of five-year OS of NPC patients following intensity-modulated radiation therapy (IMRT).

For the prediction of AOM, pathological validated NPC patients were retrospectively included from Queen Elizabeth Hospital (QEH) in Hong Kong. Radiomics features (RFs) and dosiomics features were extracted from various volume of interests

(VOIs): Gross tumor volume of primary NPC tumor (GTVp), metastatic lymph nodes area (GTVn), regions of nodal planning target volume with the prescribed dose level of 70Gy (PTVn_70Gy), PTVn with and the prescribed dose level of 60Gy (PTVn_60Gy), using contrast enhanced computed tomography (CECT), cT1-weighted imaging (cT1WI), T2-weighted imaging (T2WI), and dose-volume histogram (DVH) data. Additionally, relevant clinical variables were incorporated into the analysis. Logistic Regression (LR), Gaussian Naive Bayes (GNB) and eXtreme Gradient Boosting (XGBoost) models were developed, considering different combinations of data extracted from distinct VOIs and image modalities. The area under the curve (AUC) of the receiver operating characteristic (ROC) was used to assess the performance of the models. For the prediction of OS, patients were sourced from both a private database in Hong Kong and the publicly available RADCURE database in Canada. RFs were extracted from GTVp in computed tomography (CT) images. Perturbations of the images were employed to select robust RFs. Conventional machine learning models and a multilayer perceptron (MLP) model were developed, integrating the RFs with clinical variables to predict the 5-year OS of NPC patients. The AUC of the ROC was used as a metric to evaluate the performance of the models.

For the first study, the best performing GNB model with 10-fold cross validation AUC of 0.81±0.1 was developed for the prediction of AOM with radiomics and dosiomics features extracted from primary tumor area. For the second study, the best performing MLP score/LR model achieved an internal validation AUC of 0.734 [95% confidence interval (95% CI): 0.765-0.865] and an external validation AUC of 0.735 (95% CI: 0.681-0.783).

Our studies have shown that clinical images hold promise for predicting the occurrence of AOM and OS in NPC patients. In particular, utilizing multimodal data sources has the potential to enhance the performance of the prediction model for AOM. Additionally, deep learning model have demonstrated its effectiveness in handling data from various institutions. These findings provide valuable insights and lay the groundwork for further research in predicting outcomes for NPC patients, particularly in the context of patient screening.

Dedication

I dedicate this thesis to the people who have played a significant role in my life as a student and researcher. Their unwavering support has made my journey towards attaining a Ph.D. degree a reality.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Prof. Jing Cai, for his invaluable guidance and unwavering support throughout my PhD journey. Without his mentorship and kind advice, I would not have been able to embark on this path and successfully complete my research. During the second year of my PhD study, I experienced a significant shift in my research direction, moving from wet lab work focused on nanoparticle synthesis to the dry lab domain of image analysis, as my previous supervisor had left the university. This transition posed a considerable challenge for me. However, Prof. Cai stood by my side, providing continuous support and encouragement, especially during moments when I struggled to grasp the fundamentals of coding for data analysis. Prof. Cai not only granted me the freedom to grow and explore but also encouraged the development of my problem-solving skills and emphasized the importance of collaboration within our research group. His support created an environment that nurtured my creativity in research. His unwavering passion for his career served as a constant inspiration,

motivating me to persevere on my own research journey. I am truly grateful for the precious time I spent in Prof. Cai's research group, and I firmly believe that the experiences and knowledge I gained will continue to accompany me in my future endeavors.

I would like to express my appreciation to my family members and my beloved cat, Lucky. They have provided me with immense comfort and unwavering mental support throughout my daily life. Our regular FaceTime sessions have become a cherished part of my day, creating a sense of warmth and connection despite the physical distance. I am especially grateful to my parents. They have continuously shown their understanding and support for my career choices. I would like to extend a special thanks to my father for encouraging me to explore different perspectives. His belief in me has inspired me to broaden my horizons and strive for new experiences. I am also thankful to my mother for her loving care towards my cat. She is a quite traditional Chinese women, but she always tried her best to understand and support my life choice. Her steadfast backing and efforts to comprehend my choices mean the world to me.

I would like to express my gratitude to my science teacher, Zhang Tianmin, who passed away twenty years ago. One month prior to his departure, he found me playing in the school corridor and took the time to have a conversation with me about the future. he said "You should focus on your studies and venture far away from our hometown". I nodded, feeling confused and intrigued. "Do you know where 'far away' means?" he asked. Perplexed, I shook my head. "Leave here, leave our province, and explore the world." I nodded my head. The journey of self-growth is fraught with numerous moments of breakdown. Often, I find myself plagued with self-doubt, questioning the validity of my choices and whether I possess the capability to achieve my goals. However, in those crucial

moments, during that afternoon in the school corridor, the compassionate yet resolute gaze of my teacher ignites a renewed determination within me. I am profoundly thankful for the trust and the wisdom my teacher shared with me.

I would like to express my deep gratitude to both my previous and current colleagues and classmates. My research was based on the dataset collected by the dedicated members of the radiomics group, the feature extraction tool developed by Dr. Zhang Jiang, and the invaluable experience in model training shared by Dr. Zhang Jiang, Xiong Tianyu, Cao Jin, and Jiarui Zhu. Thanks for the help of paper polishing for my first article by Dr. Lam Saikit. I am truly grateful for their kind assistance and generous patience towards me. It is no small feat to provide guidance to a novice in computer language, especially when everyone is deeply engaged in their own research endeavors. Additionally, I cherish the enjoyable moments spent with my office colleagues during lunch and dinner breaks. These moments of camaraderie and shared experiences have brought a sense of joy and unity to my PhD journey.

I am grateful to my friends for inviting me join them to experience many different and wonderful weekend days. Their companionships have provided me with much-needed respite. I am deeply appreciative of the time spent with my friends, whether it be indulging in delicious meals or engaging in enjoyable games. I am especially thankful to my mindfulness teacher, Sharron, whose eight-week training has proven to be immensely helpful in equipping me with effective tools to manage stress and anxiety in my daily life. This training has been particularly valuable, especially during the graduation season, when pressures can be overwhelming.

I would like to express my sincere gratitude to Hong Kong Polytechnic University for the invaluable opportunities it has provided me. The university has been a catalyst for my academic and personal growth, offering scholarships, a supportive platform, and abundant resources. It has consistently demonstrated its commitment to valuing student feedback and providing exceptional resources and services. In addition, I am also grateful for my first PhD supervisor, Dr. Sixiang SHI, who initially provide me the chance to start my PhD training in the university.

Furthermore, I extend my heartfelt thanks to all the individuals who have shown me love and kindness. Their generous praise and appreciation have served as a constant source of encouragement and have bolstered my confidence on this journey. I am truly fortunate to have received such support and recognition from the people around me. Their contributions have been instrumental in my success and have made a significant impact on my personal and academic endeavors. I am deeply grateful for their continued support, love, and kindness.

Publications

1. **Dong, Y.**, Zhang, J., Lam, S., Zhang, X., Liu, A., Teng, X., ... & Cai, J. (2023). Multimodal Data Integration to Predict Severe Acute Oral Mucositis of Nasopharyngeal Carcinoma Patients Following Radiation Therapy. *Cancers*, *15*(7), 2032.

2. **Dong, Y.**, Zhang, J., Huang, Y., Cao, J., Xiong, T., Li, W., … Cai, J. (2024). CT-based generalizable prediction for overall survival of nasopharyngeal carcinoma patients treated with intensity-modulated radiation therapy: an international-wide multicentre analysis. *Radiotherapy and oncology* (submitted)

3. Du, H., Zheng, J., Li, X., **Dong, Y.**, Cheng, Y., Liu, C., ... & Chen, X. (2023). The correlation between medial pattern of intracranial arterial calcification and white matter hyperintensities. *Atherosclerosis*, 381, 117247.

4. Chen, Z., Wo, B. W. B., Chan, O. L., Huang, Y. H., Teng, X., Zhang, J., **Dong, Y.**, ... & Cai, J. (2024). Deep learning-based bronchial tree-guided semi-automatic segmentation of pulmonary segments in computed tomography images. *Quantitative Imaging in Medicine and Surgery*, 14(2), 1636.

5. Cao, J., Zhou, T., Zhi, S., Lam, S., Ren, G., Zhang, Y., Wang, Y., **Dong, Y.**, & Cai, J. (2024). Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. *Information Sciences*, 662, 120212.

6. Ma, Z.; Zhang, J.; Liu, X.; Teng, X.; Huang, Y.-H.; Zhang, X.; Li, J.; Pan, Y.; Sun, J.; **Dong, Y**., & Cai, J. (2024). Comparative Analysis of Repeatability in CT Radiomics and Dosiomics Features under Image Perturbation: A Study in Cervical Cancer Patients. *Cancers*, 16, 2872.

Conference

1. **Dong, Y**., Zhang, J., Lam, S., Zhang, X., Zhu, J., Liu, A., …& Cai, J. (2023). Predicting acute oral mucositis with data integration methods for nasopharyngeal carcinoma patients. Poster presented at the AAPM 2023 Conference.

2. **Dong, Y.**, Zhang, J., Cao, J., Huang, Y.-H., Li, W., … & Cai, J. (2024). Generalizable radiomics-based deep learning model for the survival prediction of nasopharyngeal carcinoma patients: an international multicenter analysis. Oral presentation presented at AAPM 2024 Conference.

Table of Contents

List of Tables

List of Figures

List of Acronyms

2D	Two-dimensional
3D	Three-dimensional
95% CI	95% Confidence Interval
AI	Artificial Intelligence
AOM	Acute Oral Mucositis
AUC	Area Under the Curve
BMI	Body Mass Index
CAD	Computer Assisted Detection
CCRT	Concurrent Chemoradiotherapy
CECT	Contrast Enhanced Computed Tomography
C-index	Concordance Index
CNN	Convolutional Neural Network
CT	Computed Tomography
cT1WI	Contrast Enhanced T_1 Weighted Imaging
CTCAE	Common Terminology Criteria for Adverse Events
DT	Decision Tree
DVH	Dose Volume Histogram
EBV	Epstein-Barr Virus
EMR	Electronic Medical Records

FDA	U.S. Food and Drug Administration
GAN	Generative Adversarial Network
GLCM	Gray-Level Co-occurrence Matrix
GLDM	Gray Level Dependence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
GNB	Gaussian Naive Bayes
GTVn	Lymph Node Gross Tumor Volume
GTVp	Primary Gross Tumor Volume
HE	Hematoxylin and Eosin staining
H&N	Head and Neck
HR	Hazard Ratio
HU	Hounsfield Unit
IBSI	Image Biomarker Standardisation Initiative
ICC	Intraclass Correlation Coefficient
IHC	Immunohistochemistry
IMRT	Intensity Modulated Radiation Therapy
KM	Kaplan-Meier
LASSO	Least Absolute Shrinkage and Selection Operator
LoG	Laplacian-of-Gaussian
LR	Logistic Regression
ML	Machine Learning

MLP	Multilayer Perceptron
MRI	Magnetic Resonance Imaging
MSC	Mucosa Surface Contour
NB	Naive Bayes
NCI	National Cancer Institute
NGTDM	Neighbouring Gray Tone Difference Matrix
NIH	National Institutes of Health
NPC	Nasopharyngeal Carcinoma
OAR	Organ-at-risk
OM	Oral Mucositis
OS	Overall Survival
PTVn	Regions of Nodal Planning Target Volume
PTVn_60Gy	PTVn with the Prescribed Dose Level of 60Gy
PTVn_70Gy	PTVn with the Prescribed Dose Level of 70Gy
QEH	Queen Elizabeth Hospital
QoL	Quality of Life
RFC	Random Forest Classifier
RF	Radiomic Feature
ROC	Receiver Operating Characteristic
ROI	Region-of-interest
RT	Radiotherapy

RTOG	Radiation Therapy Oncology Group
SEER	Surveillance, Epidemiology, and End Results
SHAP	Shapley Additive Explanations
SNPs	Single Nucleotide Polymorphisms
SVM	Support Vector Machine
T2WI	T_2-weighted Imaging
TNM	Tumor–node–metastasis
VOI	Volume of Interest
WHO	World Health Organization
XGBoost	eXtreme Gradient Boosting

Chapter 1.

Introduction

1.1. Background

Over the past decades, the establishment of standardized scanning protocol for the clinical imaging and the implementation of computer assisted detection (CAD) system for the diagnosis of diseases have provided a solid foundation for the development of quantitative imaging analysis [1]. Currently, in the era of artificial intelligence (AI), which involves computer systems emulating complex tasks performed by humans, such as reasoning, decision-making, and problem-solving, AI-enabled techniques have revolutionized the field of advanced quantitative imaging analysis [2]. Numerous studies have employed AI-enabled techniques to optimize the diagnosis, treatment, and outcome prediction for diseases, especially in oncology field. Among these explorations, radiomics, which involving extraction and selection of imaging biomarkers from clinical images, have emerged as mainstream methods for quantitative imaging analysis, holding tremendous potential in assisting precision treatment and personalized medicine. Radiomics can offer versatile and powerful tools for comprehensive analysis and interpretation of medical images. The application of radiomics in oncology commonly involves two categories based on the final tasks of the study: classification tasks and prediction of clinical outcomes [3, 4]. Classification tasks are focused on dividing a population into categories, such as separating individuals into low or high risk of metastasis [5]. Prediction of clinical

outcomes is based on time-to-event analysis, for example, predict the incidence of lymph node metastasis [6]. The radiomics converts the images into extractable high through-put data, offering noninvasive imaging biomarkers with significant potential to assist the screening, diagnosis, treatment and patient care in clinical practice.

1.2. Imaging Biomarker

According to the National Cancer Institute (NCI), the definition of biomarker is "A biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease" [7]. However, this definition of biomarkers with biological molecules limits their scope to the biological field, overlooking their use in other research areas such as imaging biomarkers and digital biomarkers [8-10]. In 2016, the U.S. Food and Drug Administration and the National Institutes of Health (FDA–NIH) Biomarker Working Group issued a consensus statement to promote the consistency of this terminology. The statement expanded the definition of biomarkers to include "molecular, histologic, radiographic or physiologic characteristics are examples of biomarkers" [11]. These characteristics are considered biomarkers when they are measured as indicators of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions [12].

In contrast to biological biomarkers that rely on blood or Immunohistochemistry (IHC) samples, imaging biomarkers offer a noninvasive approach for oncology patients. Especially for patients undergoing radiotherapy (RT), routine access to medical images such as CT or MRI makes imaging biomarkers highly applicable for individualized treatment. In medical imaging, the two main categories of AI-enabled imaging biomarkers

are hand-crafted radiomics features (RFs). Compared to the IHC, which could only reflect focal part of tumor, the radiomics have the capability to capture the information from entire volume of the tumor for high-throughput data analysis.

1.2.1. Hand-crafted Radiomics Features and Dosiomics

In 2012, Philippe Lambin firstly published a research article which proposed the terminology of radiomics. The radiomics is a standard quantitative image processing and analysis method for medical imaging, which including four steps: data preparation, tumor segmentation, feature extraction and data analysis [1]. The "-omics" indicated the high through-put processing method involved in the radiomics [13]. Researchers from computer science, oncology, and radiology have developed a predefined set of representations (also RFs), generated by specific algorithms, to describe the particular arrangement of image voxel intensities or morphological information of tumors in the clinical images [14]. The theoretical foundation of radiomics is that physiological properties of tissues and gene-related patterns can be captured in image intensity-based values [1]. The radiomics converts the images into mineable high-through put data for data analysis. (See **Figure 1-1**)

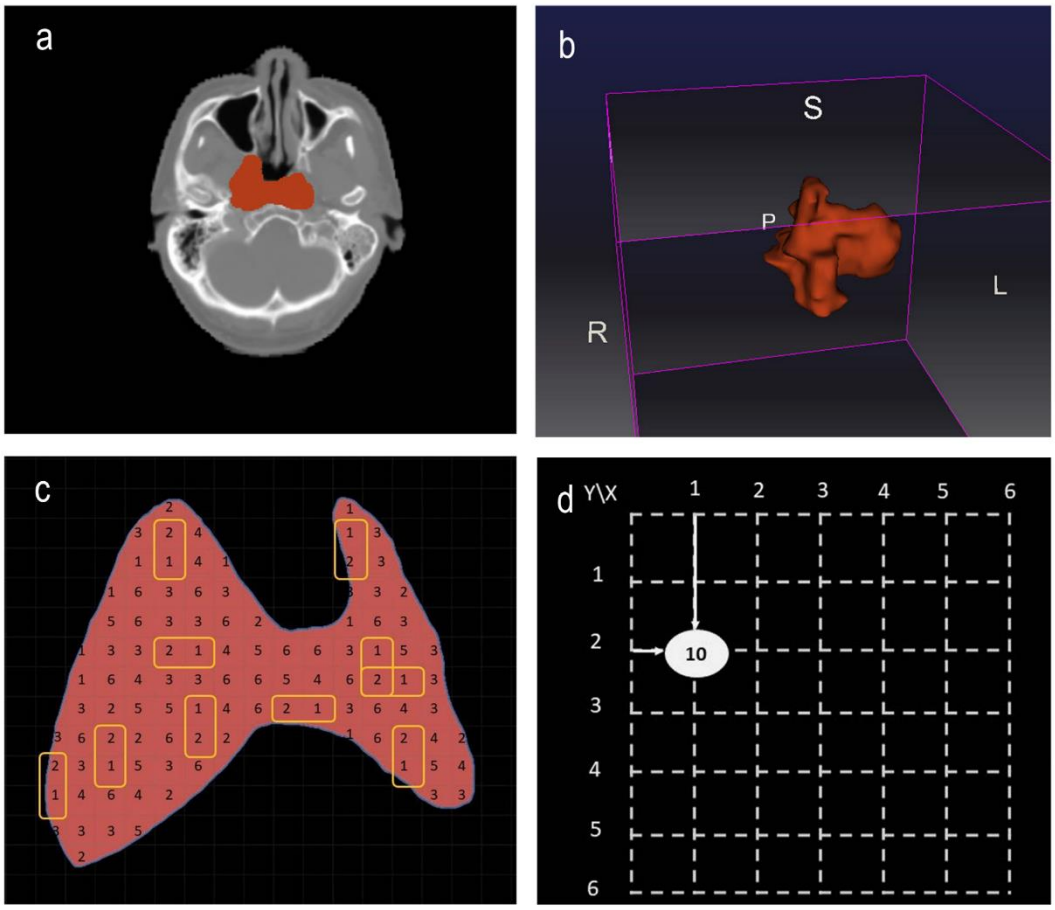

Figure 1-1. Example of calculation of radiomics features. a. Axial CT image with contour mask(red) representing the region of interest (ROI) for an NPC patient; b. Three-dimensional (3D) volume of interest (VOIs) reconstructed from the CT images with corresponding masks, highlighting the primary gross tumor volume (GTVp) area; c. Cartoon example of a two-dimensional (2D) matrix on the tumor mask used for calculating the gray-level co-occurrence matrix (GLCM). The GLCM captures the frequency and spatial relationship of different gray levels occurring together by analyzing the relationships between pixels' grayscale values. In the picture, the grey level of Hounsfield unit (HU) intensity within the tumor mask have been standardized into six grades (one to six). The yellow box encircles the paired pixels with grade one and two. d. Cartoon example of GLCM matrix, in total there are10 paired pixels with grade one and two. The value at coordinate (1,2) in the GLCM matrix is 10, and the symmetric position (2,1) also has the same value (to make the picture clearer, not explicitly indicated in the picture).

The RFs are commonly utilized to quantify morphological characteristics, first-order statistical aspects, and spatial relationships (texture) between voxels, which contain intensity values for medical images. In 2015, Philippe Lambin categorized RFs into three groups: first-order features, second-order features, and higher-order features. First-order features focus solely on the correlation of individual voxels, while second-order features consider the spatial interaction of voxels. For instance, the GLCM feature mentioned in **Figure 1-1** is a second-order features measuring the heterogeneity/disorder by examining the relationship between grey levels in specific neighborhoods. Higher-order features encompass additional filters or mathematical transformations applied to images before feature extraction, providing additional information for data analysis [15]. Since 2020, the Image Biomarker Standardization Initiative (IBSI), an independent international collaboration, has been actively working on standardizing the application of RFs. Currently, the IBSI recommends a total of 167 original RFs (first or second order features) and eight filters they were able to standardized for radiomics analysis. This combination results in a grand total of 1336 filtered RFs (167 original RFs multiplied by eight filters). Filters are commonly employed in radiomics analyses to enhance and quantify clinically relevant characteristics and textures in medical images. These filters can result in the enlargement of important information such as the peritumoral region, blood vessels, contrast agent uptake, degree of calcification, or fibrosis, among others [16, 17].

Dosiomics, introduced in 2017 as an extension of radiomics, involves extracting dose distribution characteristics to develop personalized radiotherapy plans and predict treatment outcomes [18]. By treating patients' 3D dose distributions as spatially and statistically distributed images, parameters like prescription dose, dose distribution, and

dose-volume histogram (DVH) can be utilized for treatment response assessment and prognostic analysis in cancer. Incorporating both radiology and dosimetry features provides comprehensive information that enhances prediction accuracy in tumor radiotherapy [19, 20].

Machine learning (ML) models always used for the prediction in radiomics and dosiomics. In contrast to statistical models, which rely on predefined assumptions and employ statistical inference to estimate model parameters based on observed data, ML models prioritize learning patterns and relationships directly from the data [21]. In other words, while statistical models solely validate predefined assumptions, ML models have the ability to generate personalized predictions based on the acquired patterns. Furthermore, ML models possess greater freedom in capturing intricate patterns and non-linear relationships in the data to address diverse tasks. Conversely, statistical models are typically designed with specific assumptions regarding the underlying data distribution and relationships [22].

In radiomics and dosiomics, after applying filters, thousands of features can be extracted from a single type of image within multiple VOIs. Before utilizing ML models for data analysis, it is imperative to undergo a dimension reduction (feature selection) process to ensure optimal performance. The dimension reduction is the process of searching for a low-dimensional manifold that effectively captures and represents the underlying structure of high-dimensional data [23]. As the number of features/variables in a ML model increases and/or the number of cases in the groups decreases, the likelihood of model overfitting increases. Overfit models have limited generalizability, which undermines their effectiveness. Dimension reduction can be achieved through the use of

several statistical or ML methods, either individually or in combination. These methods include feature reproducibility analysis, collinearity analysis, algorithm-based feature selection, et al. The final number of features after feature selection does not have a specific threshold. However, it is generally recommended to have at least one-tenth of the total labeled data as the minimum number of features to ensure adequate representation and prevent overfitting [24].

Some RFs are sensitive to the changes of segmentation [25]. The feature reproducibility analysis [26], especially for segmentation involving manual contouring by human experts [27]. Additionally, analysis of RFs can be influenced by different acquisition protocols. It is better to conduct the reproducibility evaluation for data with different acquisition protocols[28-30]. The primary objective of the reproducibility analysis is to reduce dimensionality by excluding features with relatively poor reproducibility. The intra-class correlation coefficient (ICC) is a widely used statistical tool for this analysis. It is essential to consider the different types of ICC when conducting the analysis [31].

The definition of collinearity is the non-independency of predictive variables. By keeping the most represent and dependent variables, the collinearity analysis could achieve the dimension reduction for the RFs [32]. Pearson's correlation coefficient is commonly employed to identify redundant features, specifically those exhibiting collinearity. When two radiomic features demonstrate high collinearity, it is advisable to exclude the one with the highest collinearity with the others from the analysis. Some algorithms such as correlation-based feature selection algorithm was designed to solve the collinearity status, such as fast correlation based feature selection algorithm [33, 34]. Except for the

collinearity problem, the algorithm also helps with selection of features with maximum correlation to the classes.

Algorithm-based feature selection is the most widely dimension reduction technique [35]. There are various algorithms with different functions such as random forest classifier (RFC) [36]. The RFC is one of the most popular and successful feature selection method. Random forest classifiers tackle dimensional redundancy by assessing the relevance of features and selecting the most informative ones with threshold settings, while the ensemble of decision trees mitigates the influence of correlated variables, resulting in improved model performance and robustness against redundant dimensions [37].

Aside from the categorize methods of reproducibility, collinearity, and algorithm-based methods, based on the working principles and interaction with the model, some articles also categorize the principal feature selection methods to three groups: filters, wrappers, and embedded methods [38]. Filter methods, wrapper methods, and embedded methods differ primarily in their evaluation criteria. Filter methods use criteria such as correlation coefficients or test statistics, which do not involve any specific learning machine, to assess feature quality. They are independent of specific learning machines, using statistical properties or correlation-based measures to evaluate feature quality. Wrapper methods utilize the performance of a learning machine trained using a given feature subset as the evaluation criterion. They embed feature selection into a specific learning machine, evaluating different feature subsets based on the performance of the trained model. They can employ search techniques but tend to be computationally expensive due to evaluating multiple subsets. Embedded methods integrate feature selection into the model training process, automatically selecting the best feature subset to

enhance model performance. They automatically select the best feature subset to enhance model performance and are performed during the model training process. Details in the **Table 1-1.**

Table 1-1. Advantages and disadvantages of dimension reduction

Types	Methods	Advantages &Disadvantages
Filter	Variance Correlation Chi-square Fisher exact/Odds ratio (OR) T-test/Mann–Whitney U test Mutual information (Information gain)	Generic of methods, not specific to ML algorithm Much faster compared with wrapper Less prone to over-fitting
Wrapper	Forward selection Backward elimination Boruta	On specific ML algorithm High computation time High chance of over-fitting
Embedded	LASSO Random Forest Embedding Ridge regression	Time consuming between filter and wrapper Generally used to reduce the over-fitting

Note: LASSO=Least Absolute Shrinkage and Selection Operator

1.2.2. Machine Learning and Deep Learning

In 1950s, John McCarthy coined the term of "AI" as using computer or machine to simulate the human intelligence and critical thinking [39]. There are numerous subfields of AI, such as ML and computer vision [40]. The definition of ML is a branch of algorithm that are designed to learn complex tasks and develop predictive model [41]. To further optimize the feature engineering steps of ML, the DL was generated [42], details in **Figure 1-2**.

