Copyright Undertaking

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

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

1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.

2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.

3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

IMPORTANT

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

MULTI-ORGAN MULTI-OMICS PREDICTION OF ADAPTIVE RADIOTHERAPY ELIGIBILITY IN PATIENTS WITH NASOPHARYNGEAL CARCINOMA

LAM, SAI KIT

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

Multi-organ Multi-omics Prediction of Adaptive Radiotherapy Eligibility in Patients with Nasopharyngeal Carcinoma

LAM, Sai Kit

A thesis submitted in partial fulfillment of the requirements

for the degree of Doctor of Philosophy

February 2022

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 aknowlegement has been made in the text.

_____(Signed)

_____Lam, Sai Kit_____(Name of Student)

Abstract

Intensity-modulated radiotherapy (IMRT) is a standard-of-care for advanced nasopharyngeal carcinoma (NPC) patients. The success of treatment relies on an assumption that patient anatomy remains throughout the entire IMRT course. In response to treatment perturbations, however, tumors and surrounding healthy organs may exhibit significant morphometric volume and/or geometric alterations, which may jointly alter patient anatomy and jeopardize treatment efficacy. Adaptive Radiotherapy (ART) can compensate for these patient-specific variations. Nevertheless, most of existing ART triggers require close monitoring throughout the IMRT course, and are deficient in capturing inter-patient disparity in intrinsic biologic tissue response. Therefore, effective pre-treatment prediction of ART eligibility is greatly demanding.

In this study, various machine learning techniques was applied to investigate capability of a variety of prediction models, developed by using different types of "-omics" features extracted from various organ structures, for pre-treatment prediction of ART demand in NPC patients, with an ultimate objective to facilitate ART clinical implementation in the future.

First, 124 and 58 NPC patients from Queen Elizabeth Hospital (QEH) and Queen Mary Hospital (QMH), respectively, were retrospectively analyzed. Radiomic features extracted from neck nodal lesions of Computed Tomography (CT) images, clinical data, and combined types of features were used for developing R, C, and RC models, respectively, for predicting ill-fitted thermoplastic mask (IfTM)-triggered ART event. Results showed that the R model performed significantly better than the C model in the external QMH testing cohort ($p<0.0001$), while demonstrating no significant difference compared to the RC model ($p=0.5773$). Second, pre-treatment contrast-enhanced T1-

4

weighted (CET1-w), T2-w magnetic resonance (MR) images of seventy NPC patients from QEH were processed for extraction of radiomic features from Gross-Tumor-Volume of primary NPC tumor, for developing CET1-w, T2-w, and joint T1-T2 models. Results indicated promising predictability of MR-based tumoral radiomics, with AUCs ranging from 0.895–0.984 in the training set and 0.750–0.930 in the testing set. Third, pre-treatment CECT and MR images, radiotherapy dose and contour data of one-hundred and thirty-five NPC patients treated at QEH were retrospectively analyzed for extraction of multi-omics features, namely Radiomics (R), Morphology (M), Dosiomics (D), and Contouromics (C), from eight organ structures. Four single-omics models (R, M, D, C) and four multi-omics models (RD, RC, RM, RMDC) were developed. Results demonstrated that the R model significantly outperformed all other three single-omics models (all p-value<0.0001), achieving an average AUC 0.918 (95%CI: 0.903-0.933) in hold-out test set, respectively. Intriguingly, Radiomic features accounted for the majority of the final selected features (64-94%) in all the studied multi-omics models.

In conclusion, a series of studies in this thesis demonstrated that CT-based neck nodal radiomics was capable of predicting IfTM-triggered ART events in NPC patients undergoing RT, showing higher predictability over clinical predictors. MRI-based tumoral radiomics was shown promising in pre-treatment identification of ART eligibility in NPC patients. Multi-organ multi-omics analyses revealed that the Radiomic model played a dominant role for ART eligibility in NPC patients. The overall findings may provide valuable insights for future study into developing an effective screening strategy for ART eligibility in NPC patients in the future.