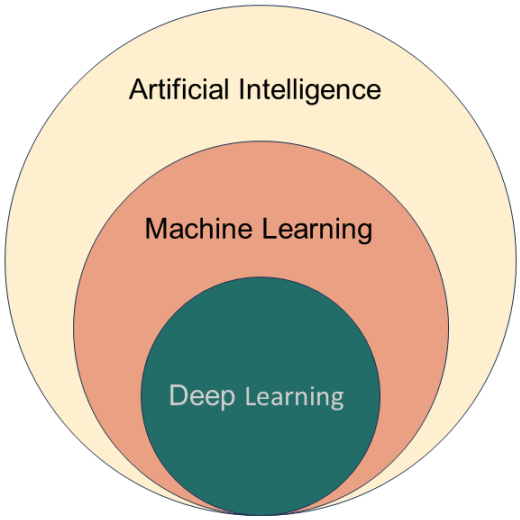

Figure 1-2. Relationship of AI, ML, and DL. Artificial intelligence (AI) aims at stimulating the human intelligence. Machine learning (ML) is a subfield of AI, intends to developing predictive model with sample data. Deep learning (DL) is a subfield of ML, designed to eliminates the feature engineering steps with optimal set of features from sample data.

Many ML algorithms have been applied on the radiomics. Here we would like to introduce some most commonly used ML algorithms in supervised learning [43].

- Logistic Regression (LR): LR is a commonly used probabilistic model for classification prediction. Depending on the context and application domain, the LR can be considered both a statistical model and a ML model. From a statistical perspective, LR is a type of generalized linear model. It involves estimating parameters through methods like maximum likelihood estimation and performing hypothesis testing and inference. In the field of statistics, LR is often used to establish probability models, analyze relationships between variables, and conduct parameter estimation and inference. From a ML perspective, LR is seen as a classification algorithm used to predict binary or multiclass

outcomes. It learns the relationship between features and the labels (also known as target variables) from training data and uses this relationship to make predictions on new data. LR exhibits a tendency to overfit high-dimensional datasets, but it performs effectively when the dataset can be linearly separated. To mitigate the risk of overfitting in such situations, regularization techniques such as L1 and L2 regularization [44] can be employed. However, it is important to note that the assumption of linearity between the dependent and independent variables is widely regarded as a significant limitation of LR [45].

- Naive Bayes (NB): The NB algorithm is based on the Bayes' theorem. NB considers each attribute features as independent feature. Mathematically, Bayes' theorem can be expressed as: $P(A|B) = (P(B|A) * P(A)) / P(B)$. Where: $P(A|B)$ represents the conditional probability of event A occurring given that event B has occurred. $P(B|A)$ represents the conditional probability of event B occurring given that event A has occurred. $P(A)$ and $P(B)$ are the probabilities of events A and B occurring independently. The key benefit of NB algorithm is that, compared to more sophisticated approaches, it needs a small amount of training data to estimate the necessary parameters and make predictions quickly. However, the performance of the NB algorithm may be influenced by its strong assumption of feature independence. Gaussian, Multinomial, Complement, Bernoulli, and Categorical are the common variants of NB classifier. Among these variants, the Gaussian NB (GNB) classifier is suitable to handle continuous features. By utilizing the assumption of Gaussian distributions, it can handle deviations from typical data patterns and still provide reliable classification outcomes [46].

-Decision tree (DT) and random forest: DT are models that use a divide and conquer approach to make decisions based on feature values. They recursively split the input space into smaller regions, where each internal node represents a decision based on a specific feature, and each leaf node represents the final decision. DTs can be combined into ensemble methods. However, they are prone to overfitting and require techniques like pruning and regularization to prevent it. Despite their limitations, DTs are widely used for their simplicity, transparency, and ability to capture non-linear relationships [47]. Random forest is a bagging ensemble model by creating multiple DTs through randomization [36]. It uses bootstrap sampling and variable selection at decision nodes to test various thresholds, capturing interactions and non-linear effects of predictors. The model's predictive performance is evaluated by aggregating the predictions of individual trees, and variable importance measures are used to assess predictor influence [48, 49]. By combining the predictions of multiple DT, random forest tends to achieve higher accuracy. The ensemble approach reduces overfitting and provides more stable and reliable predictions. In addition, random forest provides an assessment of feature importance, allowing insights into which features play a crucial role in the model's predictions. This feature importance evaluation aids in understanding the decision process of the model and identifying key features in the data [47].

- eXtreme Gradient Boosting (XGBoost): XGBoost is a scalable ML system that utilizes tree boosting, demonstrating superior performance and wide-ranging applications. The principle behind the XGBoost model involves employing the gradient boosting tree algorithm, iteratively training multiple DT to optimize the loss function. Each tree is trained based on the residuals of the previous tree, and the final prediction is generated by weighted

summation. XGBoost exhibits resilience to noise and outliers, reducing the risk of overfitting. Additionally, the model provides insights into feature importance, supports custom loss functions and evaluation metrics, and allows for fine-tuning model complexity and generalization. With its gradient boosting strategy and regularization techniques, XGBoost excels in modeling complex data relationships.

In ML, the term "error" refers to the incorrect prediction made by a model [50]. The generalization error, in turn, represents the error that arises when the model is applied to cases that it has not encountered before. In clinical medical applications, it is crucial to minimize the generalization error, as it directly reflects the accuracy of predictions. Evaluating the generalization error of models guides the choice of learning methods.

The commonly used evaluation methods to evaluate the generalization error of model in radiomics are hold-out, cross-validation methods and bootstrapping [51]. Hold-out methods involve directly splitting the dataset into two independent sets: one for training the model and the other for evaluating its generalization. Stratified sampling is often employed to ensure appropriate distribution of data between the training and test sets [52]. In addition, the cross-validation technique divides the dataset into k similar sub-datasets, ensuring the preservation of data distribution. In this method, k-1 sub-datasets are utilized for training the model, while the remaining sub-dataset is employed for testing its performance. This process is repeated k times, yielding k distinct train-test groups. Through k-fold cross-validation, the model undergoes training and testing k times, ultimately allowing for the calculation of the mean results [53]. Furthermore, bootstrapping is a resampling technique used to estimate population parameters by repeatedly sampling with replacement from the original dataset to create multiple simulated datasets. It enables the

estimation of sampling distributions and confidence intervals for statistics without relying on strict assumptions about the underlying population distribution. The cross validation and bootstrapping are recommended for radiomics study. In 2017, Philippe Lambin introduced a radiomics quality score for assessing the quality of radiomics studies. As part of this scoring system, the inclusion of cross-validation contributes an additional point to the overall score [54].

The essential idea of DL approach is from the study of artificial neurons, which was initially proposed in 1943 as a model for how neurons process information in the biological brain [55, 56]. In a neural network, the input (input including various types of data, such as numbers, text, images, audio, and more) is fed into an input layer, which then transmits its computed value to one or more hidden layers that are connected to an output layer. Each layer consists of nodes, also known as "units" or "features", connected through edges to the previous and next layers. These units transform the data in a nonlinear manner by applying an activation function. A deep neural network typically comprises multiple hidden layers, sometimes exceeding 100. During the training process, deeper layers in the network can combine high-level features or representations algorithms can automatically generate features that are well-suited for addressing the specific task at hand [57].

In recent decades, various DL neural networks are developed, such as multilayer perceptron (MLP), convolutional neural networks (CNN), generative adversarial networks (GAN) and more. Among these techniques, MLP is the most frequently used neural network, especially for small datasets [58]. MLP, a fundamental feedforward neural network, comprises multiple layers of neurons (input, hidden, and output) and is employed for classification and regression. It trains via the backpropagation algorithm, adjusting

inter-neuron connection weights to minimize the discrepancy between predicted and actual outputs. The structure and parameters of MLP can be adjusted based on specific tasks and data. For instance, different activation functions, the number and size of hidden layers, and regularization techniques can be chosen to control the model's complexity and generalization ability. Compared with other DL models, the MLP has shallow structure and fewer parameter, which can provide good performance without requiring large number of training datasets. The output layer of MLP often utilizes an activation function (such as sigmoid) to map the outputs to a range of 0 to 1, representing the probability of the positive class. These outputs could be used as model prediction results. Alternatively, they can be directed outside the network and processed by different analyzers, such as a support vector machine (SVM) or DT, for additional analysis and classification.

1.3. Nasopharyngeal Carcinoma

1.3.1. Background

Nasopharyngeal carcinoma (NPC) is a tumor that originates from the epithelial lining of the nasopharynx. The nasopharynx is anatomically defined by the posterior choanae at the front, the declivity of the sphenoid bone above, and the clivus and the first two cervical vertebrae at the back. The soft palate forms the lower boundary. The nasopharynx is well-supplied with blood vessels and has an extensive lymphatic drainage system [59]. Locally advanced NPC often involves the parapharyngeal space and the carotid artery, which significantly increases the risks associated with surgical intervention (Details in **Figure 1-3**.) [60]. The initial stages of NPC often manifest with minimal

discomfort, posing a challenge for early detection. Late-stage diagnosis and the anatomical location of the tumor make surgical intervention in NPC patients particularly challenging.

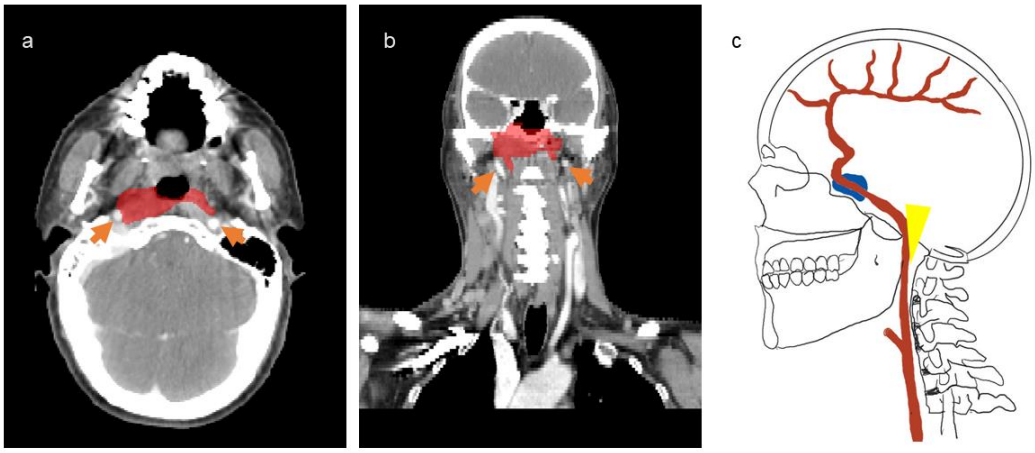

Figure 1-3. Location description of nasopharyngeal carcinoma in relation to the internal carotid artery. a. An axial CT image of one nasopharyngeal carcinoma (NPC) patient. The tumor (red area is the gross tumor volume) is close proximity to the internal carotid artery (orange arrows). b. Coronal CT image of one NPC patient. c. Cartoon illustration of sagittal image of the internal carotid artery (red) travel into the brain superiorly, and passed through the cavernous sinus (blue). The internal carotid artery is the major blood vessel that supplies oxygenated blood to the brain. The growth of tumor might invasion or compression the internal carotid artery.

In 2022, there were 120,216 new cases and 73,476 death of NPC reported globally[61]. The majority of NPC patients are concentrated in South China and Southeastern Asia, with China alone accounting for approximately 50% of the global NPC population. The age-standardized incidence rate of NPC per 100,000 person-years in China and Southeastern Asia was eight, significantly higher compared to the combined rate of 2.7 in Europe and America [62]. Due to its association with the specific geographical region, NPC is also commonly referred to as "Guangdong cancer."

1.3.2. Diagnosis and Treatment

Clinical evaluations and medical imaging play a vital role in raising suspicion of NPC, while the ultimate confirmation depends on the result of pathological histology examination, which is the gold standard for the conclusive diagnosis. The staging of NPC involves several components, including a thorough medical history, physical examination with cranial nerve assessment, complete blood count, serum biochemistry (including liver and renal function tests and lactate dehydrogenase), nasopharyngoscopy, and radiological imaging. In particular, computed tomography (CT) and magnetic resonance imaging (MRI) are standard imaging examinations routinely recommended for the diagnosis of NPC. Among the two modalities, MRI is considered the most accurate and recommended for visualizing small mucosal thickening, parapharyngeal and masticatory space involvement, retropharyngeal lymph node involvement, as well as skull base and cranial nerve infiltration. However, the MRI has long scanning time, higher cost and requirements for patients without metal implants. Conversely, CT is particularly adept at identifying bone destruction of the osteogenic skull base, lower cost, and easier to recognize small suspicious metastatic lymph nodes [63, 64].

The gold standard of the diagnosis of NPC is the result of pathological histology. In 2005, the World Health Organization (WHO) revised the classification of NPC into three categories: WHO I, which refers to Keratinizing squamous cell; WHO II, which denotes non-keratinizing carcinoma; and WHO III, which represents Basaloid squamous cell (Details in **Figure 1-4**) [30]. Among these three histological types of NPC, the non-keratinizing carcinoma is the most commonly observed, accounting for over 95% of cases in high incidence areas and approximately 75%-87% of cases in low incidence areas. [65].

The non-keratinizing carcinoma is sensitive to the radiation. Considering the anatomical localization of the tumor, the main treatment method for the NPC patients is the radiation therapy.

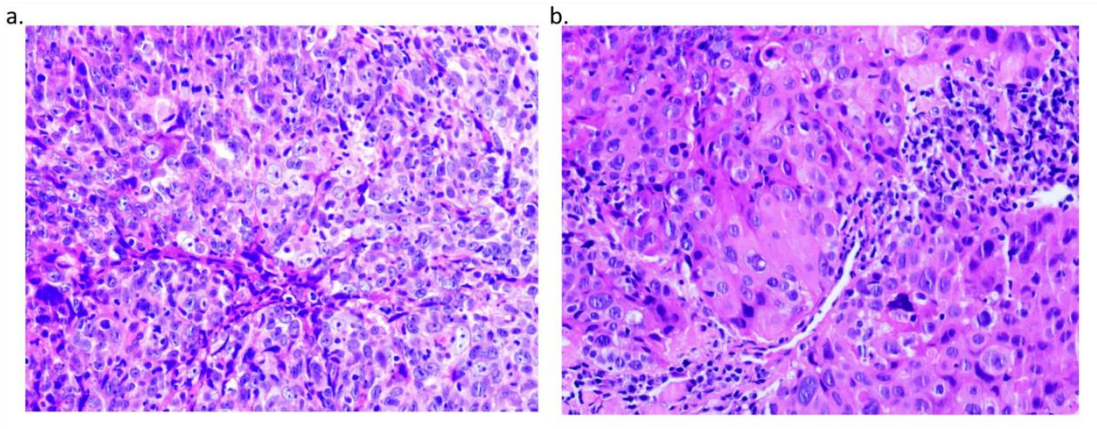

Figure 1-4. Hematoxylin and eosin stain samples for nasopharyngeal patients.

a. Non-keratinizing carcinoma, magnification 200x; b. Keratinizing squamous cell carcinoma, magnification 200x. The WHO type III is too rare to collect from the hospital.

According to the ESMO-EURACAN clinical practice guidelines in 2020 [63], the intensity-modulated radiation therapy (IMRT) is the primary treatment for NPC with or without chemotherapy. In two-dimensional (2D) conventional radiation therapy, radiation is delivered uniformly and constant without any differentiation as it penetrates through the human body. This can result in exposure of organs surrounding the tumor, also known as organs at risk (OARs), to radiation without any protection. The IMRT is an advanced three-dimensional (3D) radiation therapy technique that allows for the generation of complex 3D dose distributions, closely conforming the high dose of radiation to the target volume and reducing the radiation on OARs. IMRT utilizes computer algorithms to optimize the beam intensity (fluence) to be precisely shaped and delivered. These algorithms take into account not only the dimensions of the target and normal tissues but also user-defined constraints

such as dose limits to targets and critical organs. By incorporating these considerations, IMRT enables precise and personalized radiation treatment planning, improving the therapeutic outcome while minimizing radiation-related side effects [66, 67] . The 5-year overall was increased from 75% to over 80% for NPC patient following IMRT in some studies[68-70].

1.3.3. Survival and Toxicity of IMRT for NPC Patients

The underlying mechanism of radiotherapy is based on the ability of ionizing radiation to damage cancerous cells, ultimately resulting in their death. It is essential for clinicians to actively manage and control the spread of radiation to OARs. Even though with the highly accurate dose controlled IMRT technique, the incidence of the toxicity symptoms is inevitable for most patients. The primary toxicities associated with IMRT include oral mucositis (OM), dysphagia, dermatitis, xerostomia, and others. Among these reactions, OM is commonly experienced by the majority of NPC patients. The incidence of acute oral mucositis (AOM), symptoms lasting up to three months after IMRT, is approximately 90% [71].

Various criteria, such as those developed by the WHO, Radiation Therapy Oncology Group (RTOG), and the common terminology criteria for adverse events (CTCAE), have been established to evaluate the severity of AOM (Details in **Table 1-2**) [72-74]. These criteria are stratified patients with toxicity reaction to four or five grades and provided appropriate management methods for each grade of patients. Although there may be some discrepancies between these criteria, studies have shown that similar results

are obtained regardless of which criteria are used when dividing the grading system between ⩽grade 2 (mild AOM) and ⩾grade 3 (Severe AOM) [75].

As we have mentioned before, prediction of severe AOM for NPC patients with IMRT could help the clinical trials stratification of high risk severe AOM patients for the investigation of novel effective treatment for severe AOM. Currently, there is not a standard guidance for physicians on how to protect the OARs. The prediction model of severe AOM also could provide specific tips for clinicians to draw the RT plan contours for patients with high risk of severe AOM. This information enables the implementation of advanced protective measures and optimizes the allocation of limited medical resources.

Table 1-2. Oral Mucositis scales under various criteria

	Grade 0	Grade I	Grade II	Grade III	Grade IV	Grade V
WHO	No findings	Erythema without ulcers	Erythema and ulcers	Liquid diet	Tube feeding	NA
RTOG	No findings	Erythema without ulcers	Erythema and ulcers	Liquid diet	Tube feeding	NA
CTCAE.v4	No findings	Erythema without ulcers	Moderate pain or ulcer, not interfering oral intake	Severe pain, interfering oral intake	Life threatening consequence	Death related to toxicity

Abbreviation: World Health Organization (WHO), Radiation Therapy Oncology Group (RTOG), and the common terminology criteria for adverse events (CTCAE).

Overall survival (OS) is the primary endpoint in oncology clinical trials, and it is defined as the duration from the time of randomization or diagnosis to the occurrence of death from any cause. In oncology clinical trials, OS is commonly regarded as the gold standard endpoint, given the general objective of cancer treatment to prolong survival. It is worth emphasizing that OS is a patient-centered measure, as it directly captures the ultimate event of death, rendering measurement relatively uncomplicated. Moreover, OS is an objective measure with minimal susceptibility to researcher bias [76].

For the NPC patients with IMRT, prediction of OS plays a significant role in pretreatment risk assessment, risk-adapted clinical trials, and clinical decision-making scenarios. It helps estimate survival outcomes based on patient-specific factors, enabling informed treatment decisions. Survival prediction models aid in designing trials that target specific risk groups, optimizing treatment approaches. They also guide decisions such as escalating or de-escalating treatment strategies based on predicted risks, improving patient outcomes. Incorporating survival prediction enhances care by personalizing treatment strategies and improving overall effectiveness.

Currently, for the prediction of incidence of AOM, most studies only used single or double modal of traditional clinical data. The risk factors for AOM involving genomic background, tumor microenvironment and the radiation dose distribution. No study had considered the multiple sources of background information for the prediction.

For the prediction of OS of NPC patients with IMRT, most of the studies are restricted on centers in China. The lack of involvement of international cohorts hinders the evaluation of generalizability and robustness of the prediction models. Additionally, the current treatment decisions and prognosis of NPC heavily rely on the tumor-node-

metastasis (TNM) staging system. However, despite patients with NPC receiving similar treatments at the same stages, outcomes can vary significantly. The introduction of advanced modality of RT and treatment have increased both the survival and quality of life (QoL) for NPC patients. The dosage management in NPC is too precise to allow for significant further optimization. To further optimize the outcome for NPC with RT, it is necessary to find novel ways and new biomarkers that can accurately to predict the survival and toxicity in the context of the current treatment. Innovative strategies must be developed to manage cancer patients more effectively.

1.4. Aim and Objectives

To address the aforementioned challenges, this thesis aims to develop AI-based models utilizing multimodality data or representative international datasets for the prediction of survival and prognosis of NPC patients with IMRT. The objectives are as follows:

1. To investigate the impact of radiomics, dosiomics, extracted from multi-regions and multi-sources, and clinical data on the prediction of severe acute oral mucositis in patients undergoing radiotherapy for NPC.
2. To develop CT-based generalizable prognostic model with perturbation in an international dataset for the prediction of five-year OS of NPC patients following IMRT.

Chapter 2.

Literature Review

2.1. Acute Oral Mucositis (AOM) for NPC Patients Following IMRT

The AOM is typically characterized by atrophy, swelling, erythema, and ulceration, significantly impacting the functional status and QoL of patients. The initial soreness of AOM can lead to difficulties in opening the mouth, resulting in reduced food and water intake, weight loss, and increased risk of systemic infection. Patients experiencing severe AOM with intense pain may require a reduced dose of chemotherapy, and some may even consider discontinuing the RT regimen. The presence of severe AOM can further worsen patient morbidity, ultimately leading to a decline in QoL and increased mortality rates [77]. Currently, advanced RT machine (3D) have achieved highly precise and controlled delivery of radiation for the treatment of NPC patients. Effective management and prevention of toxicity events are crucial for improving the survival and QoL of NPC patients.

Apart from its symptomatic impact, the effective treatment of severe AOM is rare. Furthermore, the management of AOM result in a significant health economic burden. On one hand, few studies provide high levels of evidence to support the effective treatment of severe AOM. Carryn et al. conducted a phase IIb clinical trial, using manganese-containing macrocyclic complex and placebo to manage the severe AOM for head and neck (H&N) patients with concurrent RT and cisplatin. The trial demonstrated a decrease in the incidence of severe AOM from 65% to 43%, with a corresponding reduction in the severity

rate from 30% to 16% [78]. While the manganese, as a heavy metal ion, has the potential risks of long-term accumulation and toxicity effects, such as the defect of eyesight. Additionally, it is important to note that these treatment approaches may not be applicable to all patients undergoing RT [79]. On another hand, the presence of AOM is associated with various healthcare utilization factors, including increased rates of hospital admission, prolonged hospital stays, visits to the emergency department, a higher number of clinic visits, consultations with nutritionists, increased use of supplemental feeding options, as well as greater usage of opioids and antibiotics. These additional healthcare requirements and interventions contribute to the overall cost of managing AOM [80]. The incremental cost of AOM among patients receiving radiation therapy was estimated to be approximately $5,000 to $30,000. Between 15% and 55% of patients undergoing radiation for H&N cancer require enteral feedings. The management of these patients often involves the placement of feeding tubes. The cost associated with feeding tube management is approximately $30 per day, and it is typically required for a duration of four to six months. This represents a significant financial burden for both patients and healthcare systems. [81]. Early intervention is crucial for effectively managing the clinical burden of severe mucositis. However, there is still a lack of clarity regarding the individuals who are at a higher risk of developing severe symptoms. Consequently, there is a clear and unmet need to identify patients who are prone to experiencing severe mucositis. This identification is essential to implement methods of risk minimization and provide targeted supportive care. By identifying patients who are more likely to experience severe mucositis, clinicians can customize interventions and deliver personalized care to mitigate the impact of this side effect.

2.1.1. Risk Factors for AOM of NPC Patients with RT

Previous studies have primarily concentrated on analyzing specific types of data, such as genetic information or clinical variables, to predict the severity of oral mucositis. Multiple factors have been identified as correlating with the occurrence and severity of oral mucositis, including genetic background, tumor microenvironment, radiation therapy dose distribution, chemotherapy drugs, and nutritional status, et al. As for the genome studies, several of them have demonstrated a connection between the genome and AOM. However, the underlying mechanism of the incidence and severity of AOM in patients with NPC undergoing RT remains unclear. For instance, Li et al. conducted a study where blood samples were collected from 114 NPC patients prior to RT. They utilized polymerase chain reaction-restriction fragment length polymorphism to detect single nucleotide polymorphisms (SNPs). The study identified a correlation between SNPs (194Arg/Trp and 399Arg/Gln) in the X-ray cross-complementing group 1 gene and the occurrence of acute skin reactions and AOM. However, the findings were solely based on statistical investigation and lacked external or internal validation experiments to provide further support. Similarly, Yang et al. [82] recruited a large-scale of 1467 NPC patients using genome-wide association analysis. They discovered that patients carrying the minor alleles of rs117157809 were more prone to developing severe AOM, with a per allele odds ratio of 3.72 [95% confidence interval (95% CI): 2.10-6.57, $P=6.33\times10^{-6}$]. Rs117157809 is a locus in the Tankyrase gene, which is involved in protein coding. The study did not delve into the underlying pathway connecting the gene to AOM. Furthermore, a recent two-stage genome-wide association study [83] indicated that four SNPs might be associated with acute mucositis. Unfortunately, the results were not validated in the subsequent validation

stage. It is evident that, at this moment, relying solely on the variables of clinicogenomics is insufficient for accurately predicting the incidence, correlations, and severity of AOM in NPC patients after RT.

Aside the genomic background, the tumor itself may have the impact on the AOM [84]. The location and stage of tumor are associated risk factors for the AOM [85, 86]. The innate immunity is essential to the pathogenesis of oral mucositis [87]. Mahesh Kudrimoti et al. used the dusquetide to treat the severe OM patients with chemoradiation therapy. The dusquetide is a synthetic peptide drug with immunomodulatory and tissue healing properties. The drugs modulated the tumor microenvironment and achieved 50% decreased on the median duration of severe OM, which might also indicated that the tumor microenvironment might correlated with the AOM [88].

Other clinical correlated factors also contribute to the incidence and severity of AOM, such as the RT dosage, weight loss, body mass index (BMI), and concurrent chemotherapy, et al. A study [89] involving 92 patients with NPC who underwent IMRT demonstrated a correlation between severe AOM and a radiation dose exceeding V30 Gy to the oral mucosa. Additionally, the study identified weight loss as another factor associated with severe AOM. Furthermore, additional studies by Saito et al.[90] and Rosenthal et al.[91] have reported that a low BMI is a risk factor for severe AOM. Moreover, the administration of additional chemotherapy, particularly with certain agents known to be associated with AOM, such as alkylating agents and antimetabolites, can further exacerbate the occurrence and severity of AOM.

Although mucositis risk is influenced by the intricate interplay of the host, tumor microenvironment, and treatment specifications, a significant majority of studies employ

hypothesis-driven, single-candidate approaches [92]. These approaches focus on investigating individual factors or variables, driven by specific hypotheses, rather than considering the comprehensive interaction of multiple factors.

2.1.2. Radiomics and Dosiomics Models for Predicting AOM

In addition to genomics information, routine clinical data for patients with NPC undergoing RT include contrast-enhanced computed tomography (CECT), MRI, and dose files. Dosiomics, following the idea of radiomics, is a quantitative method that extracts high-throughput information from the dose files of RT patients. By leveraging intensity, textural, and shape-based features, it enables the parameterization of the dose distribution in specific VOIs. This approach provides a comprehensive understanding of the dose distribution, allowing for a detailed characterization of the RT treatment and its impact on the targeted areas. Importantly, dosiomics goes beyond traditional DVHs by describing the dose distribution at a high complexity level, capturing intricate details that are distinct from the information obtained from DVHs [18, 93]. Both radiomics and dosiomics are two quantitative methods which could provide minable texture and dose-distribution information for clinical prognosis prediction.

Numerous studies have demonstrated that the RFs could reflect the heterogeneity and microenvironment of tumor [94, 95]. Most of them selected receiver operating characteristic (ROC) curve and area under the curve (AUC) to evaluate the model performance. The AUC is a numerical value ranging from zero to one. A higher AUC indicates better model performance [96]. Few studies have indicated the association of RFs and AOM. Jiang et al.[97] developed two Light Gradient Boosting models to predict the severity of AOM. The training and internal validation AUC of the models are 0.8 and

0.85, respectively (Conference poster). However, it is worth noting that the grading cut-off in their study was grade⩾2, which may have overestimated the severity of AOM. The study reported that 26 RFs were significant for the prediction. Satheeshkumar[98] investigated the capability of ML model to predicting the incidence of mucositis for cancer patients with chemotherapy. The best performing model with testing AUC is 0.79. This study did not show the cancer types of patients and not involve the severity of AOM.

As an extension of radiomics, the dosiomics analyze the variance in dose distribution within 2D and/or 3D DVHs. Dean et al.[99] and Orlandi et al.[100] selected oral cavity and parotid gland as VOIs. The study demonstrated that the severe AOM is associated with the mean dose of parotid gland and the high dose received by small oral cavity. Meanwhile, Li et al. [101] extracted dosiomics features from oral cavity contour (OCC) and mucosa surface contour (MSC). They selected significant dosiomics features with least absolute shrinkage and selection operator (LASSO) and developed LR model to predict the incidence of severe AOM. The body mass index and retropharyngeal lymph node area irradiation status also included into the model development. The training and validation AUC of models for OCC and MSC are 0.737 and 0.767. The finding is concordance with Orlandi's studies. Many research studies have mentioned about using oral cavity as VOIs for AOM prediction. While the specific definition and delineation methods for oral cavity are different in these studies.

2.1.3. Multimodal Data Integration

The risk factors associated with the incidence and severity of AOM are multifaceted, and the contributions of these factors are complex. In clinical practice,

clinicians should integrate information from multiple sources to make informed decisions regarding the diagnosis and treatment of cancer patients. Traditional experiences with single or double sources of data have been validated to have limited prediction power for acute AOM. Integration of complementary data from multiple types of datasets can lead to an intricate outcome than a simple summation of information [102]. Integration of multimodal data from multiple sources, for instance, clinical, radiomics, and dosiomics for NPC patients, has the potential to overcome the boundaries of conventional medical analysis [103, 104].

In recent years, several multimodal data integration methods have been developed. Based on the time or stages in which data are integrated, the fusion methods including early fusion, intermediate fusion and late fusion. Early fusion involves integrating all data resources before model development. Intermediate fusion combines several types of data to generate multiple models, which are then aggregated for final analysis. Late fusion trains the model with each data resource separately and then interprets the results for the final prediction. [104]. Early fusion requires a certain level of alignment or synchronization between modalities, such as integrating wide range of electronic health records [105]. Intermediate fusion is more flexible, as it can first integrate similar modalities of data and then combine other information into the evaluation system [106]. Late fusion allows for variance in modalities, such as fusing radiomics data with clinical data [107].

The etiology of AOM in NPC patients is complex, and previous research has identified dose distribution as a significant risk factor associated with its incidence. Some previous studies have used the oral cavity as ROIs, but there is a lack of consensus on the contouring guidelines for the oral cavity. Manual contouring of the oral cavity or oral

mucosa in radiotherapy plans is a labor-intensive task, particularly when dealing with large datasets. Furthermore, it is important to acknowledge that the incidence of severe AOM can be influenced by factors beyond the nature of oral cavity, such as tumor heterogeneity, the tumor microenvironment, and the radiotherapy plan itself. Therefore, to improve the prediction of severe AOM, it is crucial to consider heterogeneity information within tumor regions and their surrounding areas, in addition to the oral cavity. Dosiomics, a high-throughput data analysis technique that focuses on dose distribution, can provide valuable information for predicting AOM. Additionally, incorporating other clinical data, such as BMI and body weight loss, can contribute to a comprehensive assessment.

Unfortunately, there is a limited number of studies exploring the potential of radiomics or dosiomics for AOM prediction. Most existing studies simply incorporate clinical data into radiomics or dosiomics models to predict the incidence of AOM. To the best of our knowledge, no studies have investigated whether the fusion of data from multiple regions and modalities could enhance the predictive capabilities for severe AOM. Therefore, there is a need for further research to explore the potential of multi-region and multimodality data fusion in enhancing the prediction of severe AOM in NPC patients. Such investigations could provide valuable insights into the complex relationship between dose distribution, tumor characteristics, and the incidence of AOM, ultimately improving patient management and treatment outcomes.

2.2. Overall Survival for NPC Patients following IMRT

Before we start to introduce the predication of OS for NPC patients, it is better to understand the word “precision medicine”. The precision medicine, also known as

stratified medicine or targeted therapy, refers to the practice of tailoring treatment to a specific subpopulation of individuals who exhibit variations in their susceptibility to develop a particular disease or response to a specific medicine [108]. In the past, traditional medicine commonly employed a uniform drug treatment for all patients diagnosed with a specific disease. However, this approach resulted in various issues. For example, some patients with the same diagnosis exhibited different responses to the identical treatment, while others experienced severe side effects from the treatment. Precision medicine, on the other hand, takes an alternative approach by stratifying patients into subpopulations, allowing for more precise sub-treatments tailored to their specific needs. This personalized approach to treatment aims to achieve greater effectiveness in patient care [109].

In the era of precision medicine, prediction of OS for cancer patients plays a crucial role in tailoring medical treatments to individual needs [110]. By utilizing prediction models, patients can be stratified into high-risk and low-risk groups, facilitating efficient resource allocation by prioritizing treatments for those who stand to benefit the most [111, 112]. Moreover, prediction of OS offers valuable insights for patients with disease backgrounds or family histories of specific illnesses, empowering them with a better understanding of what they may encounter in the future. Beyond its impact on treatment decisions, prediction of OS also holds significant benefits for researchers in terms of biomarker discovery and disease understanding [113, 114]. Through the use of prediction models, researchers can explore the correlations between different variables and clinical outcomes, thereby identifying novel biomarkers associated with specific diseases. Furthermore, these models enable the longitudinal monitoring of disease progression, allowing for the identification of critical time points and patterns of progression [115].

As we have mentioned before, the OS is a primary endpoint for cancer treatment. The definition of OS is that the duration from diagnosis with a specific disease to the end of death. In both clinical practice and trials, the primary objective for clinicians and researchers is to extend OS by striking a balance between the various interventions and standard care for patients. The measurement of OS is 100% objective and can be calculated without bias from researchers. Even though alternative survival metrics such as progression-free survival , disease-free survival and biochemical recurrence are proposed as potential surrogates for OS, the enhancement of OS remains widely recognized as the gold standard for drug approval by regulatory agencies in the United States and Europe [116].

Prediction of OS is extensively utilized in oncology research to derive precise probabilities that assist in clinical decision-making. A variety of models can be employed for the prediction of OS. Traditional Cox proportional hazards model is widely utilized as survival prediction model for survival analysis. But this model cannot cope with high-dimensional data. While precision medicine is based on the big data analysis, AI, various omics disciplines and integration of these information. In recent decades, the application of ML and DL models in survival prediction tasks has yielded impressive results, demonstrating strong model performance [117].

Numerous methods and metrics are available to evaluate the performance of prediction models, including the AUC, concordance index (C-index), accuracy, and others [115]. These methods could basically be categorized into two types: discrimination and calibration. Discrimination refers to the model's ability to accurately classify individuals into two classes: high-risk and low-risk. Calibration, on the other hand, measures the extent

to which the model's predicted probabilities align with the actual outcomes for each individual. These two aspects reflect different dimensions of prediction models. It is commonly recommended that discrimination is a more representative measure compared to calibration metrics, as recalibration is always possible to improve calibration, whereas discrimination may not be easily adjustable [118].

The most commonly used discrimination measurement is the ROC curve and AUC [119, 120]. ROC and AUC are most commonly used evaluation criteria for radiomics model [121]. ROC is a parametric plot that illustrates the relationship between the true positive rate and the false positive rate, showcasing the performance of a binary classification system as the decision threshold is adjusted across the entire range of a continuous forecast variable. The AUC is a metric that quantifies the overall performance of a model by measuring the area under the ROC curve, serving as a measure of the model's capacity to effectively distinguish between positive and negative cases and its generalization ability [122]. When evaluating a model, AUC values in the range of 0.7-0.8 are considered good, while values between 0.8-0.9 are considered excellent. AUC values above 0.9 are regarded as outstanding [123].

The C-index is also popular for the survival analysis. Instead of relying on absolute survival times for each instance, the C-index considers the relative risk of an event occurring across different instances [124]. When come up with the binary outcome task, the C-index is equal to the definition of AUC [125]. The C-index ranges from 0.5 to 1. The 0.5 represents random prediction and the 1 indicate perfect predictive ability of model. Aside from the model evaluation, hazard ratios (HR) are also a commonly used statistical measure to evaluate the risk factors in the Cox regression model [126]. The HR is derived

from regression coefficients in Cox regression model which indicated the relative risk or hazard of an event occurring between two groups or conditions. When the hazard ratio is equal to 1, it indicates that there is no difference in the hazard rates between the compared groups. A hazard ratio greater than 1 suggests an increased relative risk or hazard for the event. Conversely, a HR less than 1 corresponds to a decreased relative risk or hazard [127].

2.2.1. Conventional Prediction of OS for NPC Patients

The conventional prediction models of OS could be categorized into mainly three groups: statistic models, ML models, and DL models. Among which, the parametric Kaplan Meier (KM) and semi-parametric Cox regression are most commonly used statistic models for the prediction of OS. In these articles, the KM method and Cox regression analysis are frequently used together. The KM method is a univariate approach that allows for the estimation of survival probabilities over time, without considering the influence of other variables. On the other hand, Cox regression analysis is a multivariable approach that takes into account multiple factors simultaneously to assess their impact on survival outcomes [128]. The prediction of OS is frequently associated with the identification of prognostic or predictive factors in cancer research. Prognostic factors encompass a set of variables collected prior to treatment initiation that have the capability to forecast patient outcomes. These factors possess the ability to predict clinical outcomes irrespective of the treatment patients may undergo. On the other hand, risk factors are variables that can be influenced by the treatment itself and often exhibit quantifiable variations in response to the treatment [129].

Many studies have utilized the prediction of OS to investigate risk factors in NPC patients undergoing various treatment methods. Kwok et al. employed the Cox regression

model for multivariate analysis to identify independent risk factors associated with the prediction of 3-year OS in NPC patients experiencing local failure [130]. Their findings revealed that advanced initial T stage (HR: 1.44; $p < 0.006$) was a significant risk factor for local failure in NPC patients. Gokce et al. [131] conducted a ten-year OS study involving 58 NPC patients from a single institute. The study demonstrated no significant difference in OS based on staging ($p = 0.92$); however, males with WHO II pathological type exhibited a higher risk of metastasis ($p = 0.037$; HR = 4.132; 95% CI: 1.09-15.66). Similarly, Xiao et al. [132] conducted a study involving 299 NPC patients to investigate the correlation between various clinical variables and 5-year OS. Their findings suggested that age and gender were independent risk factors for NPC patients. The 5-OS of male patients is 70.7% compared to female patients at 94.1% ($p < 0.001$). Male patients under the age of 45 had lower 5-OS (66.8%) than same age of female (91.2%), $p = 0.008$.

KM and log-rank tests were utilized to describe patient survival, while Cox regression analysis was employed to analyze the correlation of clinical variables with 5-year OS. Additionally, Lulu Zhang et al. [133] conducted a retrospective study involving 8093 nasopharyngeal NPC patients from a single hospital to investigate the correlation between clinical variables and OS. The study employed Cox regression analysis, KM modeling, nomogram development, and evaluation metrics such as the AUC and C-index. The findings indicated that age, T stage, N stage, Epstein-Barr virus (EBV) DNA levels, lactate dehydrogenase levels, and Albumin levels were prognostic factors for the 3-year OS of NPC patients. The best performing model achieved a C-index of 0.716 (95% CI: 0.698-0.734) and an AUC of 0.717 (95% CI: 0.698-0.737) compared to the conventional TNM staging system, which had a C-index of 0.643 (95% CI: 0.624-0.661) and an AUC

of 0.643 (95% CI: 0.623-0.662) for OS prediction. However, it is important to note that the study mentioned in the previous statement was limited to a single institute.

Numerous studies also have investigated the prognostic factors for the OS in NPC patients. Liu et al. [134] conducted a survival analysis involving 83 NPC patients treated with IMRT, considering variables such as gender, age, stages, pathological types, and metastasis. The study suggested that metastasis significantly influenced OS; however, it is important to note that only six patients had distant metastasis. In another study by Li et al. [135], 202 NPC patients with metastasis were recruited, and the study identified T-N stage and metastasis in a single organ as prognostic factors for OS prediction. Randomized studies have been conducted to investigate optimal treatment methods for NPC patients under various conditions. Anthony and colleagues conducted a phase III randomized study involving 350 patients [136]. Patients were randomly assigned to either the RT alone group or the concurrent chemotherapy with RT (CCRT) group. KM and log-rank tests were utilized for 5-year OS analysis, while Cox regression was adjusted for T stage, age, and other factors to control for study variance. The 5-year OS rates for the RT and CCRT groups were 58.6% and 70.3%, respectively. The HR for CCRT was 0.71 (p = 0.049), suggesting the potential of CCRT for locoregionally advanced NPC patients. Conducting a randomized study has advantages in controlling treatment method variance and reducing bias. Similarly, in 2022, Yuan Zhang and colleagues published a paper [137] introducing the randomized phase III trial with 480 NPC patients for the prediction of 5-year OS. Patients were assigned to CCRT alone (n=238) and CCRT with induction chemotherapy (n=242) groups. KM, log-rank test and Cox regression were utilized for survival analysis. The study indicated that the induction chemotherapy before CCRT improved the OS of

NPC patients. The OS of CCRT with induction chemotherapy had higher OS of 87.9% than that for CCRT alone 78.8% (HR, 0.51; 95% CI: 0.34-0.78]; P= 0.001).

These articles utilize the KM model to describe the survival status, incorporating bivariate comparisons through the log-rank test to assess statistical significance between compared groups. The Cox regression model is used to examine the impact of risk factors and covariates on survival probabilities. The survival outcomes of cancer patients are impacted by a multitude of factors. However, most of these conventional studies are restricted in the single institute or single country with limited patient number. Only clinical variables from EMR are collected for the analysis, which naturally contains limited information related to the disease. What is more, the KM method, log-rank test and Cox regression model have their own disadvantages for the survival prediction. Specifically, the disadvantages of Cox regression: 1) The Cox regression model assumes that the hazard ratio between two groups remains constant over time. If this assumption is violated, the results can be biased. 2) The model has limitations in modeling complex interactions between covariates. It assumes that the effects of covariates are multiplicative on the hazard scale. 3) The model requires the assumption of linearity. It assumes a linear relationship between covariates and the hazard function. In addition, the disadvantages of KM and log-rank tests are: 1) The KM estimator and log-rank test assume that censoring is non-informative. However, if censoring is informative (e.g., related to the survival outcome), the estimates may be biased. 2) The KM model also lacks the covariate adjustment. The KM estimator and log-rank test do not directly account for the effects of covariates. They only compare survival curves between groups without considering other factors that may influence survival. 3) The KM estimator and log-rank test may have lower statistical power

compared to more advanced survival analysis techniques. They are less sensitive to detect more subtle differences in survival between groups, especially when sample sizes are small or event rates are low.

2.2.2. Machine Learning and Deep Learning for the Prediction of OS

With the advancement of -omics research, the landscape of outcome prediction has grown increasingly intricate. Numerous studies now integrate radiomics, clinical, and other data types to facilitate prognostic predictions. Traditional statistical models may not be adept at handling big data analysis, which entails high-dimensional inputs with intricate interrelationships among diverse data elements. Meanwhile, ML models have demonstrated superior predictive accuracy when confronted with high-dimensional datasets. ML models have found widespread application in prediction of OS within the field of oncology [138-140].

Multiple studies have employed the ML models for the prediction of OS. In 2024, Dan Hu et al. [141] recruited 420 NPC patients form one hospital to develop three ML models (LR, DT, random forest) for predicting 5-year OS. The study developed a best-performing random forest model with internal validation AUC at 0.753 (95% CI: 0.609, 0.896). In 2020, Chunyan Cui et al. [142] generated ML model (XGBoost, GNB, random forest and DL et al.) to predict the 5-year OS of MR images from 792 NPC patients. The model yielded internal validation AUC at 0.796. Similarly, Rou Jiang et al. [143] investigated 347 patients with NPC and synchronous distant metastases. Employing 30 hematological markers and 11 clinical characteristics, the study aimed to predict the two-year OS of these patients. The best performing model achieved a training AUC of 0.761

and an internal validation AUC of 0.633. However, due to an imbalanced label distribution, the resulting ROC curve for the best performing model exhibited a piecewise-linear trend. Melek Akcay and his colleagues [144] also included 72 NPC patients and applicated six models, LR, artificial neural network (ANN), XGBoost, SVM, random forest, and GNB with clinical variables for the prediction of OS. The best performing model was GNB with best AUC of 0.91. The patient number in this study is too small to be representative. The model has no 10-fold cross validation and external validation for a reliable evaluation. Changchun Lai et al. [145] collected clinical variables and blood biomarkers to predict the OS of NPC patients. They recruited 519 NPC patients from one hospital and separated them into train and validation groups with the ratio of 2:1. The best C-index of the prognostic model in training and internal validation group are 0.786 (95% CI: 0.728–0.844) and 0.697 (0.612-0.734), respectively. It is important to note that these studies solely included patients from a single institution, without external validation data to estimate the generalizability of the models.

Some studies have attempted to incorporate more diverse datasets for prediction purposes. Rasheed Omobolaji Alabi et al.[146] using NPC patients from the Surveillance, Epidemiology, and End Results (SEER) database, which involved 1,094 patients for training and internal validation, with an additional 517 patients for temporal validation. Furthermore, the study included 60 patients from Helsinki University Hospital for external validation. XGBoost models were developed using only clinical variables, achieving the best external AUC of 0.76. However, it should be noted that the SEER database primarily consists of American patients and may not be directly applicable to NPC patients in Southeast China, where the majority of cases occur. Additionally, the external validation

sample size of 60 patients is relatively small compared to the training dataset of over 1,000 patients. In "-omics" family, the genomics is also a powerful tool for the prediction of OS in cancer patients. In 2023, Yiren Wang et al. [116] examined 33 samples (18 samples for training and 15 samples for external testing from two distinct databases) using a Cox regression model to predict the OS of NPC patients [116]. The best AUC achieved was 0.769 (95% CI: 0.716-0.823) for the prediction of 3-year OS. However, genomic studies often face challenges related to sample accessibility. Genomic examinations can be invasive and expensive, particularly when employing advanced technologies such as single-cell sequencing. Furthermore, many genomic studies lack external validation and have limited sample sizes, highlighting areas for improvement in future research. Radiomics, on the other hand, involves extracting informatics information related to gene and protein expression from routine clinical imaging examinations. It holds great potential for generalizable application without adding significant financial burden.

Among the studies mentioned above, ML models such as LR, SVM, random forest, GNB, and XGBoost are frequently employed. However, models like LR and SVM may face challenges in effectively capturing non-linear relationships within the data. Their limitations lie in their inability to handle complex non-linear patterns. In contrast, random forest and XGBoost, which are ensemble tree models, demonstrate superior performance in handling non-linear relationships. Nevertheless, for highly intricate non-linear relationships, these models may necessitate deeper trees or a greater number of weak learners to accurately model the complexities present.

DL as a subset of ML, excels in capturing intricate non-linear relationships within data. Some studies have applied DL model for OS prediction of NPC patients, especially

related to the image data. Mengyun Qiang et al. [147] included 3444 locoregionally advanced NPC patients from Guangdong province, China in four hospital. 1838 patients for training, 787 patients for internal validation, 257, 248 and 314 for three external validations. 3D-CNN was used for MRI pattern recognition. The results of MRI scores then concatenate with clinical variables into XGBoost model for the prediction of 5-year OS. The AUC for the model in three external validation institutes were 0.728 (0.618 to 0.838), 0.837 (0.752 to 0.922) and 0.853 (0.789 to 0.917). The patient number included in the study is larger enough than that in previous other studies. While these patients are all from single southeast area of China. The interpretation of DL models, particularly those related to CNN studies, remains a challenge in the field. Later, Lianzhen Zhong et al. [148] included 1872 NPC patients to investigate the efficiency of induction chemotherapy for NPC patients with CCRT. The study extracted MRI patterns with ResNet model then using fully connected neural network to fuse the MRI patterns to MRI score. The MRI score was then concatenated with clinical variables into a nomogram for the prognostic prediction of NPC patients. The study only reported the C-index of progression free survival (Internal validation: 0.733, 95% CI: 0.657-0.809 and external validation:0.681, 95% CI: 0.568-0.793) without that for OS. In addition, only T3N1M0 NPC patients with CCRT or ICT plus CCRT were included in the study. Other stages of NPC patients and patients with other treatment such as adjuvant chemotherapy were not included. The selection bias might influence the representative of models.

Predicting OS for NPC patients is challenging due to the complex interactions among multiple factors that influence the survival outcomes in cancer patients. It is not feasible to accurately predict OS using a single algorithm or modeling method alone.

Regarding the aforementioned studies, it is noteworthy that many of them primarily focus on clinical variables, which offer limited information for the intricate prediction of cancer patients' survival. Additionally, a significant proportion of these studies lack external validation and solely rely on internal validation. While ML methods have demonstrated potential in enhancing our comprehension of cancer progression, it is crucial to incorporate an adequate level of validation before considering their integration into routine clinical practice [149]. DL models have demonstrated the ability to effectively handle complex high-throughput data with impressive performance. However, these models often necessitate a large amount of input data for training, and their interpretability remains a challenge.

2.3. Thesis Overview

This thesis will first review previous literatures on radiomics prediction of AOM and OS for NPC patients with RT in Chapter 2. In the next chapter, which focuses on the first objective, the prediction of AOM with the multimodal data will be explained in detail. The fourth chapter contains the study for the prediction of OS for NPC patients based on three international datasets (Two private and one open accessed datasets). Lastly, we summarize the entire thesis by revisiting the main results and discussing the limitations and future developments.

Chapter 3.

Acute Oral Mucositis

3.1. Introduction

As we have mentioned in the chapter 2, we intended to using radiomics, dosiomics and clinical data for the prediction of the severe AOM for NPC patients with IMRT. In this part, we will introduce the whole research method for the study, from data collection, image preprocessing, feature extraction to model construction. We first developed a conventional clinical LR model, and then used the model selected variables to combine with the radiomics and dosiomics data for the construction of other two ML models. We compared two data integration methods: 1) integrating radiomics and dosiomics before selection by the RFC, or 2) passing radiomics and dosiomics data separately to the RFC and then combining the results with clinical data for the prediction.