Publications

1. **<u>Lam SK</u>**, Zhang YP, Zhang J, Li B, Sun JC, Liu YT Carol, Chou PH Bosco, Teng XZ, Ma ZR, Ni RY, Zhou T, Peng T, Xiao HN, Li T, Ren G Gary, Cheung LY Andy, Lee KH Francis, Yip WY Celia, Au KH, Lee HF Victor, Chang TY Amy, Chan WC Lawrence, Cai J. <u>Multi-organ Omics-based Prediction for Adaptive Radiation Therapy Eligibility in Nasopharyngeal Carcinoma Patients undergoing Concurrent Chemoradiotherapy</u>. *Frontiers in Oncology* (2022). 11:792024. doi: 10.3389/fonc.2021.792024

2. **<u>Lam SK</u>**, Zhang, J., Zhang, Y., Li, B., Ni, R.-Y., Zhou, T., Peng, T., Cheung, A.L.-Y., Chau, T.-C., Lee, F.K.-H., Yip, C.W.-Y., Au, K.-H., Lee, V.H.-F., Chang, A.T.-Y., Chan, L.W.-C., Cai, J. <u>A Multi-Center Study of CT-Based Neck Nodal Radiomics for Predicting an Adaptive Radiotherapy Trigger of Ill-Fitted Thermoplastic Masks in Patients with Nasopharyngeal Carcinoma</u>. *Life* 2022, 12, 241. doi: 10.3390/life12020241

3. Yu T*, **<u>Lam SK*</u>**, To LH, Tse KY, Cheng NY, Fan YN, Lo CL, Or KW, Chan ML, Hui KC, Chan FC, Hui WM, Ngai LK, Lee KH, Au KH, Yip WY, Zhang Y, Cai J. <u>Pre-treatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients</u>. *Frontiers in Oncology,* 16 October 2019, doi.org/10.3389/fonc.2019.01050. **<u>(*Co-first authors)</u>**

4. Yu T*, **<u>Lam SK*</u>**, To LH, Tse KY, Cheng NY, Fan YN, Lo CL, Or KW, Chan ML, Hui KC, Chan FC, Hui WM, Ngai LK, Lee KH, Au KH, Yip WY, Zhang Y, Cai J. <u>Constructing Novel Prognostic Biomarkers of Advanced Nasopharyngeal Carcinoma from Multiparametric MRI Radiomics Using Ensemble-Model Based</u>

Iterative Feature Selection, *2019 International Conference on Medical Imaging Physics and Engineering (ICMIPE)*, Shenzhen, China, 2019, pp. 1-7, doi: 10.1109/ICMIPE47306.2019.9098211. **(*Co-first authors)**

5. Zhang YP, **Lam SK**, Yu T, Teng XZ, Zhang J, Lee KH Francis, Au KH, Yip WY Celia, Wang ST, Cai J. Integration of an Imbalance Framework with Novel High-generalizable Classifiers for Radiomics-based Distant Metastases Prediction of Advanced Nasopharyngeal Carcinoma. *Knowledge-Based Systems. 235* (2021).

6. Ren G, **Lam SK**, Zhang J, Xiao HN, Cheung LY, Ho WY, Qin J, Cai J. Investigation of a Novel Deep Learning-Based Computed Tomography Perfusion Mapping Framework for Functional Lung Avoidance Radiotherapy. *Frontiers in Oncology.* 11(466) (2021).

7. Ren G, **Lam SK**, Ni RY, Yang DR, Qin J, Cai J. The Effectiveness of Data Augmentation for Bone Suppression in Chest Radiograph using Convolutional Neural Network. *Austin Journal of Cancer and Clinical Research* (2021); 8(2): 1095.

8. Ren G, Xiao HN, **Lam SK**, Yang DR, Li T, Teng XZ, Qin J, Cai J. Deep Learning-based Bone Suppression in Chest Radiographs Using CT-derived Features: A Feasibility Study. *Quantitative Imaging in Medicine and Surgery.* 11(12):4807-4819 (2021).