3.2. Methods

3.2.1. Patient Data

All patient data were retrospectively collected from NPC patients who underwent CCRT or RT at Hong Kong Queen Elizabeth Hospital (QEH) from 2012 to 2015. Informed consent of patients was waived due to the nature of the retrospective study. NPC patients were recruited based on the following inclusion and exclusion criteria. The inclusion criteria were: (1) NPC patients with pathological validation and absence of distant metastasis and co-existing tumors of other types at diagnosis, (2) patients treated with a

total RT dose of 60–70 Gy, and (3) patients with a completed set of clinical, image, and radiation dose files. The exclusion criteria were: (1) patients aged less than 18, (2) patients without original image or clinical data, and (3) patients for whom exact standard CTCAE evaluation for AOM had not been recorded. Symptoms in grades 1 and 2 were defined as mild AOM, and grades 3 and 4 as severe AOM accordingly [150]. All the patients were negative of oral mucositis according to the CTCAE grading system before receiving radiation therapy. Clinical variables included (1) treatment information: TNM stage, treatment, past health condition, allergy history, vision condition, hearing condition, and CTCAE evaluation for AOM; and (2) demographic data: age, gender, body weight, height, BMI, and the smoking and drinking habits. All clinical variables were acquired one week before RT commencement, except the CTCAE evaluation results, which were recorded 4–5 weeks after RT commencement. The equation for BMI is defined as follows:

$$BMI = weight/height^2$$

Patients were maintained in a supine position during the imaging examination.

CECT. All patients were scanned in the CT stimulator (16-slice Brilliance Big 1 Bore CT, Philips Medical Systems, Cleveland, OH) at Queen Elizabeth Hospital in Hong Kong. The scanned regions were from vertebra to five centimeters (cm) below the sternoclavicular notch. The contrast-enhanced images were acquired at 30s after intravenous injection of 70 milliliter (mL) iodinated contrast agents. The detailed acquisition parameters are listed in **Table 3-1**.

Table 3-1. Acquisition parameters of CECT.

Parameters	CECT
Pitch	1
Voltage (kVp)	120
Exposure (mAs)	250–350
Slice thickness (mm)	3
Matrix	512 x 512
Scan time (s)	15

MRI. MRI scans were acquired with an MR scanner (1.5 Tesla, Siemens Avanto, Germany). T2-weighted imaging with short-tau inversion recovery (STIR) sequence and T1-weighted imaging with spin-echo MRI sequence were applied. The detailed acquisition parameters are listed in **Table 3-2**.

Table 3-2. Parameters of T2WI and T1WI.

Parameters	T2-STIR	cT1WI
[TR]/[TE] (ms)	7640/97	739/17
FOV (cm^2)	24 x 24	24 x 24
Number of acquisitions	1	1
Slice thickness (mm x slices)	4 x 25	3 x 28
Spacing (cm^3)	0.75 x 0.75 x 4.4	0.938 x 0.938 x 3.3
Matrix	320 x 320	256 x 256

Radiomics has been successfully applied to a range of imaging techniques, including CT, MRI, ultrasound, and others. In a hospital setting, even for the same type of imaging modality, different scanning machines and reconstruction methods may be used

to cater to various needs. This issue is not limited to multi-scale institutes but also exists within individual hospitals [151, 152]. To ensure the quality of the collected images, we employ a visualization process wherein all the data obtained from the hospital is carefully examined. This allows us to identify and exclude images that exhibit significant noise or artifacts, as these can adversely affect the consistency of radiomics analysis.

3.2.2. Imaging Preprocessing

As previously mentioned, radiomics is based on the quantitative analysis of pixel HU values in medical images. Image pre-processing plays a significant role in ensuring the accuracy and standardization of radiomics [24]. The main steps involved in image pre-processing are: 1) size of the pixel or voxels; 2) number of the gray levels [153]; 3) range of gray level values [154]; 4) nonuniformity of signal intensity for MRI [155, 156]; Numerous methods can be used for the imaging pre-processing. Details in **Table 3-3**.

Table 3-3. Commonly used image pre-processing methods

Pre-processing steps	Methods
Size of the pixel or voxels	Liner and cubic B-spline interpolation [157]
Number of the gray levels	±3 sigma normalization [158]
Range of gray level values	Discretization with fixed bin size and bin number [159]
Remove the nonuniformity of signal intensity for MRI	N3 and N4 bias field correction algorithms [160]

In this study, the imaging pre-processing steps were based on our previous work [161] and are in accordance with the IBSI guidelines [16]. Specifically, (1) voxel size resampling: all images (CECT and MRI) were resampled to a voxel size of $1 \times 1 \times 1$ mm3 ; (2) VOIs

re-segmentation: CECT images were normalized to confine the value of HU to [–150,180] to eliminate the non-soft tissue influence in the VOIs; (3) image filtering: a Laplacian of Gaussian (LoG) filter with three levels of Gaussian radius parameter was used under fine (1 mm), medium (3 mm), and coarse (6 mm) scales; (4) quantization of gray levels: gray-level intensities of the images were fixed to 50 bins; and (5) inhomogeneity correction of image pixel value: N4B bias correction in the "N4 Bias Field Correction Image Filter" in SimpleITK (v1.2.4) was implemented, in particular, to MRI images.

3.2.3. Feature Extraction

Feature extraction was performed using our in-house platform RADAR based on publicly available SimpleITK (v1.2.4) and PyRadiomics (v2.2.0) [162, 163]. All VOIs were delineated by an experienced senior clinician [164]. The gross tumor volumes were contoured based on CECT with the assistance of MRI images.

Radiomics

The GTVp and the gross tumor volume of nodal lesions (GTVn) were selected as the main VOIs for radiomics feature extraction. Features with or without LoG filters were both involved. All these features were extracted from CECT, contrast-enhanced T1 weighted (cT1WI), and T2 weighted (T2WI) images (Parameter details in **Figure 3-1**). Meaning of each VOIs for different image modalities were listed in **Table 3-4**.

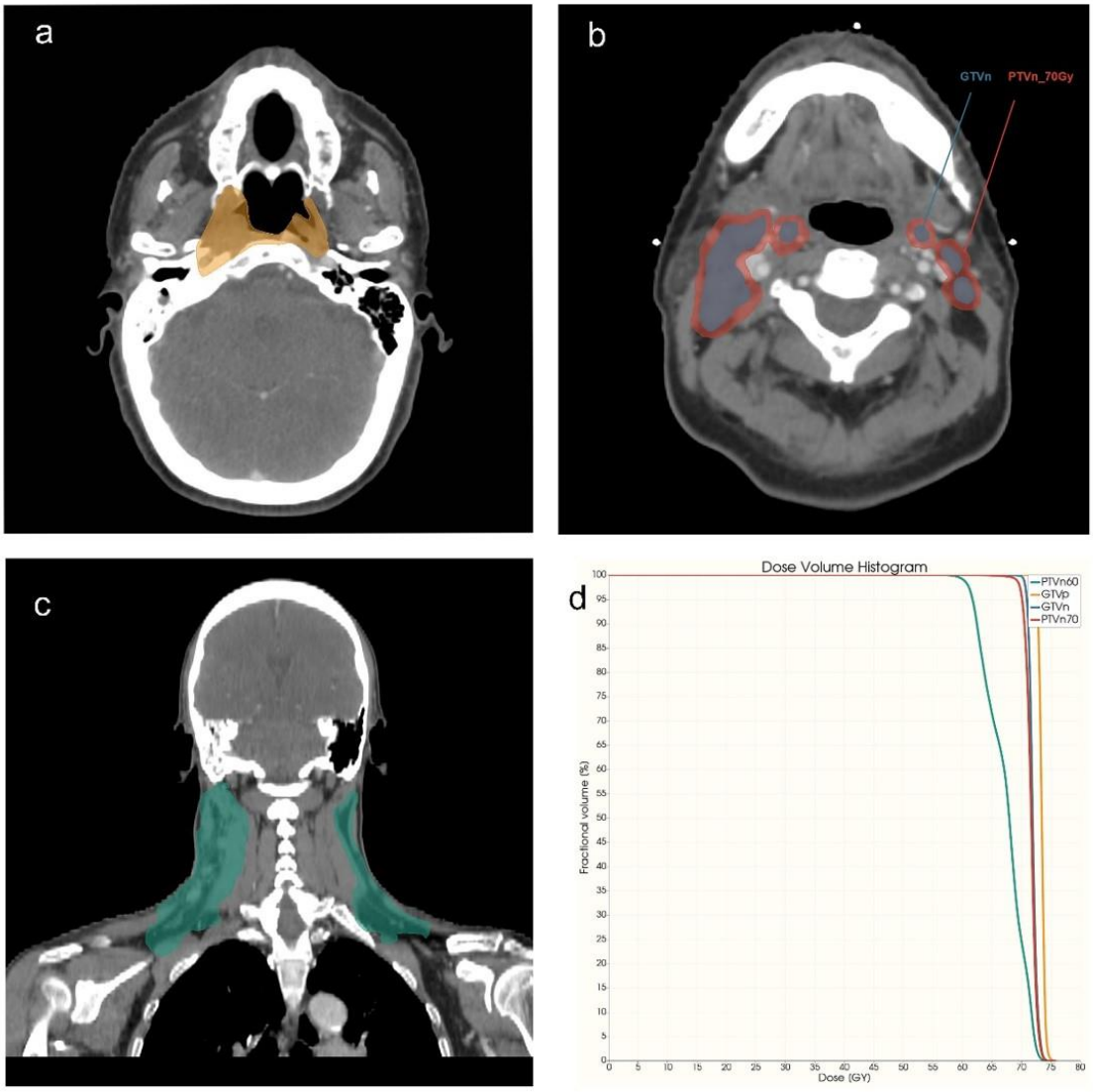

Figure 3-1. VOI examples for NPC patients with CECT examination. (a) Region of GTVp (orange), axial view. (b) Region of GTVn (blue) and PTVn_70 Gy (red), axial view. (c) Region of PTVn_60 Gy (green), coronal view. (d) DVH curve of four VOI. Three categories of radiomics features (RFs) were extracted: shape, first-order statistics, and texture features. The texture features can be further categorized into gray-level difference matrix (GLDM), GLCM, gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), and neighboring gray-tone difference matrix (NGTDM) classes.

Table 3-4. VOIs and image modalities.

VOIs	Descriptions of VOI Imaging	Modalities/Images
GTVp	Gross tumor volume of primary NPC tumor	CECT, cT1WI, T2WI, DVH
GTVn	Metastatic lymph nodes area	CECT, DVH
PTVn	Regions of nodal planning target volume	
PTVn_70Gy	PTVn with the prescribed dose level of 70Gy	DVH
PTVn_60Gy	PTVn with the prescribed dose level of 60Gy	DVH

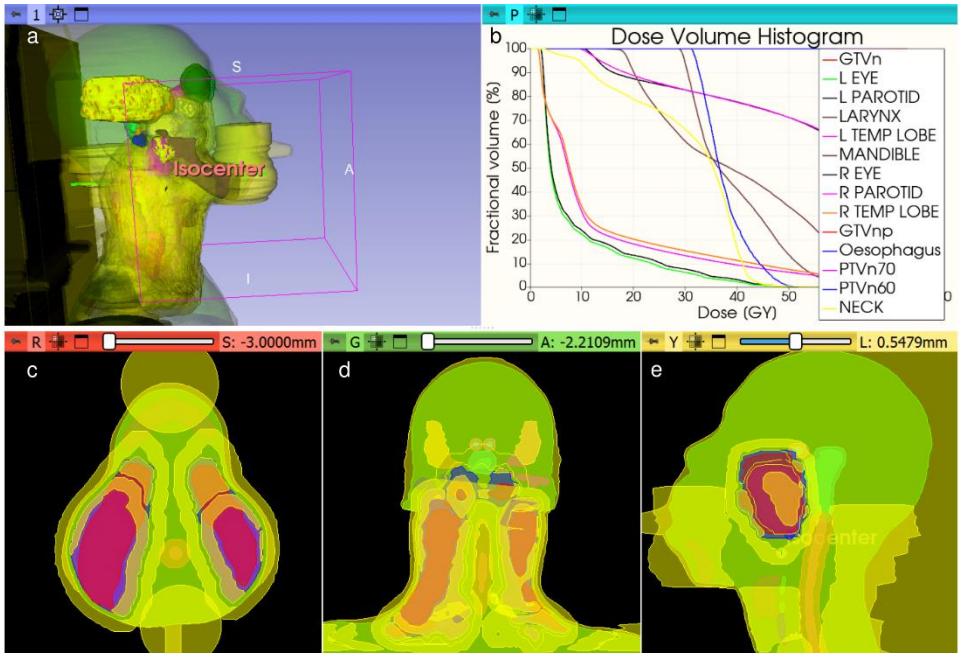

Figure 3-2. Dose maps of NPC patients undergoing IMRT. (a) 3D view of NPC patient. (b) DVH of multiple VOIs. (c) Axial view of patient in different VOIs. (d) Coronal view of patient. (e) Sagittal view of patient.

Dosiomics

Except for GTVp and GTVn, the region of the high-dose nodal planning target volume (PTVn_70Gy) and region of the low-dose nodal planning target volume (PTVn_60Gy) were also added to the dosiomics analysis (Details in **Figure 3-1, Figure 3-**

2 and **Table 3-2**). 2D and 3D DVHs of each studied VOIs were extracted from dose files for dosiomics feature extraction. All dosiomics features were extracted based on Gabry et al.'s previous study [18]. Features that reflect dose distribution, for instance, mean dose, spatial dose gradient, and spatial dose spread were extracted accordingly. All the calculation algorithms have been listed in a previous publication by Buettner et al. [165].

3.2.4. Model Development and Evaluation

Based on the training process with or without label (prediction tasks) information, the ML algorithms applied in radiomics can be classified into three groups: supervised, unsupervised, semi-supervised learning [43]. In supervised learning, the algorithm utilizes label ,for example the treatment outcomes of patients, a s input to predict future instances [166]. It should be noted that not all patients in the real world have label information. Unsupervised algorithms analyze unlabeled datasets independently, without human intervention. Semi-supervised algorithms combine both supervised and unsupervised analysis simultaneously. In our study, the grading of AOM is the label of the task. It's classified into two categories. The input label information enables the supervised ML model to make accurate prediction and complete model evaluation with the guidance of labels. Therefore, here we only introduce the background information related to the supervised tasks. The most common supervised tasks are "classification" that separates the data, and "regression" that fits the data.

In this study, the statistical analysis, model training, and evaluation were conducted in Jupyter 6.4.12 and SPSS 25. The ROC curve and AUC with 10-fold cross validation was performed to evaluate model performance. The CTCAE grade scale of patients in mucositis was dichotomized between severe AOM (grade ≥ 3) and mild AOM (grade ≤ 2) as the

prediction endpoint. Patients were stratified based on CTCAE grade to training and testing groups at a 7:3 ratio (Details in **Figure 3-3**).

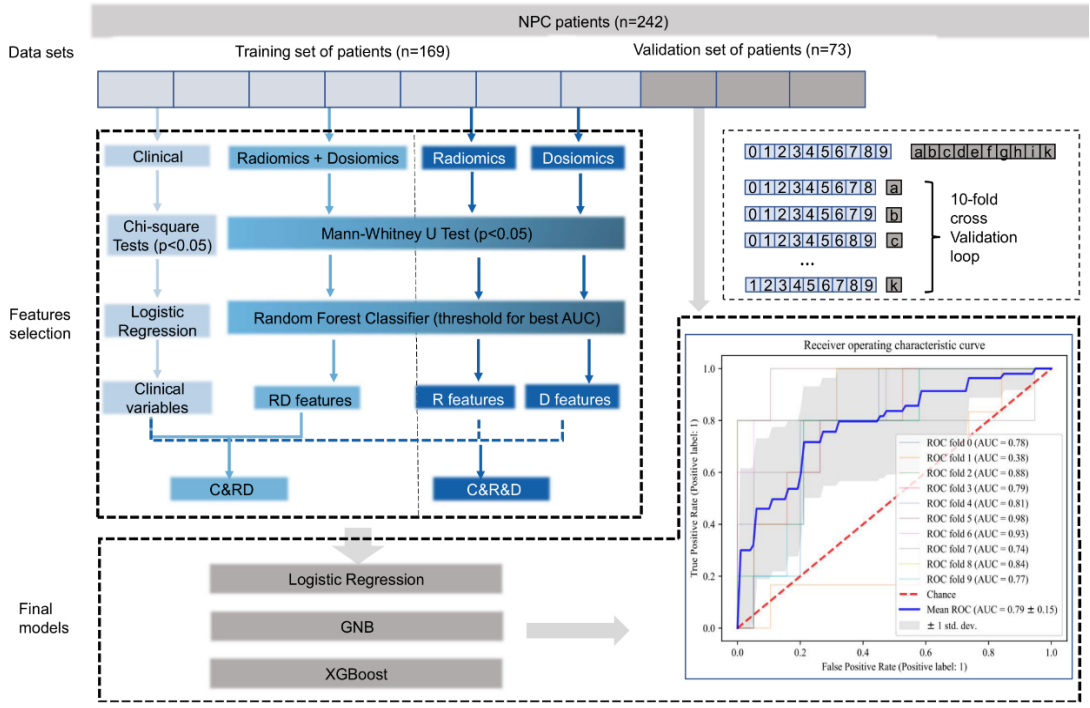

Figure 3-3. Scheme of feature selection and modeling. Training and validation sets were separated before data analysis. The training set of data was used for feature selection. The validation set of data was used for model evaluation. To further manipulate the numerical and categorical data, reduce the interactions, and solve the collinearity problems, random forest selection was applicated for radiomics, dosiomics, and integrated data. Three linear or non-linear models were developed with independent validation data sets with selected features. The AUC was set as the main evaluation method for the model performance. performance.

Models trained with Single-modal data

The single-modal data sources in this study were restricted to single type of data (radiomics, dosiomics, or clinics), single modality of imaging (CECT, cT1WI, T2WI, and DVH), or single region of patients (GTVp, GTVn, PTVn_60Gy, and PTVn_70Gy). Each single data source experienced two steps: (1) feature selection and model training in the

training dataset and (2) AUC evaluation in validation datasets. For clinical data, chi-square and Mann–Whitney U tests were employed for binary and non-binary variables for univariate analysis. *p* values < 0.05 was considered to be statistically significant. All radiomics and dosiomics data were standardized with the MinMax scaler before selection. For radiomics and dosiomics data, we first identified significant features between severe and mild AOM patients in the training set with Mann–Whitney U tests. After that, random fores was used to rank the importance of the significant features considering both feature interactions and nonlinearities. The optimal feature number was set according to the best RF training model score. Three models, including LR, GNB, and XGBoost, were applied to evaluate the combined predictive value of these selected features in the independent validation set. All VOIs data were analyzed separately at the single model stage.

Models trained with multi-modal data

The multi-modal data sources in this study indicated the dataset including equal to or more than two single-modal data sources. Specifically, clinical variables after multivariant analysis (LR) with *p* value < 0.05 were selected for data integration. Dosiomics and radiomics data from different VOIs and image modalities were integrated with two methods: (1) dosiomics and radiomics data were combined together before feature selection and (2) the features selected after the RFC were merged and directly combined without a further feature-selection step (please refer to **Figure 3-3** for more details). All the data-integration methods are listed in **Table 3-5.** Shapley Additive Explanations (SHAP), an explainable AI-based tool, was applied for further explanation of feature importance for the model with the best AUC result and specific features [167]. The feature importance method in tree ensemble algorithms like XGBoost is inconsistent, as it can

underestimate a feature's importance even when its true impact increases. In contrast, SHAP utilizes game theory and provides fast and precise attribution values for individual trees. With its unique consistency and local accuracy, SHAP is a superior approach for explaining feature importance in XGBoost models [168].

Table 3-5. Data resources and integration/combination methods.

Name of Model	Methods
GTVp_RD	Integration of radiomics and dosiomics GTVp data before feature selection
GTVp_R_CECTcT1T2	Integration of radiomics GTVp data from CECT, cT1WI, T2WI before feature selection
GTVp_R_CECTcT1	Integration of radiomics GTVp data from CECT and cT1WI before feature selection
GTVp_R_cT1T2	Integration of radiomics GTVp data from cT1WI and T2WI before feature selection
GTVp_R_cT1	Single radiomics data from cT1WI
GTVp_R_CECT	Single radiomics data from CECT
GTVp_R_T2	Single radiomics data from T2WI
GTVp_D	Single dosiomics data from GTVp
GTVn_RD	Integration of radiomics and dosiomics data from GTVn before feature selection
GTVn_R	Single radiomics data from GTVn
GTVn_D	Single dosiomics data from GTVn
PTVn_D	Integration of 60 and 70 Gy dosiomics data before feature selection
PTVn_60Gy_D	Single dosiomics data from PTVn_60Gy

Continue…

Name of Model	Methods
PTVn_70Gy_D	Single dosiomics data from PTVn_70Gy
D	Integration of all dosiomics data before feature selection
C	Single clinical data
C&D	Combine selected clinical and dosiomics data for modeling
C&R	Combine selected clinical and radiomics data for modeling
RD	Integration of radiomics and dosiomics data before feature selection
C&RD	Combine selected clinical and RD data for modeling
C&GTVp_RD	Combine selected clinical and GTVp_RD data for modeling
R&D	Combine selected radiomics and dosiomics data for modeling
C&R&D	Combine selected clinical, radiomics and dosiomics data for modeling

3.3. Results

3.3.1. Patient's Characteristics

A total of 397 continuous patients were collected based on their final diagnosis with pathological validation. Of these patients, with a median age of 54 (range 26–86 years), 242 were enrolled for further analysis following the inclusion and exclusion criteria (details in **Figure 3-4**). All patients were negative for oral mucositis with CTCAE graded 0 before radiation therapy. Univariate analysis results of demographic and clinical characteristics for those patients are listed in **Table 3-6**.

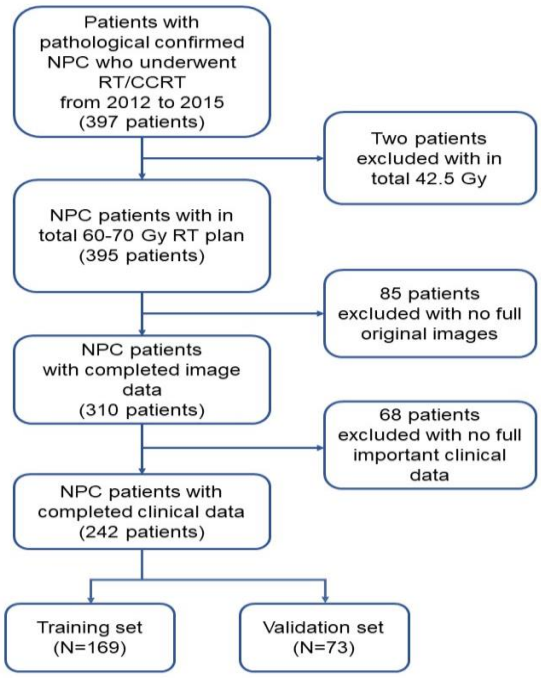

Figure 3-4. Schematic diagram of patient selection.

Table 3-6. Demographic and clinical characteristics for all patients.

Characteristics	**AOM < Grade 3 (Mild AOM)**	**AOM ≥ Grade 3 (Severe AOM)**	***p* Value**
Total Number	191 (78.9%)	51 (21.1%)	
Age, mean ± SD, years	54.89 ± 12.25	50.9 ± 10.60	0.036 *
18–65	149 (61.6%)	44 (18.1%)	
≥65	42 (17.4%)	7 (2.9%)	0.192
Gender Male	135 (55.8%)	41 (16.9%)	
Female	56 (23.1%)	10 (4.1%)	0.167
Treatment			0.004 *
RT alone	27 (11.2%)	0	
CCRT	164 (67.8%)	51 (21.1%)	

Continue…

Characteristics	**AOM < Grade 3 (Mild AOM)**	**AOM ≥ Grade 3 (Severe AOM)**	***p* Value**
T stage			0.031 *
T1	15 (6.2%)	3 (0.1%)	
T2	8 (3.3%)	5 (2.1%)	
T3	137 (56.6%)	28 (11.6%)	
T4	31 (12.8%)	15 (6.2%)	
N stage			0.091
N1	28 (11.2%)	1 (0.4%)	
N2	142 (58.7%)	45 (18.6%)	
N3	20 (8.2%)	5 (2.1%)	
Pathology			
Non-keratinizing squamous cell	175 (72.3%)	48 (19.8%)	0.556
Past health condition			
Past health good	92 (38.0%)	27 (11.2%)	
Basic diseases/cancer	99 (40.9%)	24 (9.9%)	0.545
Allegory of History			
No known drug allergies	176 (72.7%)	46 (19.0%)	
Allergy history	15 (6.2%)	5 (2.1%)	0.653
Vision			
Normal	189 (78.1%)	51 (21.1%)	
With eye impairment	2 (0.8%)	0	0.463
Hearing			
Normal	186 (76.9%)	48 (19.8%)	
With hearing impairment	5 (2.1%)	3 (1.2%)	0.247

Continue…

Characteristics	AOM < Grade 3 (Mild AOM)	AOM ≥ Grade 3 (Severe AOM)	*p* Value
Habits			
Smoking	9 (3.7%)	6 (2.5%)	0.044 *
Non-smoker	182 (75.2%)	45 (18.6%)	
Drinking	4 (1.7%)	1 (0.4%)	
No alcohol consumption	187 (77.3%)	50 (20.7%)	0.953
Height, mean ± SD, cm	163.4 ± 8.5	165.0 ± 8.0	0.561
Body weight,			
mean ± SD, kg			
1st week of RT	63.1 ± 11.9	66.2 ± 14.6	1.599
2nd week of RT	62.0 ± 11.8	64.9 ± 14.5 1.5	
3rd week of RT	61.2 ± 11.4	63.9 ± 14.1 0.116	
BMI			
1st week of RT			
<25	131 (54.1%)	32(13.2%)	
≥25	60(24.8%)	19(7.9%)	0.429
2nd week of RT			
<25	131 (54.1%)	32 (13.2%)	
≥25	60 (24.8%)	19 (7.9%)	0.116
3rd week of RT			
<25	131 (54.1%)	31 (12.8%)	
≥25	55 (22.7%)	20 (8.3%)	0.153
4th week of RT			
<25	142 (58.7%)	34 (14.0%)	
≥25	49 (20.2%)	17 (7.0%)	0.274

* $p < 0.05$. All the above data are derived from biopsy-proven primary NPC patients without the existence of distant metastasis or co-existing tumors of other type at diagnosis.

3.3.2. Feature Extraction and Model Development

Feature Extraction

In this study, a total of 1544 RFs, 386 features each for four modalities of imaging, were extracted from raw and LoG-filtered images. A total of 836 dosiomics features (210 for GTVn, 211 for GTVp, 204 for PTVn_60Gy, and 211 for PTVn_70Gy) were extracted from dose images.

Dimension Reduction

For the clinical data, four variables, including age, RT treatment alone, T stage, and smoking habits, were selected after univariate analysis. The LR model was established with these variables. T stage and smoking habits had statistical significance in the LR model with a *p*-value < 0.05 (Details in **Table 3-7**).

Table 3-7. LR results for single clinical data model.

Variables	*p*-Value	95% Confidence Interval	
		Lower 95% Bound	Upper 95% Bound
Age (18, 65)	0.802	0.345	2.274
T	0.007 *		
T 1	0.591	0.149	2.96
T 2	0.069	0.881	29.854
T 3	0.024 *	0.195	0.891
RT alone	0.998	0	.
Smoker	0.043 *	1.037	10.683

* $P < 0.05$.

Radiomics and dosiomics features extracted from various VOIs were put into Mann–Whitney U tests and RFC step by step. RFC selection results of the threshold and feature numbers are listed in **Table 3-8**.

Table 3-8. Threshold of RFC and feature numbers for further model development.

Modal of data	Threshold	Number of features
GTVp_RD	0.014	5
GTVp_R_CECTcT1T2	0.01	8
GTVp_R_CECTcT1	0.0125	5
GTVp_R_cT1T2	0.125	5
GTVp_R_cT1	0.015	4
GTVp_R_CECT	0.01	19
GTVnp_R_T2	0.03	2
GTVnp_D	0.024	6
GTVn_RD	0.02	7
GTVn_R	0.03	7
GTVn_D	0.06	7
PTVn_D	0.03	3
PTVn_60Gy_D	0.03	12
PTVn_70Gy_D	0.042	1
R	0.012	2
D	0.016	13
RD	0.005	13

Models

Nine categories of single-modal models (C, PTVn_70Gy_D, PTVn_60Gy_D, GTVn_D, GTVp_D, GTVn_R, GTVp_R_T2, GTVp_R_CECT, and GTVp_R_cT1) were established with single modal, single modality, and single VOI data. The best validation AUC was at 0.75 ± 0.12 (training AUC = 0.73 ± 0.01) of a GNB model (GTVp_R_cT1) with radiomics data from GTVp of cT1WI. Seven groups of models with data integrated before feature selection (raw-data integration) were generated with the best AUC of a GNB

model (GTVp_RD) at 0.81 ± 0.01 (training AUC = 0.79 ± 0.01). This best-performing model was constructed with features selected from radiomics and dosiomics data in the region of GTVp. In addition, six sets of combined data after feature selection were also used for modeling. A best LR model (C&R&D) with AUC at 0.79 ± 0.14 (training AUC = 0.81 ± 0.02) was set with the simply combined data of selected clinical, dosiomics, and RFs (Details of mean 10-fold validation AUC results are listed in **Figure 3-5**).

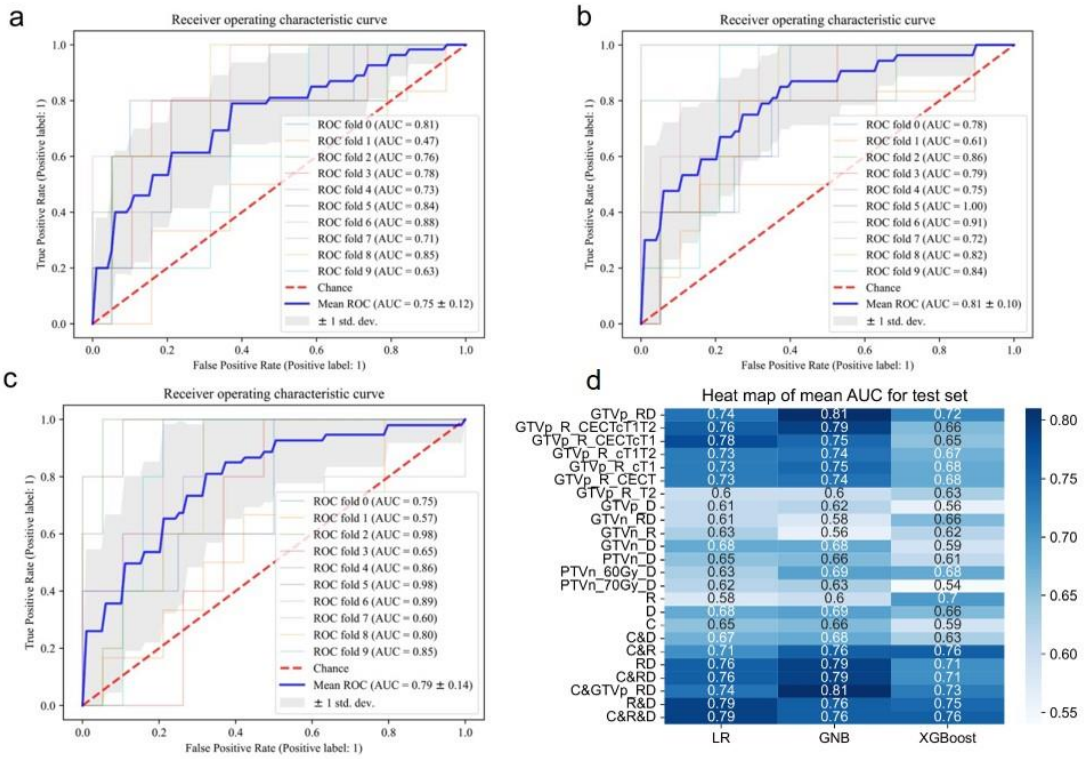

Figure 3-5. 10-fold validation AUC results for the internal validation dataset. (a) The AUC plot of GNB model for the GTVp_R_cT1 data set. (b) The AUC plot of GNB model for the GTVp_RD data set. (c) The AUC plot of LR model for the C&R&D data set. (d) The heatmap of mean AUC results for all models**.**

The SHAP analysis showed the importance of the five features in the GTVp_RD model for prediction of severe AOM. Four of the five features were derived from cT1WI.

All five features are texture features. No dosiomics features were selected after the feature selections (details in **Figure 3-6**).

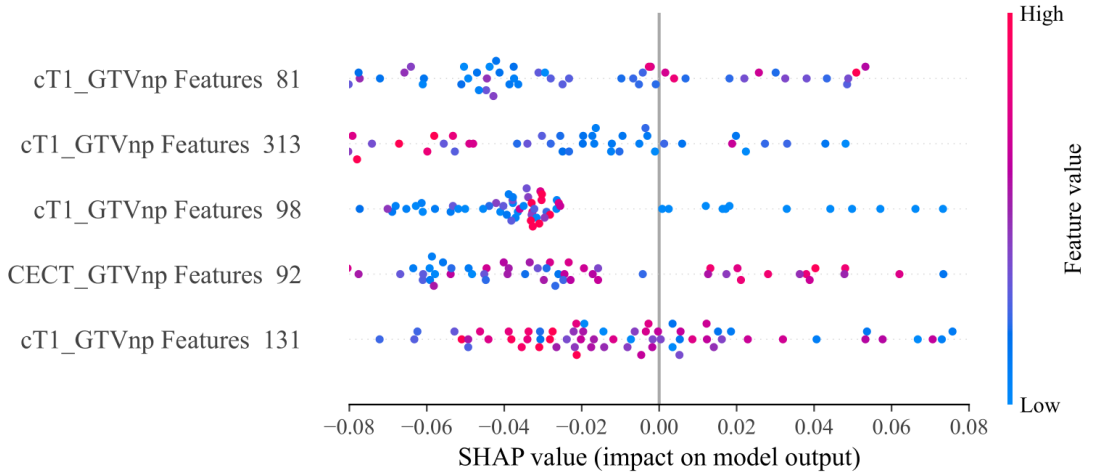

Figure 3-6. Feature importance of SHAP for XGBoost model of GTVp_RD. From the highest to the lowest level, the features are categorized in GLSZM, log sigma 6 0 mm 3D GLCM, original GLDM, GLDM, and log-sigma 1.0 mm 3D GLCM.

3.4. Discussion

In our study, we used simply combined and data-fusion methods to manage multi-modalities of data (clinical, radiomics, and dosiomics), multimodalities of imaging (CECT, cT1WI, and T2WI), and multi-regional information (GTVn, GTVp, and PTVn) to predict the incidence of severe AOM. Multiple models were established to evaluate and determine which method was effective for clinical decision-making. Comparison of the AUC between models showed that the simple combination of single-modal data of selected features had the most stable performance (C&R&D), with an average AUC of 0.77 ± 0.17. In addition, data-fusion methods, integrating radiomics and dosiomics data before selection procedures, resulted in the best-performing model (GTVp_RD), with the best validation

AUC of 0.81 ± 0.01. This is also the best AUC among the existing AOM prediction models from previous studies. The feature numbers in the C&R&D model and GTVp_RD were 29 and five, respectively. Obviously, data fusion was more efficient for training a model with one-sixth the number of features to achieve stronger model predictability.

To better explain the correlations of the selected features and severe AOM for NPC patients, a SHAP plot was applied for the GTVp_RD XGBoost model. In this model, RFs extracted from GTVp in cT1WI images yielded the highest and majority prediction value for severe AOM. Poolakkad and his colleagues established a ML model of 253 H&N patients' clinical data with the best AUC of 0.79 for AOM prediction [169]. Most clinical data selected in their study were late after the CCRT scheme, for example, the anti-neoplastic chemotherapy-induced pancytopenia, co-morbidity score, and agranulocytosis. It is worth noting that the features and variables selected in our study were all from the data collected before implementation of the RT regimen. Clinicians could predict the severe AOM before the commencement of RT planning. Personalized treatment strategies adjustment could be achieved using the developed prediction model. Strictly speaking, the concept of dosiomics is originated from radiomics. The data for dosiomics and radiomics are similar in terms of the feature calculation algorithm.[93] The clinical data are different from the "-omics" data in nature. Therefore, instead of integrating raw clinical data with other data, the simple concatenation of selected clinical data was only applied in this study.

Compared with less increase of AUC for the combination of clinical data with integrated RD data, the clinical data could enhance the prediction capability of single-modal models. The single radiomics models (R) and single clinical models (C) have limited prediction performance with average AUC of three models (LR, GNB, and XGBoost) at

0.63 ± 0.06 and 0.63 ± 0.64, respectively. When combining the clinical data with the selected RFs, the model (C&R) outperformed both R and C models with average AUC at 0.74 ± 0.03. The single dosiomics models yielded poor performance, with most AUCs under 0.7 in the validation data set. In a previous study, dose distribution correlated with the incidence and severity of AOM [170]. Dean et al.[171] developed an random forest model with a testing AUC of 0.71 ± 0.09, using a dose–volume histogram, spatial dose metrics from the oral cavity, and clinical data. In the current study, the best dosiomics model had the mean testing AUC of 0.69 ± 0.14. Different tumor-related VOIs may present different prediction value for severe AOM. The difference in VOI selection between the two studies might shed some light on the discrepancy in the findings.

When analyzing the influence of the dose distribution, the oral cavity contours directly represent the dose distribution in the oral mucosa, which might be more accurate than the GTVn, GTVp, PTVn_60_Gy, or PTVn_70_Gy. The VOI of the oral cavity requires specific contouring. It is worth noting that contouring of the oral cavity is not a common practice in the participating hospital of this study. Extra contouring is labor-intensive work in daily clinical practice. Our study only selected the routine VOI broadly used for RT planning, which could support our model to be applied from bench to bedside for clinical decision-making. Besides, the DVH is prone to over-simplifying the dose distribution [172]. It is recommended to combine or integrate dosiomics data with other modalities of data. When incorporating dosiomics data with other data types, the best mean validation AUC could surge to 0.81 ± 0.01. At present, there exists no effective preventive measures for the occurrence of severe AOM in NPC patients undergoing RT. Nevertheless, it is feasible to mitigate the severity of this affliction:

(1) Use of alternative radiation techniques, such as proton therapy, may be considered to reduce the risk of oral mucositis while maintaining treatment efficacy [150, 173].

(2) Shortening the duration of chemotherapy. For advanced NPC patients who need to accept both radiotherapy and chemotherapy, shortening the exposition time to chemotherapy agents has shown lower mucosal toxicity [174].

(3) Photo-biomodulation is a supportive treatment for the protection of high-risk mucositis patients [175].

(4) Supportive care interventions: preemptive or proactive use of supportive care interventions, such as oral hygiene measures, pain management, or nutritional support, may be considered to prevent or reduce the severity of AOM [176].

The limitations of our study were:

(1) The mucositis grade levels of our patients had an imbalanced distribution. This might have had a negative influence on the data analysis work. The imbalanced results were the nature of the clinical situation. Patients were stratified into the training and validation groups according to the severity of OM, which could offset the imbalance problem [150, 177, 178].

(2) Potential bias of smoking information: in our study, the number of smoking patients might be underestimated due to the nature of this patient-reported outcome. This data was reported by patients at the time of their hospital visit and recorded in the nursing consultation notes.

(3) The severity of AOM was scaled with CTCAE in v3 or v4.03, almost equivalent for evaluation of mucositis. Various criteria are available for mucositis grading, such as the those of the RTOG and the WHO. These scales have excellent concordance with bundled scores of 3 and 4 to describe severe AOM [75]. The CTCAE is easily conducted by clinicians and nurses and broadly applied in the hospital.

(4) The correlations of contributors under AOM for NPC patients are complex. For clinical decision-making, genome information, other clinical information such as fermented-food consumption and EBV infection, and pathological image may also play critical roles. The limited data resources for multimodal data integration are common challenges in the data-mining field. The radiomics data in our study also provided relevant genomic information. Compared with gene test results, the CECT and MRI examination images collected in our research are clinical routines used by clinicians to set the RT plan for NPC patients. These noninvasive examinations could serve as high-throughput screening tools for further application of severe AOM prediction in the future.

(5) Other selection of VOIs: for practical consideration, we have not added the VOIs of the oral cavity, tongue, pharyngeal muscles, etc., which may hold potential predictive value for AOM. Further investigation is recommended to incorporate this information to enhance the accuracy of the analysis.

3.5. Conclusion

AOM is a challenging and distressing complication in NPC patients following RT. Prediction of severe AOM is necessary for timely prevention and intervention, which would further improve the QoL and survival of patients. In this study, we adopted

multimodal data (clinical, radiomics, and dosiomics), multimodality of imaging (CECT, cT1WI, and T2WI), and multi-regional information (GTVn, GTVp, and PTVn) to develop a best-performance model for severe AOM prediction. The simple combination of selected information and data fusion were applied in our work. The results demonstrated that the fusion of radiomics and dosiomics data from the primary tumor could generate the most effective and best-performing model (mean AUC = 0.81 ± 0.01). The data resources and VOIs selected in this study are routinely used in clinical practice, which has excellent potential for further clinical support. Further validation work on a large cohort is warranted to validate model generalizability. Our current integration work only considers two methods for integrating radiomics and dosiomics data. However, we found that the clinical data had little contribution to the model performance, possibly due to the small number of clinical data compared to radiomics and dosiomics data. Other potential multimodal data integration frameworks to balance the weight of clinical variables and radiomics/dosiomics features are worth exploring.

Chapter 4.

Overall Survival

4.1. Introduction

The OS is the most commonly used outcome measure for cancer patients. Prognostic prediction of OS for NPC patients using ML models optimizes resource allocation, improves prognosis assessment, guides clinical decision-making, and provides valuable insights for researchers, fostering further research and development. In this part, we intended to using CT-based radiomics and clinical data to develop a generalizable prognostic DL model to predict the 5-year OS of NPC patients in Hong Kong and Canada. The entire study design is introduced following the data collection, imaging preprocessing, feature extraction, feature redundancy to model development.

4.2. Methods

4.2.1. Patient Data

The study follows the MI-CLAIM (Minimum Information about CLinical Artifcial Intelligence Modelling) checklist (2020) [179, 180] and the TRIPOD+AI (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis + Artificial Intelligence) statement (2024) [181]. In view of the retrospective nature of study, the study informed consent and ethical approval were waived by the ethnic committee of Queen Elizabeth Hospital (QEH) and University Health Network (UHN).

The CT images, GTVp contour mask files and clinical variables were searched and collected from a private QEH database in Hong Kong and a public RADCURE database in Canada (https://www.cancerimagingarchive.net/collection/radcure). Details are presented in **Figure 4-1**. The acquisition parameters of CT scanning are listed in **Table 4-1**. Patient data from RADCURE was randomly split into a train and an independent internal validation cohort, stratified by the survival outcome, with the ratio of eight to two. Patients from QEH was set as an external validation cohort. The inclusion criteria included: 1) Pathological validated NPC patients and no prior surgery at diagnosis; 2) Patients with total IMRT dose around 60-70 Gy; and 3) Patients with GTVp contour mask delineated by registered radiation oncologists, pretreatment CT images, and clinical variables. The exclusion criteria included: 1) Patients without completed CT images or GTVp contour files; 2) Age of patients less than 18 years old. Clinical variables included the following: age, gender, Eastern cooperative oncology group performance status (ECOG PS), TNM stage, Clinical stage, Pathological types (based on World Health Organization classification, WHO), smoking habits, treatment methods (radiotherapy with/without chemotherapy).

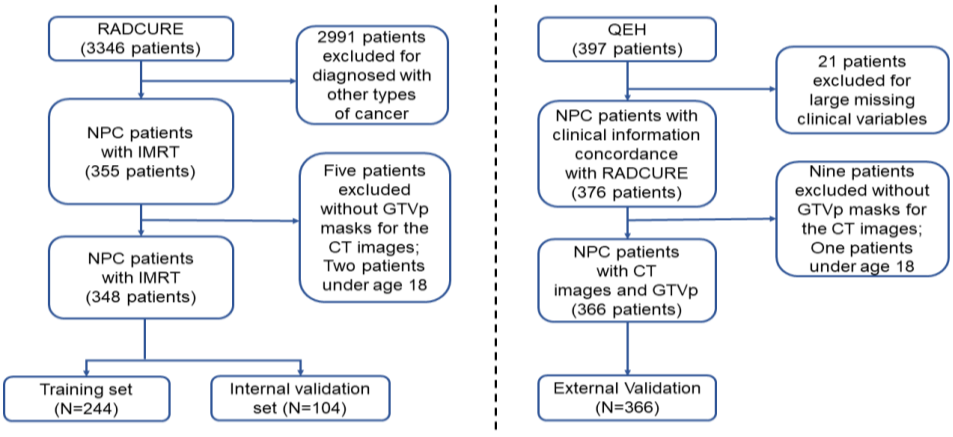

Figure 4-1. Flowchart of patient selection process QEH and RADCURE datasets

Table 4-1. The acquisition parameters of CT scanning.

Parameter	RADCURE	QEH
Position	Supine	Supine
Slice thickness [mm]	2 [2-2.5]	3 [3-3]
Voltage [kVp]	120 [120,120]	120 [120,120]
Exposure [mAs]	300 [121-540]	300 [250-350]
Pixel spacing [mm]	0.976 [0.702-1.17]	1.102 [0.941-1.330]

4.2.2. Imaging Preprocessing

In this study, CT images and GTVp masks were converted to digital imaging and communications in medicine (DICOM) and radiation therapy structure (RTSTRUCT) formats, respectively. All DICOM and RTSTRUCT files were transferred to MetaImage (MHA) files for further processing. All imaging preprocessing methods were in accordance with Image Biomarker Standardization Initiative (IBSI) guidelines [16, 17]. In particular, all images were resampled to 1x1x1 mm^3 and normalized to [-150,180] HU. The number of intervals into which the pixel values of an image are fixed at 32 bin count.

4.2.3. Feature Extraction

All feature extraction procedures were conducted in our in-house platform which is developed based on SimpleITK (v1.2.4) and PyRadiomics (v2.2.0). The GTVp masks was delineated by experienced clinicians as RT planning masks. The minimum dimensions required for the images is at least three dimensions (VOIs). Three categories of RFs including Shape, first-order, GLCM, GLRLM, GLSZM, GLDM, and NGTDM were extracted from the CT images. Perturbations of images were utilized to delineate the

influence of variations from human contours based on our previous studies [182]. Specifically, a total of 40 perturbations were performed, resulting in 40 times of calculations for RFs under different perturbation conditions. The rotation angles used were -20°, 0°, and 20°, indicating that the images were rotated by these angles for feature calculation. Additionally, three translation distances (0, 0.4, and 0.8) and a smoothing sigma of 10 were employed.

4.2.4. Model Development and Evaluation

The Jupyter (v6.5.4), python (v3.11.7) were utilized for statistical analysis, model training, and evaluation. The intraclass correlation coefficient (ICC) [183] and coefficient of variant (CV) [184] were calculated based on the 40 times of perturbations. The cutoff value for ICC>0.9 and CV<15% were set for the selection of robust RFs [185]. For clinical variables, nearest neighbour imputation was used for the missing values. The categorical variables were compared with the Chi-square test. The continuous and ordinal variables were compared with the Mann-Whitney U test.

For ML model with RFs, Random Forest Classifier (RFC) was utilized for the important feature selection. Four conventional ML models, namely Logistic Regression (LR), Random Forest, Gaussian Naive Bayes (GNB), and eXtreme Gradient Boosting (XGBoost), were developed using selected robust RFs and clinical variables. Additionally, a Multilayer Perceptron (MLP) neural network model was constructed. The robust RFs were directly input into the MLP model to obtain the MLP score. Subsequently, the score was concatenated with the clinical variables to generate a nomogram based on LR model.

Receiver Operating Characteristic (ROC), AUC, Accuracy (ACC) were applied for evaluation of models. Details are presented in **Figure 4-2**.

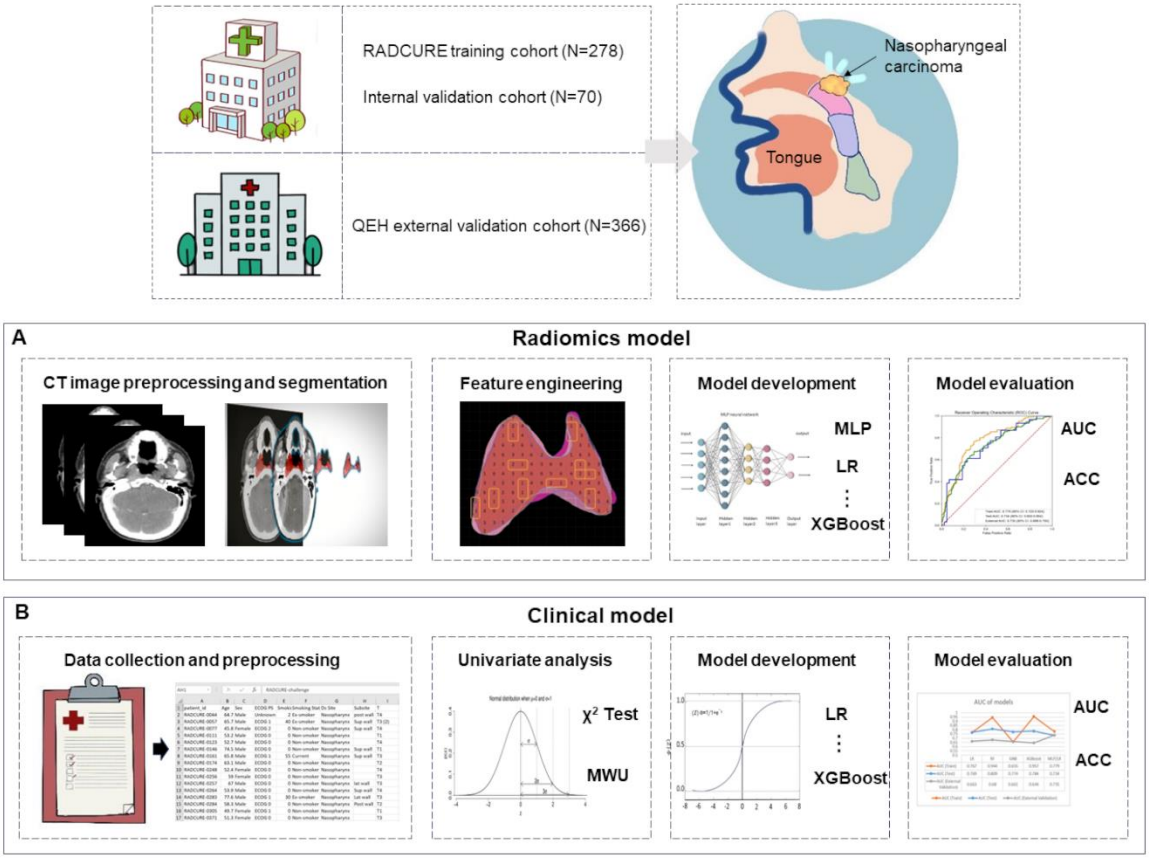

Figure 4-2. Study layout. 348 and 366 NPC patients were retrospective searched from the RADCURE and QEH datasets. All patients underwent pretreatment CT imaging. **A)** Radiomics workflow: Imaging preprocessing and segmentation; Feature engineering; model development and model evaluation; **B)** Clinical variable workflow: Data collection and preprocessing; univariate analysis; model development; and model evaluation. **C)** The development and evaluation of the nomogram. MLP=multilayer perceptron; LR=logistic regression; XGBoost= eXtreme Gradient Boosting; AUC = area under the curve; ACC= accuracy; MWU=Mann Whitney U test; Decision Curve Analysis (DCA); Kaplan-Meier (K-M) analysis;

4.3. Results

4.3.1. Patient Data

A total of 397 and 3346 NPC patients were collected from QEH (Hong Kong, China) and RADCURE (Toronto, Canada) databases, respectively. The date of diagnosis for NPC patients in QEH and RADCURE were from 2012 to 2015 and from 2005 to 2017, respectively. 348 NPC patients and 366 NPC patients from RADUCRE and QEH with CT images, GTVp mask, and clinical variables were finally recruited into this study (Details are provided in **Figure 4-1**). The datasets for training, internal validation, and external validation consisted of 244, 104, and 366 patients, respectively. The mean follow-up time for QEH and RADCURE are 4.5 and 5.3 years, respectively. In RADCURE datasets, ten patients without smoking status and three patients without ECOG PS information. In QEH datasets, pathological types were unknown for four patients; 16 patients without smoking status. All these information was imputed accordingly. Adjusted characteristics of patients in training, internal and external cohorts are listed in the **Table 4-2**. The continuous variables are reported as median and standard deviation. The categorial and ordinal variables are summarized as frequencies and percentage.