9. Li W, Xiao HN, Li T, Ren G, **Lam SK**, Teng XZ, Liu CY, Zhang J, Lee KH Francis, Au KH, Lee HF Victor, Chang TY Amy, Cai J. Virtual Contrast-enhanced Magnetic Resonance Images Synthesis for Patients with Nasopharyngeal Carcinoma using Multimodality-guided Synergistic Neural Network. *International Journal of Radiation Oncology Biology Physics.* S0360-3016(21)03115-1 (2021).

10. Cheung LY, Zhang L, Liu CY, Li T, Cheung HY, Leung C, Leung KC, **Lam SK**, Lee HF, Cai J. <u>Evaluation of Multisource Adaptive MRI Fusion for Gross Tumor Volume Delineation of Hepatocellular Carcinoma.</u> *Frontiers in Oncology* (2021).

11. Tie X, **Lam SK**, Zhang Y, Lee KH, Au KH, Cai J. <u>Pseudo-CT Generation from Multi-parametric MRI using a Novel Multi-channel Multi-path Conditional Generative Adversarial Network for Nasopharyngeal Carcinoma Patients.</u> *Medical Physics*. 2020 Feb 3. doi: 10.1002/mp.14062.

12. Wang MQ, Zhang Q, **Lam SK**, Cai J, Yang RJ. <u>A Review on Application of Deep Learning Algorithms in External Beam Radiotherapy Automated Treatment Planning.</u> *Frontiers in Oncology*. 10:580919 (2020).

Conference Presentation

1. **Lam SK**, Zhang YP, Zhang J, Suen JC, Teng XZ, Ma ZR, Li B, Ni RY, Zhou T, Peng T, Chou PH Bosco, Liu YT Carol, Cheung LY Andy, Lee KH Francis, Au KH, Yip WY Celia, Chan WC Lawrence, CAI J. <u>Investigating the Role of Multi-organ Multi-omics Prediction Models for Adaptive Radiation Therapy Eligibility in Nasopharyngeal Carcinoma Patients undergoing Concurrent Chemoradiotherapy</u>. *The 17th Chinese Society of Therapeutic Radiation Oncology (CSTRO)*, 2021, Xiamen, China. (Abstract)

2. Li W, **Lam SK**, Xiao HN, Li T, Ren G, Zhi SH, Teng XZ, Liu CY, Zhang J, Lee KH Francis, Au KH, Lee HF Victor, Chang TY Amy, Cai J. <u>Gadolinium-free Contrast-enhanced MRI (GFCE-MRI) Synthesis via Generalizable MHDgN-Net for Patients with Nasopharyngeal Carcinoma.</u> Joint Annual Meeting ISMRM-ESMRMB & SMRT 31st Annual Meeting, 2022. (Abstract)

3. Li B, Zheng XL, Guo W, Zhang J, Teng XZ, Wang YH, **Lam SK**, Zhang YP, Ma ZR, Ge H, Cai J. Improving Pre-treatment Multi-omics Prediction of Radiation Pneumonitis using Dose-threshold Sub-regions in Non-Small Cell Lung Cancer Patients. *The 17th Chinese Society of Therapeutic Radiation Oncology (CSTRO)*, 2021, Xiamen, China. (Abstract)

4. Yu T, Teng X, **Lam SK**, Zhang J, Lee F, Au K, Yip W, Cai J. Investigation of Discriminability of Sub-volumes using Radiomics Predictor of Distant Metastasis in Advanced Nasopharyngeal Carcinoma. *2020 AAPM Annual Meeting*, Vancouver, BC, Canada e-Poster, 2020. (Abstract)

5. Zhang J, Teng XZ, Ma ZR, Yu TT, **Lam SK**, Lee KH Francis, Au KU, Yi WY, Cai J. Radiomics Feature Robustness Under Different Image Perturbation Combinations and Intensities: A Study On Nasopharyngeal Carcinoma CT Images. *2020 AAPM Annual Meeting*, Vancouver, BC, Canada. (Abstract)

6. Duarte I, **Lam SK**, Cui T, Miller G, Garrison W, Mugler J, Shim M, Cates G, Yin F, Cai J. Evaluation of DIR-Based Lung Ventilation Imaging Against Hyperpolarized Gas Ventilation MRI and Hyperpolarized Gas Tagging MRI. *Medical Physics*. June 2019. 46(6), E502-E502, TH-A-225BC. (Oral Presentation).