Table 4-2. Characteristics of patients in training, internal and external cohorts

Characteristics	Training	Internal Validation	External Validation
No. of patients	278	70	366
Age	52.9±12.0	54.0±13.8	54.2±12.6
Male	194 (69.8)	55 (78.6)	273 (74.6)

Continue…

Characteristics	Training	Internal Validation	External Validation
Female	84 (30.2)	15 (21.4)	93 (25.4)
ECOG PS			
0	198 (71.2)	51 (72.9)	58 (15.8)
1	76 (27.3)	17 (24.3)	292 (79.8)
2	3 (1.1)	1 (1.4)	13 (3.6)
3	1 (0.4)	1 (1.4)	3 (0.8)
Smoking Status			
Non-smoker	155 (55.8)	42 (60)	288 (78.7)
Ex-smoker	77 (27.7)	15 (21.4)	54 (14.8)
Current	46 (16.5)	13 (18.6)	24 (6.6)
T			
1	74 (26.6)	25 (35.7)	24 (6.6)
2	46 (16.5)	4 (5.7)	21 (5.7)
3	86 (30.9)	27 (38.6)	246 (67.2)
4	82 (29.5)	14 (20.0)	75 (20.5)
N			
0	31 (11.2)	11 (15.7)	3 (0.8)
1	71 (25.5)	16 (22.9)	44 (12)
2	141 (50.7)	32 (45.7)	275 (75.1)
3	86 (30.9)	11 (15.7)	44 (12)
M			
0	276 (99.3)	70 (100)	365 (99.7)
1	2 (0.7)	0 (0)	1 (0.3)

Continue…

Characteristics	Training	Internal Validation	External Validation
Clinical Stage			
1	9 (3.2)	4 (5.7)	2 (0.5)
2	32 (11.5)	10 (14.3)	9 (2.5)
3	127 (45.7)	34 (48.6)	239 (65.3)
4	110 (39.6)	22 (31.4)	116 (31.7)
Pathological Type			
WHO I	8 (2.9)	3 (4.3)	20 (5.5)
WHO II	268 (96.4)	67 (95.7)	345 (94.3)
WHO III	2 (0.7)	0 (0)	1 (0.3)
Treatment			
IMRT	32 (11.5)	10 (14.3)	52 (14.2)
IMRT+chemo	246 (88.5)	60 (95.7)	314 (85.8)

Note: ECOG PS= Eastern Cooperative Oncology Group Performance Status; WHO= World Health Organization; IMRT= Intensity-Modulated Radiation Therapy.

4.3.2. Feature Extraction and Model Development

Feature Extraction

1130 RFs, including 107 original RFs, 278 log-sigma RFs, and 743 wavelet RFs were extracted from GTVp. Among these RFs, 170 RFs were robust RFs with ICC >0.9 and CV<15%.

Models

In the training dataset, three clinical variables had significant *p* value <0.05 after univariate analysis: age, sex and smoking status (Detailed results are displayed in **Table 4-3**). The best ML model with clinical variables was XGBoost, which achieved internal and

external validation AUC around 0.55. (Results can be found in **Table 4-4**). For ML models with RFs, RFC was applied for further feature selection. Important feature number of RFC were set at 6, 8, 10, 12, 14. The ML models with best internal AUC were listed in **Table 4-5.** Other details are listed in **Table 4-6, 10)**. The RF model had best internal AUC of 0.840 (95% CI: 0.728, 0.938). The LR model with MLP score (the nomogram) achieved the best internal validation and external AUC of 0.796 (95% CI: 0.689-0.881) and 0.700 (95% CI: 0.642-0.753), respectively.

Table 4-3. The results of univariate analysis for clinical variables in training cohort.

Characteristics	**Univariate (*p* value)**
Age	0.029*
Sex	0.038*
ECOG PS	0.187
Smoking Status	0.168
T	0.726
N	0.980
M	1.000
Clinical Stage	0.771
Pathological Type	0.383
Treatment	0.116

**p*<0.05

Table 4-4. AUC and ACC of ML models for clinical variables

Models	Training		Internal		External	
	AUC (95%CI)	ACC	AUC (95%CI)	ACC	AUC (95%CI)	ACC
LR	0.647(0.581, 0.712)	0.594	0.491(0.357, 0.626)	0.529	0.566 (0.507, 0.624)	0.574
RF	0.799(0.744, 0.846)	0.723	0.462(0.328, 0.609)	0.486	0.576 (0.519, 0.631)	0.612
GNB	0.656(0.591, 0.717)	0.59	0.506(0.372, 0.642)	0.514	0.551 (0.489, 0.609)	0.505
XGBoost	0.751(0.691, 0.802)	0.662	0.554(0.417, 0.692)	0.543	0.547 (0.492, 0.601)	0.582

Table 4-5. AUC and ACC of ML models with robust RFs as input

Models	No.	Training		Internal		External	
		AUC (95%CI)	ACC	AUC (95%CI)	ACC	AUC (95%CI)	ACC
LR	8	0.831 (0.781, 0.876)	0.788	0.788 (0.668, 0.9)	0.786	0.585 (0.524, 0.642)	0.566
RF	12	0.985(0.973, 0.994)	0.928	0.84(0.728, 0.938)	0.757	0.560 (0.502, 0.619)	0.552
GNB	14	0.849 (0.804, 0.891)	0.784	0.825 (0.712, 0.921)	0.771	0.575(0.515, 0.633)	0.574
XGBoost	6	0.998 (0.996, 1.0)	0.982	0.812(0.696, 0.912)	0.714	0.453(0.395, 0.513)	0.552
Nomogram	5	0.827 (0.776, 0.874)	0.774	0.786 (0.688, 0.871)	0.762	0.705 (0.646, 0.749)	0.623

Note: No.=numbers of variables as input; AUC=Area Under the Curve; ACC=Accuracy; ML=Machine Learning; LR=Logistic Regression; RF=Random Forest; GNB= Gaussian Naive Bayes; XGBoost= Extreme Gradient Boosting.

Table 4-6. The AUC and ACC of ML models with six robust RFs

Model	Train			Internal			External		
	AUC	95% CI	ACC	AUC	95% CI	ACC	AUC	95% CI	ACC
LR	0.809	(0.758, 0.856)	0.745	0.745	(0.619, 0.867)	0.743	0.588	(0.527, 0.646)	0.571
RF	0.985	(0.973, 0.993)	0.928	0.828	(0.715, 0.925)	0.743	0.503	(0.444, 0.562)	0.546
GNB	0.835	(0.787, 0.881)	0.766	0.797	(0.678, 0.897)	0.786	0.558	(0.497, 0.616)	0.566
XGBoost	0.998	(0.996, 1.0)	0.982	0.812	(0.696, 0.912)	0.714	0.453	(0.395, 0.513)	0.552

Table 4-7. The AUC and ACC of ML models with eight robust RFs

Model	Train			Internal			External		
	AUC	95% CI	ACC	AUC	95% CI	ACC	AUC	95% CI	ACC
LR	0.831	(0.781, 0.876)	0.788	0.788	(0.668, 0.9)	0.786	0.585	(0.524, 0.642)	0.566
RF	0.982	(0.967, 0.991)	0.921	0.806	(0.691, 0.906)	0.729	0.438	(0.379, 0.492)	0.549
GNB	0.846	(0.801, 0.888)	0.763	0.783	(0.669, 0.884)	0.686	0.488	(0.426, 0.544)	0.527
XGBoost	0.979	(0.963, 0.99)	0.917	0.805	(0.687, 0.911)	0.714	0.441	(0.38, 0.498)	0.552

Table 4-8. The AUC and ACC of ML models with ten robust RFs

Model	Train			Internal			External		
	AUC	95% CI	ACC	AUC	95% CI	ACC	AUC	95% CI	ACC
LR	0.831	(0.781, 0.876)	0.788	0.788	(0.668, 0.9)	0.771	0.586	(0.524, 0.642)	0.563
RF	0.97	(0.952, 0.985)	0.903	0.833	(0.721, 0.93)	0.786	0.524	(0.466, 0.582)	0.56
GNB	0.844	(0.8, 0.888)	0.781	0.819	(0.71, 0.913)	0.757	0.522	(0.466, 0.58)	0.536
XGBoost	0.966	(0.945, 0.983)	0.899	0.774	(0.648, 0.881)	0.714	0.506	(0.445, 0.562)	0.552

Table 4-9. The AUC and ACC of ML models with 12 robust RFs

Model	Train			Internal			External		
	AUC	95% CI	ACC	AUC	95% CI	ACC	AUC	95% CI	ACC
LR	0.81	(0.756, 0.864)	0.773	0.776	(0.651, 0.893)	0.800	0.572	(0.513, 0.627)	0.546
RF	0.985	(0.973, 0.994)	0.928	0.840	(0.728, 0.938)	0.757	0.560	(0.502, 0.619)	0.552
GNB	0.854	(0.81, 0.895)	0.777	0.804	(0.682, 0.904)	0.786	0.508	(0.449, 0.566)	0.525
XGBoost	0.992	(0.98, 0.999)	0.964	0.786	(0.666, 0.892)	0.686	0.503	(0.445, 0.559)	0.560

Table 4-10. The AUC and ACC of ML models with 14 robust RFs

Model	Train			Internal			External		
	AUC	95% CI	ACC	AUC	95% CI	ACC	AUC	95% CI	ACC
LR	0.81	(0.757, 0.863)	0.766	0.781	(0.656, 0.896)	0.800	0.576	(0.516, 0.631)	0.552
RF	0.982	(0.968, 0.992)	0.928	0.838	(0.721, 0.94)	0.786	0.553	(0.495, 0.608)	0.546
GNB	0.849	(0.804, 0.891)	0.784	0.825	(0.712, 0.921)	0.771	0.575	(0.515, 0.633)	0.574
XGBoost	1	(0.998, 1.0)	0.996	0.797	(0.675, 0.9)	0.757	0.457	(0.4, 0.515)	0.536

The MLP model was constructed with five hidden layers. The model architecture consists of multiple dense layers with different activation functions. Dropout regularization was added after the hidden layer to prevent the overfitting of the model. The input layer, with an input dimension of 'featureNum', was followed by a dense layer with 512 units and a ‘sigmoid’ activation function. A dropout layer with a rate of 0.5 was applied after the first dense layer to prevent overfitting. The subsequent hidden layer was consisted with a dense layer (256 units, activation= ‘Leaky_relu’), and following a dropout layer with a rate of 0.5. Two additional dense layers followed, each with 256 units and ‘Leaky_relu’ activation. The dropout layer with a rate of 0.3 was added for the second one to further regularize the model. Another two hidden layers with 128 units were included. Leaky_relu and sigmoid were added as activation, respectively. Finally, the output layer consisted of a single unit with a ‘sigmoid’ activation function, facilitating binary classification. The RMSprop optimizer was employed to optimize the model's performance, with specific parameters for learning rate, rho, momentum, and epsilon. For the loss function, we utilized the Binary Focal Crossentropy, which was customized with alpha and gamma values of 0.25 and 2.0, respectively, to address class imbalance. The model was evaluated using accuracy and AUC metrics. During the training phase, the model was trained on the provided training dataset for 200 epochs, with a batch size of 16 and a validation split of 0.2. The training history was recorded for further analysis and evaluation. MLP score was concatenated with significant clinical variables to construct the nomogram with logistic regression model.

The nomogram (MLP/LR) achieved the training, internal and external validation AUC of 0.818 (95% CI: 0.765-0.865), 0.734 (95% CI: 0.609-0.839), and 0.735 (95% CI: 0.681-0.783), respectively. The ACC of the model are 0.770, 0.686 and 0.697 for train, internal and external validation. Nomogram was established based on the coefficient of MLP/LR model with

simpleNomo package [186]. Details were displaced in **Figure 4-3**. The *p*-value of log-rank test for training, internal validation and external validation cohorts were all less than 0.05.

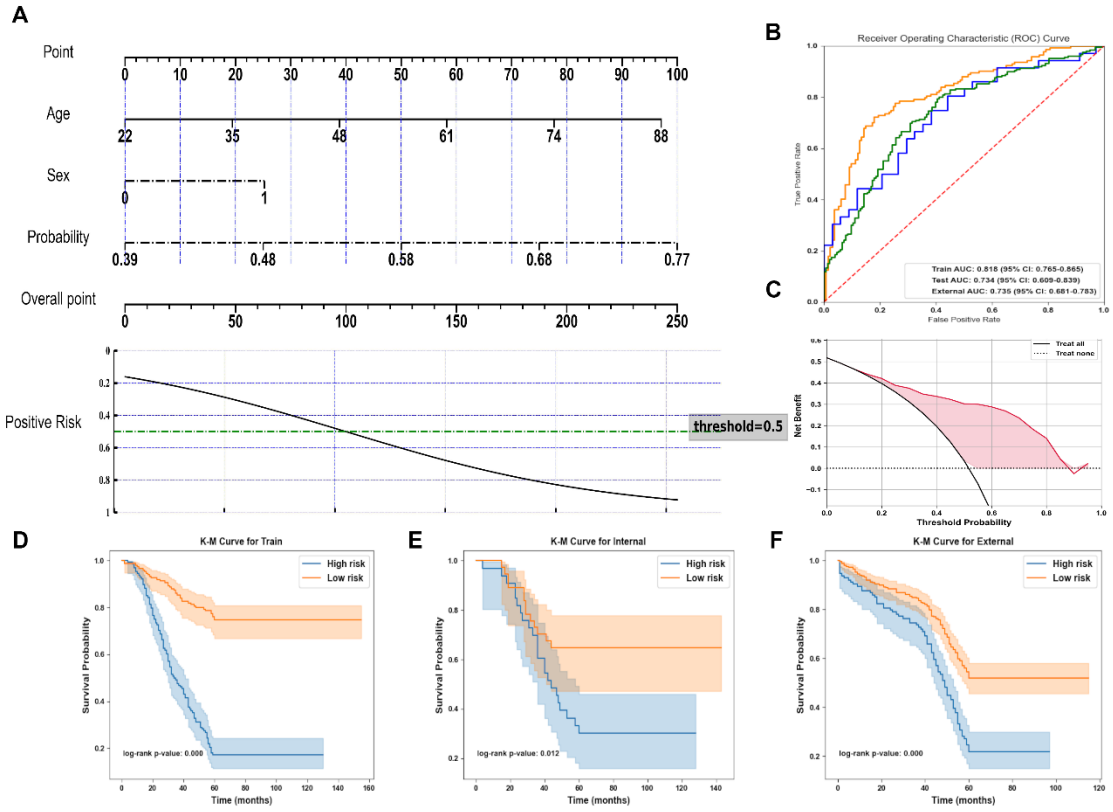

Figure 4-3. Development and evaluation of Nomogram. A. Clinical variables and MLP score were included for the construction of nomogram; B. ROC of nomogram for training, internal and external validation cohorts. C. The decision curve for the training validation cohort. The dot and solid black lines represent non-responders and responders, respectively. The calculation algorithm of net benefit: net_benefit = (tp / n) - (fp / n) * (thresh / (1 - thresh)), in which tp=number of true positive samples, fp=number of false positive samples, thresh= threshold, total= number of true and false positive samples; D. Kaplan-Meier (K-M) curve for the training cohort; E. K-M curve for the internal validation cohort; F. K-M curve for the external validation cohort. Sex: 0=female, 1=male; Smoking status: 0=non-smoker, 1=ex-smoker, 2=current smoker.

4.4. Discussion

The nomogram with CT-based RFs for the prediction of 5-year OS for NPC patients was developed and evaluated on an international dataset in this study. The results of our study demonstrated that the nomogram exhibited the best and most generalizable performance for predicting the 5-year OS of NPC patients in both internal （AUC: 0.734, 95% CI: 0.609-0.839), and external (AUC: 0.735, 95% CI: 0.681-0.783) validation cohorts.

In contrast, the clinical variables showed limited predictive power for the 5-year OS, with best internal and external AUC values of only around 0.55. In a previous study by Mengyun Qiang [187], the information of TNM stage along with other clinical variables, including blood test results, were utilized to predict the 5-year OS of NPC patients. The AUC values for the training, internal validation, and three external validation cohorts were reported as follows: 0.642 (95% CI: 0.602-0.682), 0.621 (95% CI: 0.563-0.679), 0.618 (95% CI: 0.517-0.718), 0.658 (95% CI: 0.552-0.764), and 0.672 (95% CI: 0.584-0.761), respectively. In our study, only three clinical variables, the age, sex and smoking status, exhibited significant differences. Blood test results were not included due to the lack of information in the two datasets. These might contribute to the lower model performance. In addition, other clinical information also has an impact on patients' survival status, such as detailed treatment drugs and dosages, which were not included in the study. However, the results of univariate analysis are in concordance with previous studies. Xiao's study indicated that younger male patients (age<45) have lower survival rate than female. Male patients had poor 5-year OS [132].

In line with previous studies, which have extensively examined the 5-year survival of NPC patients, our ML models' internal validation AUC values are comparable to or even

better than those reported, especially for CT-based radiomics. Best results of studies in recent ten years predicting the OS for NPC patients with internal or external validation cohorts were listed in **Table 4-11.**

Table 4-11. **Summary of previous studies predicting the OS of patients with NPC**

Year	Data	Author	Patients	Best Model	Validation	AUC/C	95% CI
2024 [141]	Clinical	Dan Hu	420	RF	Internal	0.753	0.609-0.896
2024 [188]	R/MRI	Da Fenglin	921	Cox	External	0.731	-
2024 [189]	R/CT	Yinbing Lin	99	Cox	Internal	0.82	-
2023 [146]	Clinical	Rasheed	1094/60	XGBoost	External	0.76	-
2022 [190]	Clinical	Rong Zhao	1304	Cox	Internal	0.717	-
2021 [145]	Clinical	Changchun Lai	519	LASSO	Internal	0.697	0.612-0.734
2020 [142]	R/MRI	Chunyan Cui	792	AutoML	Internal	0.796	-
2020 [191]	R/MRI	Marco Bologna	136	COX	Internal	0.68	-
2020 [192]	R/CT	Linyan Chen	136	LASSO	Internal	0.752	0.614-0.891
2020 [193]	Clinical	Melek Akcay	72	GNB	Internal	0.91	-
2020 [187]	DL/MRI	Mengyun Qiang	2625	3D-CNNs	External	0.757	0.695-0.819
2019 [194]	R/MRI	Xue Ming	303	Cox	Internal	0.845	0.752-0.939
2016 [143]	Clinical	Rou Jiang	347	SVM	Internal	0.633	-

Note: C=C index; RF=Random Forest; LASSO= Least Absolute Shrinkage and Selection Operator; R=radiomics; DL=deep learning with images as input; AutoML= a series combination of ML; SVM= Support vector machine.

Yingbing Lin [189] and Linyan Chen [192] conducted studies that utilized CT-based RFs for 5-year OS prediction. The internal validation AUC (0.82) of Yingbing Lin's work was similar to our's study (RF, 0.84, 95% CI: 0.728, 0.938). However, these studies had a limited number of patients and were restricted to internal validation only. In Melek Akcay's study [193], the GNB model with clinical variables achieved the best internal AUC of 0.91. This study included a total of only 72 NPC patients and did not provide 95% confidence intervals or cross-validation results. Other radiomics studies primarily focused

on extracting RFs from magnetic resonance imaging (MRI) images, especially for local recurrent NPC patients. MRI images have higher resolution compared to CT images, potentially providing more information for OS prediction. However, not all NPC patients had MRI images available. For primary NPC patients, CT is the more commonly used imaging modality for pretreatment RT planning [195, 196]. Two studies have employed DL models and utilized MRI images as input for predicting survival in NPC patients. Lianzhen Zhong's study [148] generated a nomogram to visualize the contribution of clinical variables and DL score for OS prediction. The study only provided the prediction C-index for disease-free survival. Patients in this study were all at stage $T_3N_1M_0$ which has limited clinical application. Another study by Mengyun Qiang [187] also used the DL model and achieved best external validation AUC of 0.853 (95% CI: 0.789- 0.917). Similarly, the study only included NPC patients at stage III or IVA at the first visit. In our study, patients from all clinical stages were included. The DCA and K-M analyses demonstrated that the nomogram exhibited strong capability in stratifying NPC patients into high-risk and low-risk groups. This provides clinicians with a valuable tool for patient screening.

The study does have several limitations that should be acknowledged:

(1) the retrospective nature of the study resulted in the exclusion of patients with incomplete clinical variables, CT images, or GTVp masks. In cases where patients had a few missing clinical variables, data imputation techniques were employed to fill in the gaps. This approach helped to ensure a larger sample size, while it is important to note that it may have affected the representativeness of the study sample. However, it is worth mentioning that clinical variables contribute less to the final models compared to RFs.

Therefore, it may have been more beneficial to retain patients with CT images rather than excluding them based on minor missing clinical variables.

(2) The inherent variance of patients between the QEH and RADCURE datasets is a factor that could potentially influence the predictive performance. The proportion of patients in late stage (clinical stage ⩾3) is obvious higher in QEH (97%) than that in RADCURE (84.29%). The VOIs in the study only included GTVp, while for late status patients, other areas surrounding the primary tumor also has predictive information. This might limit the predictive performance of models.

(3) This study is the first try for the inclusion of two datasets from both southeast China and western country to predict the 5-year OS of NPC patients. The data resource is rare. Further prospective studies with multiple cohorts from different localizations with comprehensive and standardized EMR are needed for the evaluation of generalizability for prognostic models. The development of future radiomics studies would greatly benefit from the availability of large open access resources that provide comprehensive information and various types of images.

Conclusions

In this study, we use CT-based MLP score with late fusion of clinical variables to construct a LR based nomogram predicting the 5-year OS of NPC patients. The AUC of the nomogram achieved 0.734, 95% CI: 0.609-0.839 in internal and 0.735, 95% CI: 0.681-0.783 in external cohorts, demonstrating its generalizability with an international dataset. However, further investigations using more comprehensive patient data are necessary to enhance the model's performance. Nonetheless, our generalizable model exhibited

significant potential for clinical application across various healthcare institutes. Notably, patients with lower probabilities of 5-year OS should be considered for more aggressive treatment options beyond the scope of current clinical guidelines.

Chapter 5.

Summary

This thesis presents the development of ML and DL models for predicting the outcomes of NPC patients using clinical images and EMR. Chapter 1 provides background information on radiomics, precision medicine, and NPC. Chapter 2 reviews previous articles on AOM and OS of NPC patients treated with RT. Chapters 3 and 4 introduce two studies related to the prediction of AOM and OS for NPC patients.

For the prediction of AOM, multimodal data consisting of CT and MRI images, dose files, and EMR were utilized to develop ML models. This study is the first to include multimodal information, including radiomics, dosiomics and EMR, for predicting severe AOM in NPC patients. The model achieved the best performance in predicting severe AOM compared with previous studies, making it a valuable tool for identifying NPC patients following RT at high risk of developing severe AOM. This information can act in suggesting advanced prevention treatment and management, as well as guide clinicians and physicians in RT planning, where AOM is strongly correlated with RT dose. Patients at high risk of severe AOM should be considered for advanced prevention treatment and management, and dose reduction near the oral area.

For the prediction of 5-year OS, international datasets from Southeast Asia (Hong Kong) and North America (Canada) were used for model evaluation. Majority of previous correlated studies restricted their study in internal validation. Previous studies mostly focused on internal validation, with only a few incorporating external validation. However, these studies were often limited to a single region, such as Southeast China, which may not

effectively assess the generalization ability of the models. Radiomics and clinical variables (mainly TNM stages related variables) were used for the model development. The results indicated that MLP/LR model achieved best and consistent AUC performance for the prediction of 5-year OS of NPC patients within two distinct populations. This model exhibits good generalizability, reliability and applicability for screening NPC patients. Moreover, it suggests the potential application of advanced treatment methods beyond the current TNM stage treatment guidelines for NPC patients with lower probabilities of surviving beyond five years.

Several limitations still exist in our studies, necessitating further investigation and development. In the AOM prediction study, we employed 10-fold cross-internal validation to assess model performance. However, the inclusion of external validation datasets is necessary for additional exploration and validation. The current model only incorporates data from the primary tumor and correlated lymph nodes, while the addition of OARs, such as the oral cavity, might enhance its performance. The establishment of a standard definition for delineating the oral mucosa is imperative for accurate contouring. In the OS prediction study, only one contour mask was available for RFs extraction. Developing a generalizable automatic segmentation tool for CT and MRI images of NPC is essential. Furthermore, the available EMR information is not comprehensive enough for accurate OS prediction. Standardizing EMR practices in real-world clinical setting is of utmost importance for future studies. It is worth noting that two of the studies conducted were retrospective, which may introduce patient bias, information bias, selection bias into the investigation. Stratifying patients based on their outcome results was used in both two studies to make sure the distribution balance of patients in train and internal validation

groups. In addition, to address this limitation, prospective studies focusing on radiomics and outcome prediction in NPC are urgently needed to provide more robust and reliable insights.

In conclusion, the integration of clinical variables and RFs shows promise for predicting outcomes in NPC patients following IMRT. The developed models have the potential to serve as adjunct tools for precision medicine, enabling patient screening for severe AOM and 5-year OS. Additionally, the utilization of pathomics and bioinformatics provides valuable insights into protein and gene expression levels, offering micro-level understanding of NPC patients. The combination of macro-level radiomics and micro-level pathomics and bioinformatics data holds potential for providing comprehensive information in the development of precision medicine.

Future work is crucial for further validating models and findings. In the context of the AOM study, it is highly recommended to conduct a comparison and incorporate the VOI of the oral cavity. The initial study results indicated that data derived from the gross tumor area are most effective in predicting AOM. Furthermore, the direct impact of radiation dosage on the oral cavity significantly influences the occurrence of AOM. The implementation of auto-segmentation techniques and the establishment of a standardized definition of the oral cavity for CT images would be advantageous for future investigations. Comparing or combining information extracted from both the primary tumor area and the oral cavity has the potential to enhance the model's performance further. External validation is paramount to evaluate the models' generalizability.

Moreover, additional evidence is required to explore the relationship between the tumor microenvironment and the incidence of severe AOM in NPC patients, which could

offer new insights into AOM treatment. In the case of the OS study, only radiomics features extracted from the primary tumor area and a few clinical variables were considered. To enhance the model's generalizability, it is recommended to include dosiomics, pathomics, genomics, proteomics, and other essential clinical variables, such as EBV infection. Furthermore, integrating information on deep learning patterns has the potential to boost the model's performance.

References

[1] P. Lambin, E. Rios-Velazquez, and R. Leijenaar, "Radiomics: extracting more information from medical images using advanced feature analysis," (in eng), *Eur J Cancer,* vol. 48, no. 4, pp. 441-6, Mar 2012, doi: 10.1016/j.ejca.2011.11.036.

[2] D. J. Hunter and C. Holmes, "Where Medical Statistics Meets Artificial Intelligence," (in eng), *N Engl J Med,* vol. 389, no. 13, pp. 1211-1219, Sep 28 2023, doi: 10.1056/NEJMra2212850.

[3] K. Bera, N. Braman, and A. Gupta, "Predicting cancer outcomes with radiomics and artificial intelligence in radiology," (in eng), *Nat Rev Clin Oncol,* vol. 19, no. 2, pp. 132-146, Feb 2022, doi: 10.1038/s41571-021-00560-7.

[4] J. D. Shur *et al.*, "Radiomics in Oncology: A Practical Guide," (in eng), *Radiographics,* vol. 41, no. 6, pp. 1717-1732, Oct 2021, doi: 10.1148/rg.2021210037.

[5] D. Dong, L. Tang, and Z. Y. Li, "Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer," (in eng), *Ann Oncol,* vol. 30, no. 3, pp. 431-438, Mar 1 2019, doi: 10.1093/annonc/mdz001.

[6] Y. Q. Huang *et al.*, "Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer," (in eng), *J Clin Oncol,* vol. 34, no. 18, pp. 2157-64, Jun 20 2016, doi: 10.1200/jco.2015.65.9128.

[7] S. Das, M. K. Dey, R. Devireddy, and M. R. Gartia, "Biomarkers in Cancer Detection, Diagnosis, and Prognosis," (in eng), *Sensors (Basel),* vol. 24, no. 1, Dec 20 2023, doi: 10.3390/s24010037.

[8] H. Pandha, J. Waxman, and K. Sikora, "Tumour markers," *British journal of hospital medicine,* vol. 51, no. 6, pp. 297-303, 1994.

[9] N. L. Henry and D. F. Hayes, "Cancer biomarkers," (in eng), *Mol Oncol,* vol. 6, no. 2, pp. 140-6, Apr 2012, doi: 10.1016/j.molonc.2012.01.010.

[10] S. S. Arya, S. B. Dias, H. F. Jelinek, L. J. Hadjileontiadis, and A. M. Pappa, "The convergence of traditional and digital biomarkers through AI-assisted biosensing: A new era in translational diagnostics?," (in eng), *Biosens Bioelectron,* vol. 235, p. 115387, Sep 1 2023, doi: 10.1016/j.bios.2023.115387.

[11] D. N. Cagney, J. Sul, R. Y. Huang, K. L. Ligon, P. Y. Wen, and B. M. Alexander, "The FDA NIH Biomarkers, EndpointS, and other Tools (BEST) resource in neuro-oncology," *Neuro-oncology,* vol. 20, no. 9, pp. 1162-1172, 2018, doi: 10.1093/neuonc/nox242.

[12] J. P. O'Connor *et al.*, "Imaging biomarker roadmap for cancer studies," (in eng), *Nat Rev Clin Oncol,* vol. 14, no. 3, pp. 169-186, Mar 2017, doi: 10.1038/nrclinonc.2016.162.

[13] M. Porcu *et al.*, "Radiomics and "radi-…omics" in cancer immunotherapy: a guide for clinicians," (in eng), *Crit Rev Oncol Hematol,* vol. 154, p. 103068, Oct 2020, doi: 10.1016/j.critrevonc.2020.103068.

[14] M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present… any future?," (in eng), *Eur J Nucl Med Mol Imaging,* vol. 44, no. 1, pp. 151-165, Jan 2017, doi: 10.1007/s00259-016-3427-0.

[15] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data," (in eng), *Radiology,* vol. 278, no. 2, pp. 563-77, Feb 2016, doi: 10.1148/radiol.2015151169.

[16] A. Zwanenburg *et al.*, "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping," (in eng), *Radiology,* vol. 295, no. 2, pp. 328-338, May 2020, doi: 10.1148/radiol.2020191145.

[17] P. Whybra *et al.*, "The Image Biomarker Standardization Initiative: Standardized Convolutional Filters for Reproducible Radiomics and Enhanced Clinical Insights," (in eng), *Radiology,* vol. 310, no. 2, p. e231319, Feb 2024, doi: 10.1148/radiol.231319.

[18] H. S. Gabryś, F. Buettner, F. Sterzing, H. Hauswald, and M. Bangert, "Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia," (in eng), *Front Oncol,* vol. 8, p. 35, 2018, doi: 10.3389/fonc.2018.00035.

[19] D. Tan, N. M. Nasir, H. A. Manan, and N. Yahya, "Prediction of toxicity outcomes following radiotherapy using deep learning-based models: a systematic review," *Cancer/Radiothérapie,* 2023, doi: 10.1016/j.canrad.2023.05.001.

[20] M. Santoro *et al.*, "Recent applications of artificial intelligence in radiotherapy: where we are and beyond," *Applied Sciences,* vol. 12, no. 7, p. 3223, 2022, doi: 10.3390/app12073223.

[21] D. Bzdok, Altman, N. & Krzywinski, M., "Statistics versus machine learning," *Nat Methods,* vol. 15, no. 4, pp. 233–234, 2018, doi: 10.1038/nmeth.4642.

[22] A. L. Boulesteix and M. Schmid, "Machine learning versus statistical modeling," *Biometrical Journal,* vol. 56, no. 4, pp. 588-593, 2014, doi: 10.1002/bimj.201300226.

[23] M. A. Carreira-Perpinán, "A review of dimension reduction techniques," *Department of Computer Science. University of Sheffield. Tech. Rep. CS-96-09,* vol. 9, pp. 1-69, 1997.

[24] B. Koçak, E. Durmaz, E. Ateş, and Ö. Kılıçkesmez, "Radiomics with artificial intelligence: a practical guide for beginners," (in eng), *Diagn Interv Radiol,* vol. 25, no. 6, pp. 485-495, Nov 2019, doi: 10.5152/dir.2019.19321.

[25] R. W. Granzier *et al.*, "MRI-based radiomics in breast cancer: feature robustness with respect to inter-observer segmentation variability," *Scientific reports,* vol. 10, no. 1, p. 14163, 2020, doi: 10.1038/s41598-020-70940-z.

[26] B. Kocak, E. Ates, E. S. Durmaz, M. B. Ulusan, and O. Kilickesmez, "Influence of segmentation margin on machine learning-based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas," (in eng), *Eur Radiol,* vol. 29, no. 9, pp. 4765-4775, Sep 2019, doi: 10.1007/s00330-019-6003-8.

[27] B. Kocak *et al.*, "Textural differences between renal cell carcinoma subtypes: Machine learning-based quantitative computed tomography texture analysis with

independent external validation," (in eng), *Eur J Radiol*, vol. 107, pp. 149-157, Oct 2018, doi: 10.1016/j.ejrad.2018.08.014.

[28] T. Hodgdon, M. D. McInnes, N. Schieda, T. A. Flood, L. Lamb, and R. E. Thornhill, "Can Quantitative CT Texture Analysis be Used to Differentiate Fat-poor Renal Angiomyolipoma from Renal Cell Carcinoma on Unenhanced CT Images?," (in eng), *Radiology*, vol. 276, no. 3, pp. 787-96, Sep 2015, doi: 10.1148/radiol.2015142215.

[29] S. J. Ahn, J. H. Kim, S. M. Lee, S. J. Park, and J. K. Han, "CT reconstruction algorithms affect histogram and texture analysis: evidence for liver parenchyma, focal solid liver lesions, and renal cysts," (in eng), *Eur Radiol*, vol. 29, no. 8, pp. 4008-4015, Aug 2019, doi: 10.1007/s00330-018-5829-9.

[30] S. Leng *et al.*, "Subjective and objective heterogeneity scores for differentiating small renal masses using contrast-enhanced CT," (in eng), *Abdom Radiol (NY)*, vol. 42, no. 5, pp. 1485-1492, May 2017, doi: 10.1007/s00261-016-1014-2.

[31] T. K. Koo and M. Y. Li, "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research," (in eng), *J Chiropr Med*, vol. 15, no. 2, pp. 155-63, Jun 2016, doi: 10.1016/j.jcm.2016.02.012.

[32] C. F. Dormann *et al.*, "Collinearity: a review of methods to deal with it and a simulation study evaluating their performance," *Ecography*, vol. 36, no. 1, pp. 27-46, 2013, doi: 10.1111/j.1600-0587.2012.07348.x.

[33] M. A. Hall, "Correlation-based feature selection for machine learning," The University of Waikato, 1999.

[34] N. Gopika and A. M. K. ME, "Correlation based feature selection algorithm for machine learning," in *2018 3rd international conference on communication and electronics systems (ICCES)*, 2018: IEEE, pp. 692-695, doi: 10.1109/CESYS.2018.8723980.

[35] B. Mwangi, T. S. Tian, and J. C. Soares, "A review of feature reduction techniques in neuroimaging," (in eng), *Neuroinformatics*, vol. 12, no. 2, pp. 229-44, Apr 2014, doi: 10.1007/s12021-013-9204-3.

[36] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5-32, 2001, doi: 10.17849/insm-47-01-31-39.1.

[37] A. T. Azar, H. I. Elshazly, A. E. Hassanien, and A. M. Elkorany, "A random forest classifier for lymph diseases," *Computer methods and programs in biomedicine*, vol. 113, no. 2, pp. 465-473, 2014, doi: 10.1016/j.cmpb.2013.11.004.

[38] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature extraction: foundations and applications*. Springer, 2008.

[39] S. L. Andresen, "John McCarthy: father of AI," *IEEE Intelligent Systems*, vol. 17, no. 5, pp. 84-85, 2002, doi: 10.1109/MIS.2002.1039837.

[40] D. A. Hashimoto, G. Rosman, D. Rus, and O. R. Meireles, "Artificial intelligence in surgery: promises and perils," *Annals of surgery*, vol. 268, no. 1, pp. 70-76, 2018, doi: 10.1097/SLA.0000000000002693.

[41] I. El Naqa and M. J. Murphy, *What is machine learning?* Springer, 2015.

[42] I. Castiglioni *et al.*, "AI applications to medical images: From machine learning to deep learning," *Physica medica*, vol. 83, pp. 9-24, 2021, doi: 10.1016/j.ejmp.2021.02.006.

[43] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN computer science,* vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.

[44] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research,* vol. 12, pp. 2825-2830, 2011, doi: 10.5555/1953048.2078195.

[45] L. M. Gladence, M. Karthi, and V. M. Anu, "A statistical comparison of logistic regression and different Bayes classification methods for machine learning," *ARPN Journal of Engineering and Applied Sciences,* vol. 10, no. 14, pp. 5947-5953, 2015.

[46] H. Kamel, D. Abdulah, and J. M. Al-Tuwaijari, "Cancer classification using gaussian naive bayes algorithm," in *2019 international engineering conference (IEC)*, 2019: IEEE, pp. 165-170, doi: 10.1109/IEC47844.2019.8950650.

[47] M. Fratello and R. Tagliaferri, "Decision trees and random forests," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics,* vol. 1, no. S 3, 2018, doi: 10.1016/b978-0-12-809633-8.20337-3.

[48] S. J. Rigatti, "Random forest," *Journal of Insurance Medicine,* vol. 47, no. 1, pp. 31-39, 2017, doi: 10.17849/insm-47-01-31-39.1.

[49] A. Parmar, R. Katariya, and V. Patel, "A review on random forest: An ensemble classifier," in *International conference on intelligent data communication technologies and internet of things (ICICI) 2018*, 2019: Springer, pp. 758-763, doi: 10.1007/978-3-030-03146-6_86.

[50] M. Markatou, H. Tian, S. Biswas, and G. M. Hripcsak, "Analysis of variance of cross-validation estimators of the generalization error," 2005, doi: 10.7916/D86D5R2X.

[51] S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning," *arXiv preprint arXiv:1811.12808*, 2018, doi: 10.48550/arXiv.1811.12808.

[52] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine learning: algorithms and applications*. Crc Press, 2016.

[53] I. Tougui, A. Jilbab, and J. E. Mhamdi, "Impact of the Choice of Cross-Validation Techniques on the Results of Machine Learning-Based Diagnostic Applications," (in eng), *Healthc Inform Res,* vol. 27, no. 3, pp. 189-199, Jul 2021, doi: 10.4258/hir.2021.27.3.189.

[54] P. Lambin *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine," (in eng), *Nat Rev Clin Oncol,* vol. 14, no. 12, pp. 749-762, Dec 2017, doi: 10.1038/nrclinonc.2017.141.

[55] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity. 1943," (in eng), *Bull Math Biol,* vol. 52, no. 1-2, pp. 99-115; discussion 73-97, 1990.

[56] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointestinal endoscopy,* vol. 92, no. 4, pp. 807-812, 2020, doi: 10.1016/j.gie.2020.06.040.

[57] H. Shimizu and K. I. Nakayama, "Artificial intelligence in oncology," *Cancer science,* vol. 111, no. 5, pp. 1452-1460, 2020, doi: 10.1111/cas.14377.

[58] W. H. Delashmit and M. T. Manry, "Recent developments in multilayer perceptron neural networks," in *Proceedings of the seventh annual memphis area engineering and science conference, MAESC*, 2005, vol. 7, p. 33.

[59] E. E. Vokes, D. N. Liebowitz, and R. R. Weichselbaum, "Nasopharyngeal carcinoma," *The Lancet*, vol. 350, no. 9084, pp. 1087-1091, 1997, doi: 10.1016/S0140-6736(97)07269-3.

[60] K. Liu, Y. Wang, L. Ma, S. Yang, and X. Zhang, "Efficacy and safety of a treatment in patients with locoregionally advanced nasopharyngeal carcinoma (LANC) involving carotid artery invasion," (in eng), *Eur Arch Otorhinolaryngol*, vol. 279, no. 12, pp. 5791-5799, Dec 2022, doi: 10.1007/s00405-022-07446-z.

[61] J. F. ME, R. L. Siegel, M. Isabelle Soerjomataram, and D. Ahmedin Jemal, "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," 2024, doi: 10.3322/caac.21834.

[62] Y. Zhang, H. Rumgay, M. Li, S. Cao, and W. Chen, "Nasopharyngeal Cancer Incidence and Mortality in 185 Countries in 2020 and the Projected Burden in 2040: Population-Based Global Epidemiological Profiling," *JMIR public health and surveillance*, vol. 9, no. 1, p. e49968, 2023, doi: 10.2196/49968.

[63] P. Bossi *et al.*, "Nasopharyngeal carcinoma: ESMO-EURACAN Clinical Practice Guidelines for diagnosis, treatment and follow-up(†)," (in eng), *Ann Oncol*, vol. 32, no. 4, pp. 452-465, Apr 2021, doi: 10.1016/j.annonc.2020.12.007.

[64] L. L. Tang *et al.*, "The Chinese Society of Clinical Oncology (CSCO) clinical guidelines for the diagnosis and treatment of nasopharyngeal carcinoma," *Cancer Communications*, vol. 41, no. 11, pp. 1195-1227, 2021, doi: 10.1002/cac2.12218.

[65] Z. Li and Y. Zong, "Review of the histological classification of nasopharyngeal carcinoma," *J Nasopharyng Carcinoma*, vol. 1, no. 15, p. e15, 2014, doi: 10.15383/jnpc.15.

[66] J. C. H. Cheng, K. C. Chao, and D. Low, "Comparison of intensity modulated radiation therapy (IMRT) treatment techniques for nasopharyngeal carcinoma," *International journal of cancer*, vol. 96, no. 2, pp. 126-132, 2001, doi: 10.1002/ijc.1004.

[67] P. M. Teo and A. T. Chan, "Treatment strategy and clinical experience," in *Seminars in cancer biology*, 2002, vol. 12, no. 6: Elsevier, pp. 497-504, doi: 10.1016/S1044579X02000925.

[68] G. Peng *et al.*, "A prospective, randomized study comparing outcomes and toxicities of intensity-modulated radiotherapy vs. conventional two-dimensional radiotherapy for the treatment of nasopharyngeal carcinoma," *Radiotherapy and oncology*, vol. 104, no. 3, pp. 286-293, 2012, doi: 10.1016/j.radonc.2012.08.01.

[69] X. Sun *et al.*, "Long-term outcomes of intensity-modulated radiotherapy for 868 patients with nasopharyngeal carcinoma: an analysis of survival and treatment toxicities," *Radiotherapy and oncology*, vol. 110, no. 3, pp. 398-403, 2014, doi: 10.1016/j.radonc.2013.10.020.

[70] M.-X. Zhang *et al.*, "Intensity-modulated radiotherapy prolongs the survival of patients with nasopharyngeal carcinoma compared with conventional two-dimensional radiotherapy: a 10-year experience with a large cohort and long follow-up," *European journal of cancer*, vol. 51, no. 17, pp. 2587-2595, 2015, doi: 10.1016/j.ejca.2015.08.006.

[71] S. Ozdemir, M. Akin, Y. Coban, C. Yildirim, and O. Uzel, "Acute toxicity in nasopharyngeal carcinoma patients treated with IMRT/VMAT," *Asian Pacific Journal of Cancer Prevention,* vol. 16, no. 5, pp. 1897-1900, 2015, doi: 10.7314/APJCP.2015.16.5.1897.

[72] W. H. Organization, *WHO handbook for reporting results of cancer treatment*. World Health Organization, 1979.

[73] S. T. Sonis *et al.*, "Validation of a new scoring system for the assessment of clinical trial research of oral mucositis induced by radiation or chemotherapy," *Cancer,* vol. 85, no. 10, pp. 2103-2113, 1999, doi: 10.1002/(sici)1097-0142(19990515)85:10<2103::aid-cncr2>3.0.co;2-0.

[74] T. M. Atkinson *et al.*, "The association between clinician-based common terminology criteria for adverse events (CTCAE) and patient-reported outcomes (PRO): a systematic review," *Supportive Care in Cancer,* vol. 24, pp. 3669-3676, 2016, doi: 10.1007/s00520-016-3297-9.

[75] A. Villa, M. Vollemans, A. De Moraes, and S. Sonis, "Concordance of the WHO, RTOG, and CTCAE v4.0 grading scales for the evaluation of oral mucositis associated with chemoradiation therapy for the treatment of oral and oropharyngeal cancers," (in eng), *Support Care Cancer,* vol. 29, no. 10, pp. 6061-6068, Oct 2021, doi: 10.1007/s00520-021-06177-x.

[76] A. Delgado and A. K. Guddati, "Clinical endpoints in oncology - a primer," (in eng), *Am J Cancer Res,* vol. 11, no. 4, pp. 1121-1131, 2021.

[77] K. Wang and J. E. Tepper, "Radiation therapy-associated toxicity: Etiology, management, and prevention," (in eng), *CA Cancer J Clin,* vol. 71, no. 5, pp. 437-454, Sep 2021, doi: 10.3322/caac.21689.

[78] C. M. Anderson *et al.*, "Phase IIb, Randomized, Double-Blind Trial of GC4419 Versus Placebo to Reduce Severe Oral Mucositis Due to Concurrent Radiotherapy and Cisplatin For Head and Neck Cancer," (in eng), *J Clin Oncol,* vol. 37, no. 34, pp. 3256-3265, Dec 1 2019, doi: 10.1200/jco.19.01507.

[79] S. Elad, N. Yarom, Y. Zadik, M. Kuten-Shorrer, and S. T. Sonis, "The broadening scope of oral mucositis and oral ulcerative mucosal toxicities of anticancer therapies," (in eng), *CA Cancer J Clin,* vol. 72, no. 1, pp. 57-77, Jan 2022, doi: 10.3322/caac.21704.

[80] K. Berger *et al.*, "Burden of oral mucositis: a systematic review and implications for future research," *Oncology research and treatment,* vol. 41, no. 6, pp. 399-405, 2018, doi: 10.1159/000487085.

[81] L. S. Elting and Y.-C. Chang, "Costs of oral complications of cancer therapies: estimates and a blueprint for future study," *JNCI Monographs,* vol. 2019, no. 53, p. lgz010, 2019, doi: 10.1093/jncimonographs/lgz010.

[82] D. W. Yang *et al.*, "Genome-wide association study identifies genetic susceptibility loci and pathways of radiation-induced acute oral mucositis," (in eng), *J Transl Med,* vol. 18, no. 1, p. 224, Jun 5 2020, doi: 10.1186/s12967-020-02390-0.

[83] Y. Wang *et al.*, "A two-stage genome-wide association study to identify novel genetic loci associated with acute radiotherapy toxicity in nasopharyngeal carcinoma," (in eng), *Mol Cancer,* vol. 21, no. 1, p. 169, Aug 23 2022, doi: 10.1186/s12943-022-01631-8.

[84] C. Scully, S. Sonis, and P. Diz, "Oral mucositis," *Oral diseases,* vol. 12, no. 3, pp. 229-241, 2006, doi: 10.1111/j.1601-0825.2006.01258.x.
[85] G. B. van den Broek, A. J. Balm, M. W. van den Brekel, M. Hauptmann, J. H. Schornagel, and C. R. Rasch, "Relationship between clinical factors and the incidence of toxicity after intra-arterial chemoradiation for head and neck cancer," (in eng), *Radiother Oncol,* vol. 81, no. 2, pp. 143-50, Nov 2006, doi: 10.1016/j.radonc.2006.09.002.
[86] M. Vera-Llonch, G. Oster, M. Hagiwara, and S. Sonis, "Oral mucositis in patients undergoing radiation treatment for head and neck carcinoma," (in eng), *Cancer,* vol. 106, no. 2, pp. 329-36, Jan 15 2006, doi: 10.1002/cncr.21622.
[87] J. Bowen and C. Cross, "The Role of the Innate Immune Response in Oral Mucositis Pathogenesis," *International Journal of Molecular Sciences,* vol. 24, no. 22, p. 16314, 2023, doi: 10.3390/ijms242216314.
[88] M. Kudrimoti *et al.*, "Dusquetide: reduction in oral mucositis associated with enduring ancillary benefits in tumor resolution and decreased mortality in head and neck cancer patients," *Biotechnology reports,* vol. 15, pp. 24-26, 2017, doi: 10.1016/j.btre.2017.05.002.
[89] K. Li *et al.*, "Impact of dose volume parameters and clinical factors on acute radiation oral mucositis for locally advanced nasopharyngeal carcinoma patients treated with concurrent intensity-modulated radiation therapy and chemoradiotherapy," (in eng), *Oral Oncol,* vol. 72, pp. 32-37, Sep 2017, doi: 10.1016/j.oraloncology.2017.06.026.
[90] N. Saito, Y. Imai, T. Muto, and T. Sairenchi, "Low body mass index as a risk factor of moderate to severe oral mucositis in oral cancer patients with radiotherapy," (in eng), *Support Care Cancer,* vol. 20, no. 12, pp. 3373-7, Dec 2012, doi: 10.1007/s00520-012-1620-7.
[91] D. I. Rosenthal and A. Trotti, "Strategies for managing radiation-induced mucositis in head and neck cancer," in *Seminars in radiation oncology*, 2009, vol. 19, no. 1: Elsevier, pp. 29-34, doi: 10.1016/j.semradonc.2008.09.00.
[92] H. Wardill *et al.*, "Prediction of mucositis risk secondary to cancer therapy: a systematic review of current evidence and call to action," *Supportive Care in Cancer,* vol. 28, no. 11, pp. 5059-5073, 2020, doi: 10.1007/s00520-020-05579-7.
[93] L. Rossi *et al.*, "Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy," (in eng), *Radiother Oncol,* vol. 129, no. 3, pp. 548-553, Dec 2018, doi: 10.1016/j.radonc.2018.07.027.
[94] R. Sun *et al.*, "Imaging approaches and radiomics: toward a new era of ultraprecision radioimmunotherapy?," *Journal for Immunotherapy of Cancer,* vol. 10, no. 7, 2022, doi: 10.1136/jitc-2022-004848.
[95] J. Dai, H. Wang, Y. Xu, X. Chen, and R. Tian, "Clinical application of AI-based PET images in oncological patients," in *Seminars in Cancer Biology*, 2023: Elsevier, doi: 10.1016/j.semcancer.2023.03.005.
[96] A. R. Van Erkel and M. Peter, "Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology," *European Journal of radiology,* vol. 27, no. 2, pp. 88-94, 1998, doi: 10.1016/S0720-048X(97)00157-5.
[97] Z. Jiang, Y. Liang, X. Wang, M. Zhuang, M. Feng, and Y. Kuang, "A Radiomics-Based Light Gradient Boosting Machine to Predict Radiation-Induced Toxicities in

Nasopharynx Cancer Patients Receiving Chemoradiotherapy," *International Journal of Radiation Oncology, Biology, Physics,* vol. 114, no. 3, p. e118, 2022, doi: 10.1016/j.ijrobp.2022.07.933.