Acknowledgement

I am truly grateful to the support received from various people and organizations, without which this thesis project cannot be performed smoothly and completed on time. First and foremost, I would like to express my sincere gratitude and appreciation to my chief supervisor, Prof. Jing Cai, whose patience, kindness and trust have always been great support and motivation throughout my entire Ph.D. journey. Special thanks are also given to my co-supervisor, Dr. Lawrence Chan, whose instructive guidance and encouragement have driven me to improve quality of this study from technical perspectives.

I would like to especially acknowledge Dr. Au Kwok Hung, Dr. Francis Lee Kar Ho, Dr. Celia Yip Wai Yi from Hong Kong Queen Elizabeth Hospital, Dr. Victor Lee, Ho Fun, Mr. Andy Cheung Lai Yin from Hong Kong Queen Mary Hospital, and Dr. Amy Chang Tien Yee from Hong Kong Sanatorium & Hospital, for their professional advice from clinical perspectives that has allowed me to enhance the methodology of this study, and for their constant administrative support and coordination during data acquisition.

I would like to thank my colleagues for their constructive advice and technical support throughout my thesis project. Dr. Gloria Li, Dr. Zhang Yuanpeng, Dr. Li Bing, Dr. Zhou Ta, Dr. Li Tain, Dr. Ren Ge, Mr. Zhang Jiang, Mr. Teng Xinzhi, and Mr. Sun Jiachen have offered valuable suggestions on methodology of study design and technical support. Special thanks to Mr. Zhang Jiang who developed a sophisticated software that have greatly facilitated the research progress and always showed his keen willingness in supporting my study despite his tight schedule, and to Dr. Zhang

Table of Contents

List of Tables

List of Figures

Abbreviations

18F-FDG	18F-fluorodeoxyglucose
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
AC	Adjuvant chemotherapy
AJCC	American joint committee on cancer
ART	Adaptive radiotherapy
AUC	Area under receiver operating characteristics
BMI	Body mass index
BS	Brainstem
BW	Body weight
CCRT	Concurrent chemoradiotherapy
CECT	Contrast-enhanced computed tomography
CET1-w	Contrast-enhanced T1-weighted

CI	Confidence interval
ContraPG	Contra-lateral parotid gland
CR	Complete response
CRT	Chemo-radiotherapy
CT	Computed tomography
CV	Cross validation
DICOM	Digital imaging and communications in medicine
DNA	Deoxyribonucleic acid
DVH	Dose-volume histogram
EBV	Epstein-Barr virus
FOV	Field-of-view
FS	Feature selection
GLCM	Gray level cooccurrence matrix
GLDM	Gray level difference matrix
GLRLM	Gray level run length matrix

GRSZM	Gray level size zone matrix
GTVn	Gross-tumor-volume of NPC nodal lesions
GTVnp	Gross-tumor-volume of primary NPC tumor
HNC	Head-and-neck cancer
HU	Hounsfield unit
IBSI	Image biomarker standardization initiative
IC	Induction chemotherapy
IfTM	Ill-fitted thermoplastic mask
IMRT	Intensity-modulated radiotherapy
IpsiPG	Ipsi-lateral parotid gland
IV	Intravenous
KRR	Kernel ridge regression
LASSO	Least absolute shrinkage and selection operator
LNs	Lymph nodes
LoG	Laplacian-of-Gaussian

MKL	Multi-kernel learning
MLC	Multi-lead collimator
MR	Magnetic resonance
NCCN	National comprehensive cancer network
NGTDM	Neighboring gray tone difference matrix
NPC	Nasopharyngeal carcinoma
OARs	Organs at risk
OVH	Overlap-volume histogram
PACs	Picture archiving and communication system
PCC	Pearson correlation coefficient
PET	Positron emission tomography
PG	Parotid glands
POV	Projection-overlap-volume
PR	Partial response
PTV	Planning target volume