[98] P. S. Satheeshkumar, M. El-Dallal, and M. P. Mohan, "Feature selection and predicting chemotherapy-induced ulcerative mucositis using machine learning methods," *International Journal of Medical Informatics,* vol. 154, p. 104563, 2021, doi: 10.1016/j.ijmedinf.2021.104563.

[99] J. Dean *et al.*, "Normal tissue complication probability (NTCP) modelling of severe acute mucositis using a novel oral mucosal surface organ at risk," *Clinical oncology,* vol. 29, no. 4, pp. 263-273, 2017, doi: 10.1016/j.clon.2016.12.001.

[100] E. Orlandi *et al.*, "Multivariable model for predicting acute oral mucositis during combined IMRT and chemotherapy for locally advanced nasopharyngeal cancer patients," *Oral oncology,* vol. 86, pp. 266-272, 2018, doi: 10.1016/j.oraloncology.2018.10.006.

[101] P.-J. Li *et al.*, "Predictive model and precaution for oral mucositis during chemo-radiotherapy in nasopharyngeal carcinoma patients," *Frontiers in Oncology,* vol. 10, p. 596822, 2020, doi: 10.3389/fonc.2020.596822.

[102] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE,* vol. 103, no. 9, pp. 1449-1477, 2015, doi: 10.1109/JPROC.2015.2460697.

[103] K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao, and S. P. Shah, "Harnessing multimodal data integration to advance precision oncology," *Nature Reviews Cancer,* vol. 22, no. 2, pp. 114-126, 2022, doi: 10.1038/s41568-021-00408-3.

[104] J. Lipkova *et al.*, "Artificial intelligence for multimodal data integration in oncology," *Cancer cell,* vol. 40, no. 10, pp. 1095-1110, 2022, doi: 10.1016/j.ccell.2022.09.012.

[105] J. Chu, W. Dong, J. Wang, K. He, and Z. Huang, "Treatment effect prediction with adversarial deep learning using electronic health records," *BMC Medical Informatics and Decision Making,* vol. 20, pp. 1-14, 2020, doi: 10.1186/s12911-020-01151-9.

[106] L. A. Vale-Silva and K. Rohr, "Long-term cancer survival prediction using multimodal deep learning," *Scientific Reports,* vol. 11, no. 1, p. 13505, 2021, doi: 10.1038/s41598-021-92799-4.

[107] K. M. Boehm *et al.*, "Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer," *Nature cancer,* vol. 3, no. 6, pp. 723-733, 2022, doi: 10.1038/s43018-022-00388-9.

[108] N. Naithani, S. Sinha, P. Misra, B. Vasudevan, and R. Sahu, "Precision medicine: Concept and tools," (in eng), *Med J Armed Forces India,* vol. 77, no. 3, pp. 249-257, Jul 2021, doi: 10.1016/j.mjafi.2021.06.021.

[109] M. Farrokhi *et al.*, "Role of precision medicine and personalized medicine in the treatment of diseases," *Kindle,* vol. 3, no. 1, pp. 1-164, 2023, doi: 10.5281/zenodo.8176216.

[110] C. Compton, "Precision medicine core: progress in prognostication—populations to patients," *Annals of surgical oncology,* vol. 25, no. 2, pp. 349-350, 2018, doi: 10.1245/s10434-017-6024-y.

[111] D. Billheimer, E. W. Gerner, C. E. McLaren, and B. LaFleur, "Combined benefit of prediction and treatment: a criterion for evaluating clinical prediction models," *Cancer informatics,* vol. 13, p. CIN. S13780, 2014, doi: 10.4137/CIN.S13780.

[112] A.-M. Tsimberidou *et al.*, "Long-term overall survival and prognostic score predicting survival: the IMPACT study in precision medicine," *Journal of Hematology & oncology,* vol. 12, pp. 1-12, 2019, doi: 10.1186/s13045-019-0835-1.

[113] A. Sarma, C. S. Calfee, and L. B. Ware, "Biomarkers and precision medicine: state of the art," *Critical care clinics,* vol. 36, no. 1, pp. 155-165, 2020, doi: 10.1016/j.ccc.2019.08.012.

[114] Y. C. Chen, U. J. Lee, C. A. Tsai, and J. J. Chen, "Development of predictive signatures for treatment selection in precision medicine with survival outcomes," *Pharmaceutical Statistics,* vol. 17, no. 2, pp. 105-116, 2018, doi: 10.1002/pst.1842.

[115] A. F. Abbasi, M. N. Asim, S. Ahmed, S. Vollmer, and A. Dengel, "Survival Prediction Landscape: An In-Depth Systematic Literature Review on Activities, Methods, Tools, Diseases, and Databases," *medRxiv,* p. 2024.01. 05.24300889, 2024, doi: 10.1101/2024.01.05.24300889.

[116] J. J. Driscoll and O. Rixe, "Overall survival: still the gold standard: why overall survival remains the definitive end point in cancer clinical trials," (in eng), *Cancer J,* vol. 15, no. 5, pp. 401-5, Sep-Oct 2009, doi: 10.1097/PPO.0b013e3181bdc2e0.

[117] P. Wang, Y. Li, and C. K. Reddy, "Machine learning for survival analysis: A survey," *ACM Computing Surveys (CSUR),* vol. 51, no. 6, pp. 1-36, 2019, doi: 10.1145/3214306.

[118] A. C. Alba *et al.*, "Discrimination and calibration of clinical prediction models: users' guides to the medical literature," *Jama,* vol. 318, no. 14, pp. 1377-1384, 2017, doi: 10.1001/jama.2017.12126.

[119] D. Bamber, "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *Journal of mathematical psychology,* vol. 12, no. 4, pp. 387-415, 1975, doi: 10.1016/0022-2496(75)90001-2.

[120] B.-H. Nam and R. B. D'Agostino, "Discrimination index, the area under the ROC curve," *Goodness-of-fit tests and model validity,* pp. 267-279, 2002, doi: 10.1007/978-1-4612-0103-8_20.

[121] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern recognition,* vol. 30, no. 7, pp. 1145-1159, 1997, doi: 10.1016/S0031-3203(96)00142-.

[122] D. J. Hand, "Evaluating diagnostic tests: the area under the ROC curve and the balance of errors," *Statistics in medicine,* vol. 29, no. 14, pp. 1502-1510, 2010, doi: 10.1002/sim.3859.

[123] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. John Wiley & Sons, 2013, p. 177.

[124] M. J. Pencina, *Overall C as a measure of discrimination in survival analysis*. Boston University, 2003.

[125] Y. Li, K. S. Xu, and C. K. Reddy, "Regularized parametric regression for high-dimensional survival analysis," in *Proceedings of the 2016 SIAM International Conference on Data Mining*, 2016: SIAM, pp. 765-773, doi: 10.1137/1.9781611974348.86.

[126] S. Abd ElHafeez, G. D'Arrigo, D. Leonardis, M. Fusaro, G. Tripepi, and S. Roumeliotis, "Methods to analyze time-to-event data: the Cox regression analysis," *Oxidative medicine and cellular longevity,* vol. 2021, pp. 1-6, 2021, doi: 10.1155/2021/1302811.

[127] J. De Neve and T. A. Gerds, "On the interpretation of the hazard ratio in Cox regression," (in eng), *Biom J,* vol. 62, no. 3, pp. 742-750, May 2020, doi: 10.1002/bimj.201800255.

[128] W. N. Dudley, R. Wickham, and N. Coombs, "An introduction to survival statistics: Kaplan-Meier analysis," *Journal of the advanced practitioner in oncology,* vol. 7, no. 1, p. 91, 2016, doi: 10.6004/jadpro.2016.7.1.8.

[129] L. Simms, H. Barraclough, and R. Govindan, "Biostatistics primer: what a clinician ought to know--prognostic and predictive factors," (in eng), *J Thorac Oncol,* vol. 8, no. 6, pp. 808-13, Jun 2013, doi: 10.1097/JTO.0b013e318292bdcd.

[130] K. H. Yu *et al.*, "Survival outcome of patients with nasopharyngeal carcinoma with first local failure: a study by the Hong Kong Nasopharyngeal Carcinoma Study Group," *Head & Neck: Journal for the Sciences and Specialties of the Head and Neck,* vol. 27, no. 5, pp. 397-405, 2005, doi: 10.1002/hed.20161.

[131] T. Gokce, I. Unlu, and C. Akcay, "Evaluation of overall survival of nasopharyngeal carcinoma patients treated in ten years at a single institution," *J BUON,* vol. 15, no. 1, pp. 36-42, 2010.

[132] G. Xiao, Y. Cao, X. Qiu, W. Wang, and Y. Wang, "Influence of gender and age on the survival of patients with nasopharyngeal carcinoma," *BMC cancer,* vol. 13, pp. 1-8, 2013, doi: 10.1186/1471-2407-13-226.

[133] L.-L. Zhang *et al.*, "Development of a nomogram model for treatment of nonmetastatic nasopharyngeal carcinoma," *JAMA Network Open,* vol. 3, no. 12, pp. e2029882-e2029882, 2020, doi: 10.1001/jamanetworkopen.2020.29882.

[134] M.-T. Liu, C.-Y. Hsieh, T.-H. Chang, J.-P. Lin, C.-C. Huang, and A.-Y. Wang, "Prognostic factors affecting the outcome of nasopharyngeal carcinoma," *Japanese journal of clinical oncology,* vol. 33, no. 10, pp. 501-508, 2003, doi: 10.1093/jjco/hyg092.

[135] J.-X. Li, S.-M. Huang, B.-X. Wen, and T.-X. Lu, "Prognostic factors on overall survival of newly diagnosed metastatic nasopharyngeal carcinoma," *Asian Pacific Journal of Cancer Prevention,* vol. 15, no. 7, pp. 3169-3173, 2014, doi: 10.7314/APJCP.2014.15.7.3169.

[136] A. T. Chan *et al.*, "Overall survival after concurrent cisplatin–radiotherapy compared with radiotherapy alone in locoregionally advanced nasopharyngeal carcinoma," *Journal of the National Cancer Institute,* vol. 97, no. 7, pp. 536-539, 2005, doi: 10.1093/jnci/dji084.

[137] Y. Zhang *et al.*, "Final Overall Survival Analysis of Gemcitabine and Cisplatin Induction Chemotherapy in Nasopharyngeal Carcinoma: A Multicenter, Randomized Phase III Trial," (in eng), *J Clin Oncol,* vol. 40, no. 22, pp. 2420-2425, Aug 1 2022, doi: 10.1200/jco.22.00327.

[138] Deepti and S. Ray, "A survey on application of machine learning algorithms in cancer prediction and prognosis," in *Data Management, Analytics and Innovation: Proceedings of ICDMAI 2020, Volume 1*, 2021: Springer, pp. 349-361, doi: 10.1007/978-981-15-5616-6_25.

[139] B. Zhang, H. Shi, and H. Wang, "Machine learning and AI in cancer prognosis, prediction, and treatment selection: a critical approach," *Journal of multidisciplinary healthcare*, pp. 1779-1791, 2023, doi: 10.2147/JMDH.S410301.
[140] P. Dhiman *et al.*, "Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review," *BMC medical research methodology*, vol. 22, no. 1, p. 101, 2022, doi: 10.1186/s12874-022-01577-x.
[141] D. Hu, Y. Wang, G. Ji, and Y. Liu, "Using machine learning algorithms to predict the prognosis of advanced nasopharyngeal carcinoma after intensity-modulated radiotherapy," *Current Problems in Cancer*, vol. 48, p. 101040, 2024, doi: 10.1016/j.currproblcancer.2023.10104.
[142] C. Cui *et al.*, "Machine learning analysis of image data based on detailed MR image reports for nasopharyngeal carcinoma prognosis," *BioMed research international*, vol. 2020, 2020, doi: 10.1155/2020/8068913.
[143] R. Jiang *et al.*, "Development of a ten-signature classifier using a support vector machine integrated approach to subdivide the M1 stage into M1a and M1b stages of nasopharyngeal carcinoma with synchronous metastases to better predict patients' survival," *Oncotarget*, vol. 7, no. 3, p. 3645, 2016, doi: 10.18632/oncotarget.6436.
[144] M. Akcay, D. Etiz, O. Celik, and A. Ozen, "Evaluation of prognosis in nasopharyngeal cancer using machine learning," *Technology in Cancer Research & Treatment*, vol. 19, p. 1533033820909829, 2020, doi: 10.1177/153303382090982.
[145] C. Lai *et al.*, "A novel prognostic model predicts overall survival in patients with nasopharyngeal carcinoma based on clinical features and blood biomarkers," *Cancer Medicine*, vol. 10, no. 11, pp. 3511-3523, 2021, doi: 10.1002/cam4.3839.
[146] R. O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, and A. A. Mäkitie, "Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP," *Scientific Reports*, vol. 13, no. 1, p. 8984, 2023, doi: 10.1038/s41598-023-35795-0.
[147] M. Qiang *et al.*, "A prognostic predictive system based on deep learning for locoregionally advanced nasopharyngeal carcinoma," *JNCI: Journal of the National Cancer Institute*, vol. 113, no. 5, pp. 606-615, 2021, doi: 10.1093/jnci/djaa149.
[148] L. Zhong *et al.*, "A deep learning-based radiomic nomogram for prognosis and treatment decision in advanced nasopharyngeal carcinoma: A multicentre study," *EBioMedicine*, vol. 70, 2021, doi: 10.1016/j.ebiom.2021.103522.
[149] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," *Computational and structural biotechnology journal*, vol. 13, pp. 8-17, 2015, doi: 10.1016/j.csbj.2014.1.
[150] M. J. Dodd *et al.*, "Risk factors for chemotherapy-induced oral mucositis: dental appliances, oral hygiene, previous oral lesions, and history of smoking," (in eng), *Cancer Invest*, vol. 17, no. 4, pp. 278-84, 1999, doi: 10.3109/07357909909040598.

[151] B. A. Varghese, S. Y. Cen, D. H. Hwang, and V. A. Duddalwar, "Texture Analysis of Imaging: What Radiologists Need to Know," (in eng), *AJR Am J Roentgenol,* vol. 212, no. 3, pp. 520-528, Mar 2019, doi: 10.2214/ajr.18.20624.

[152] A. J. Buckler, L. Bresolin, N. R. Dunnick, and D. C. Sullivan, "A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging," (in eng), *Radiology,* vol. 258, no. 3, pp. 906-14, Mar 2011, doi: 10.1148/radiol.10100799.

[153] R. Larue *et al.*, "Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study," (in eng), *Acta Oncol,* vol. 56, no. 11, pp. 1544-1553, Nov 2017, doi: 10.1080/0284186x.2017.1351624.

[154] R. Larue *et al.*, "4DCT imaging to assess radiomics feature stability: An investigation for thoracic cancers," (in eng), *Radiother Oncol,* vol. 125, no. 1, pp. 147-153, Oct 2017, doi: 10.1016/j.radonc.2017.07.023.

[155] P. Hu *et al.*, "Reproducibility with repeat CT in radiomics study for rectal cancer," (in eng), *Oncotarget,* vol. 7, no. 44, pp. 71440-71446, Nov 1 2016, doi: 10.18632/oncotarget.12199.

[156] H. J. Aerts *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," (in eng), *Nat Commun,* vol. 5, p. 4006, Jun 3 2014, doi: 10.1038/ncomms5006.

[157] Y. Balagurunathan *et al.*, "Test-retest reproducibility analysis of lung CT image features," (in eng), *J Digit Imaging,* vol. 27, no. 6, pp. 805-23, Dec 2014, doi: 10.1007/s10278-014-9716-x.

[158] Y. Balagurunathan *et al.*, "Reproducibility and Prognosis of Quantitative Features Extracted from CT Images," (in eng), *Transl Oncol,* vol. 7, no. 1, pp. 72-87, Feb 2014, doi: 10.1593/tlo.13844.

[159] D. V. Fried *et al.*, "Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer," (in eng), *Int J Radiat Oncol Biol Phys,* vol. 90, no. 4, pp. 834-42, Nov 15 2014, doi: 10.1016/j.ijrobp.2014.07.020.

[160] L. A. Hunter *et al.*, "High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images," (in eng), *Med Phys,* vol. 40, no. 12, p. 121916, Dec 2013, doi: 10.1118/1.4829514.

[161] S. K. Lam *et al.*, "Multi-Organ Omics-Based Prediction for Adaptive Radiation Therapy Eligibility in Nasopharyngeal Carcinoma Patients Undergoing Concurrent Chemoradiotherapy," (in eng), *Front Oncol,* vol. 11, p. 792024, 2021, doi: 10.3389/fonc.2021.792024.

[162] J. J. M. van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," (in eng), *Cancer Res,* vol. 77, no. 21, pp. e104-e107, Nov 1 2017, doi: 10.1158/0008-5472.Can-17-0339.

[163] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The Design of SimpleITK," (in eng), *Front Neuroinform,* vol. 7, p. 45, 2013, doi: 10.3389/fninf.2013.00045.

[164] V. Grégoire *et al.*, "CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC,RTOG consensus guidelines," (in eng), *Radiother Oncol,* vol. 69, no. 3, pp. 227-36, Dec 2003, doi: 10.1016/j.radonc.2003.09.011.

[165] F. Buettner *et al.*, "Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial," (in eng), *Radiother Oncol,* vol. 103, no. 1, pp. 82-7, Apr 2012, doi: 10.1016/j.radonc.2012.02.006.

[166] J. Han, J. Pei, and H. Tong, *Data mining: concepts and techniques*. Morgan kaufmann, 2022.

[167] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems,* vol. 30, 2017, doi: 10.48550/arXiv.1705.07874.

[168] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888,* 2018, doi: 10.48550/arXiv.1802.03888.

[169] P. S. Satheeshkumar, M. El-Dallal, and M. P. Mohan, "Feature selection and predicting chemotherapy-induced ulcerative mucositis using machine learning methods," (in eng), *Int J Med Inform,* vol. 154, p. 104563, Oct 2021, doi: 10.1016/j.ijmedinf.2021.104563.

[170] G. Sanguineti *et al.*, "Effect of radiotherapy and chemotherapy on the risk of mucositis during intensity-modulated radiation therapy for oropharyngeal cancer," (in eng), *Int J Radiat Oncol Biol Phys,* vol. 83, no. 1, pp. 235-42, May 1 2012, doi: 10.1016/j.ijrobp.2011.06.2000.

[171] J. A. Dean *et al.*, "Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy," (in eng), *Radiother Oncol,* vol. 120, no. 1, pp. 21-7, Jul 2016, doi: 10.1016/j.radonc.2016.05.015.

[172] P. Giraud *et al.*, "Radiomics and Machine Learning for Radiotherapy in Head and Neck Cancers," (in eng), *Front Oncol,* vol. 9, p. 174, 2019, doi: 10.3389/fonc.2019.00174.

[173] A. C. Moreno *et al.*, "Intensity modulated proton therapy (IMPT) - The future of IMRT for head and neck cancer," (in eng), *Oral Oncol,* vol. 88, pp. 66-74, Jan 2019, doi: 10.1016/j.oraloncology.2018.11.015.

[174] P. Plevová, "Prevention and treatment of chemotherapy- and radiotherapy-induced oral mucositis: a review," (in eng), *Oral Oncol,* vol. 35, no. 5, pp. 453-70, Sep 1999, doi: 10.1016/s1368-8375(99)00033-0.

[175] W. W. P. Melo, W. A. B. Aragão, D. C. Baia-da-Silva, P. C. Nascimento, R. R. Lima, and R. D. de Souza-Rodrigues, "Effects of Photobiomodulation on Oral Mucositis: Visualization and Analysis of Knowledge," (in eng), *Life (Basel),* vol. 12, no. 11, Nov 21 2022, doi: 10.3390/life12111940.

[176] S. N. M. Araújo, M. H. B. A. Luz, G. R. F. d. Silva, E. M. L. R. Andrade, L. C. C. Nunes, and R. O. Moura, "Cancer patients with oral mucositis: challenges for nursing care," *Revista latino-americana de enfermagem,* vol. 23, pp. 267-274, 2015, doi: 10.1590/0104-1169.0090.2551.

[177] Z. Shu *et al.*, "Nutritional Status and Its Association With Radiation-Induced Oral Mucositis in Patients With Nasopharyngeal Carcinoma During Radiotherapy: A Prospective Study," (in eng), *Front Oncol,* vol. 10, p. 594687, 2020, doi: 10.3389/fonc.2020.594687.

[178] L. S. Elting, C. D. Cooksley, M. S. Chambers, and A. S. Garden, "Risk, outcomes, and costs of radiation-induced oral mucositis among patients with head-and-neck malignancies," (in eng), *Int J Radiat Oncol Biol Phys,* vol. 68, no. 4, pp. 1110-20, Jul 15 2007, doi: 10.1016/j.ijrobp.2007.01.053.

[179] B. Norgeot *et al.,* "Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist," *Nature Medicine,* vol. 26, no. 9, pp. 1320-1324, 2020/09/01 2020, doi: 10.1038/s41591-020-1041-y.

[180] C. Hurkmans *et al.,* "A joint ESTRO and AAPM guideline for development, clinical validation and reporting of artificial intelligence models in radiation therapy," (in eng), *Radiother Oncol,* vol. 197, p. 110345, Jun 3 2024, doi: 10.1016/j.radonc.2024.110345.

[181] "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," (in eng), *Bmj,* vol. 385, p. q902, Apr 18 2024, doi: 10.1136/bmj.q902.

[182] J. Zhang *et al.,* "Comparing effectiveness of image perturbation and test retest imaging in improving radiomic model reliability," *Scientific Reports,* vol. 13, no. 1, p. 18263, 2023, doi: 10.1038/s41598-023-45477-6.

[183] T. K. Koo and M. Y. Li, "A guideline of selecting and reporting intraclass correlation coefficients for reliability research," *Journal of chiropractic medicine,* vol. 15, no. 2, pp. 155-163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[184] H. Abdi, "Coefficient of variation," *Encyclopedia of research design,* vol. 1, no. 5, pp. 169-171, 2010.

[185] R. Berenguer *et al.,* "Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters," *Radiology,* vol. 288, no. 2, pp. 407-415, 2018, doi: 10.1148/radiol.2018172361.

[186] H. Hong and S. Hong, "simpleNomo: A python package of making nomograms for visualizable calculation of logistic regression models," *Health Data Science,* vol. 3, p. 0023, 2023, doi: 10.34133/hds.0023.

[187] M. Qiang *et al.,* "A Prognostic Predictive System Based on Deep Learning for Locoregionally Advanced Nasopharyngeal Carcinoma," (in eng), *J Natl Cancer Inst,* vol. 113, no. 5, pp. 606-615, May 4 2021, doi: 10.1093/jnci/djaa149.

[188] D. F. Lin *et al.,* "Radiomic signatures associated with tumor immune heterogeneity predict survival in locally recurrent nasopharyngeal carcinoma," (in eng), *J Natl Cancer Inst,* Apr 19 2024, doi: 10.1093/jnci/djae081.

[189] Y. Lin *et al.,* "A contrast-enhanced CT radiomics-based model to identify candidates for deintensified chemoradiotherapy in locoregionally advanced nasopharyngeal carcinoma patients," *European Radiology,* vol. 34, no. 2, pp. 1302-1313, 2024, doi: 10.1007/s00330-023-09987-1.

[190] R. Zhao, Z. Liang, K. Chen, and X. Zhu, "Nomogram based on inflammatory biomarkers and nutritional indicators for predicting overall survival in locoregionally advanced nasopharyngeal carcinoma," *Journal of Inflammation Research,* pp. 2971-2981, 2022, doi: 10.2147/JIR.S366299.

[191] M. Bologna *et al.,* "Baseline MRI-radiomics can predict overall survival in non-endemic EBV-related nasopharyngeal carcinoma patients," *Cancers,* vol. 12, no. 10, p. 2958, 2020, doi: 10.3390/cancers12102958.

[192] L. Chen, H. Wang, H. Zeng, Y. Zhang, and X. Ma, "Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma," *Cancer Imaging*, vol. 20, pp. 1-9, 2020, doi: 10.1186/s40644-020-00310-5.
[193] M. Akcay, D. Etiz, O. Celik, and A. Ozen, "Evaluation of prognosis in nasopharyngeal cancer using machine learning," *Technology in Cancer Research & Treatment*, vol. 19, 2020, doi: 10.1177/1533033820909829.
[194] X. Ming *et al.*, "MRI-based radiomics signature is a quantitative prognostic biomarker for nasopharyngeal carcinoma," *Scientific reports*, vol. 9, no. 1, p. 10412, 2019, doi: 10.1038/s41598-019-46985-0.
[195] A. W. Lee *et al.*, "International guideline for the delineation of the clinical target volumes (CTV) for nasopharyngeal carcinoma," (in eng), *Radiother Oncol*, vol. 126, no. 1, pp. 25-36, Jan 2018, doi: 10.1016/j.radonc.2017.10.032.
[196] L. Gilbeau, M. Octave-Prignot, T. Loncol, L. Renard, P. Scalliet, and V. Grégoire, "Comparison of setup accuracy of three different thermoplastic masks for the treatment of brain and head and neck tumors," (in eng), *Radiother Oncol*, vol. 58, no. 2, pp. 155-62, Feb 2001, doi: 10.1016/s0167-8140(00)00280-2.