QEH	Queen Elizabeth Hospital
QMH	Queen Mary Hospital
RECIST	Response evaluation criteria in solid tumors
ROC	Receiver operator characteristic
ROI	Region-of-interest
RT	Radiotherapy
SCC	Spearman's correlation coefficient
SC	Spinal cord
T2-w	T2-weighted
TM	Thermoplastic mask
UICC	Union for International Cancer Control
VOI	Volume-of-interest
WHO	World Health Organisation

1.   Chapter 1: Introduction

1.1   *Advance in Omics Features from Radiotherapy Data*

Omics refers to a discipline of biological sciences that ends with -omics, such as genomics. The addition of -omics to a discipline implies a comprehensive analysis of a set of features, such as genetic features in genomics. Over the past decades, the omics field has been tremendously improved by technological advances which enables cost-effective extraction and analysis of high-throughput features, and it has been largely expanded to a variety of new disciplines in medical research. In the context of radiotherapy (RT), a series of medical resources have been utilized for generation of -omics features for cancer research. Specifically, medical images, such as computed tomography (CT), magnetic resonance (MR) and positron emission tomography (PET) images, have been adopted for massive extraction of Radiomic features; radiation treatment plan data, such as radiation dose and organ contours, have been used for extraction of multi-dimensional dose information, termed Dosiomics features. These high-throughput technologies have been extensively studied in the cancer research for a variety of predictions, including early detection of diseases, recurrence risk stratification, treatment responses, tumor grading, etc.

1.1.1 Radiomics

Radiomics has recently become an emerging field for personalized medicine. Coupled with the advanced machine learning technologies, radiomics has gained tremendous attention in the field of medical oncology [1]. Radiomics involves transformation of digitally encrypted medical images into mineable high-dimensional features, including morphologic shape of the volume of interest, distribution of pixel

intensity and the spatial relationship between adjacent pixel intensity. Application of various imaging filters, such as Laplacian-of-Gaussian filter and Wavelet filter will yield extra features with a wide range of scales and resolutions. A simplified diagram showing overview of different classes of radiomic features, with and without application of imaging filter (Laplacian-of-Gaussian or Wavelet) before feature extraction is given in **Figure 1.** Such multifarious features can subsequently be processed for decoding concealed genetic and molecular traits. Panth et al. conducted an *in vivo* preclinical experiment to investigate relationship between genetic changes and radiomic features in HCT116 doxycycline (dox) inducible GADD34 tumor cells, and reported that the radiomic features was capable of identifying early effects of variation in genetic changes following RT treatment [2]. The capability of radiomics have also been extensively studied in areas of cancer differentiation, prognosis, treatment response and toxicity prediction. For instance, Zhuo et al. combined patient's clinical data and MR-based radiomic features to stratify nasopharyngeal carcinoma (NPC) patients into different survival risk groups, and reported that the developed radiomic model outperformed traditional factor of T-stage (C-index: 0.814 vs 0.803) and Tumor-Nodal-Metastasis staging system (C-index: 0.814 vs 0.765) [3]. Dong et al. developed an individualized nomogram for detecting occult peritoneal metastasis in advanced gastric cancer patients using CT-based radiomic features, and reported a promising model performance with area under the receiver characteristics curve (AUC) of 0.958 (95% confidence interval (CI): 0.923-0.993) in the training cohort, 0.941 (95%CI: 0.904-0.977) in the internal validation cohort, 0.928 (95% CI: 0.886-0.971) and 0.920 (95% CI: 0.862-0.978) in two external validation cohorts [4]. Besides, Wang et al. developed a series of multi-parametric MR-based radiomic signatures for predicting early response of NPC tumor following induction chemotherapy, and

reported that the radiomic signature developed jointly from T1-weighted, contrast-enhanced T1-weighted, T2-weighted and T2-weighted fat-suppressed MR images was the best performing one among those developed solely from either one of the imaging modalities, with the AUC of 0.822 (95% CI: 0.809-0.835), sensitivity of 0.980, specificity of 0.529, positive predictive value of 0.593 and negative predictive value of 0.949, demonstrating capability of radiomics in predicting NPC tumor shrinkage following induction chemotherapy [5].
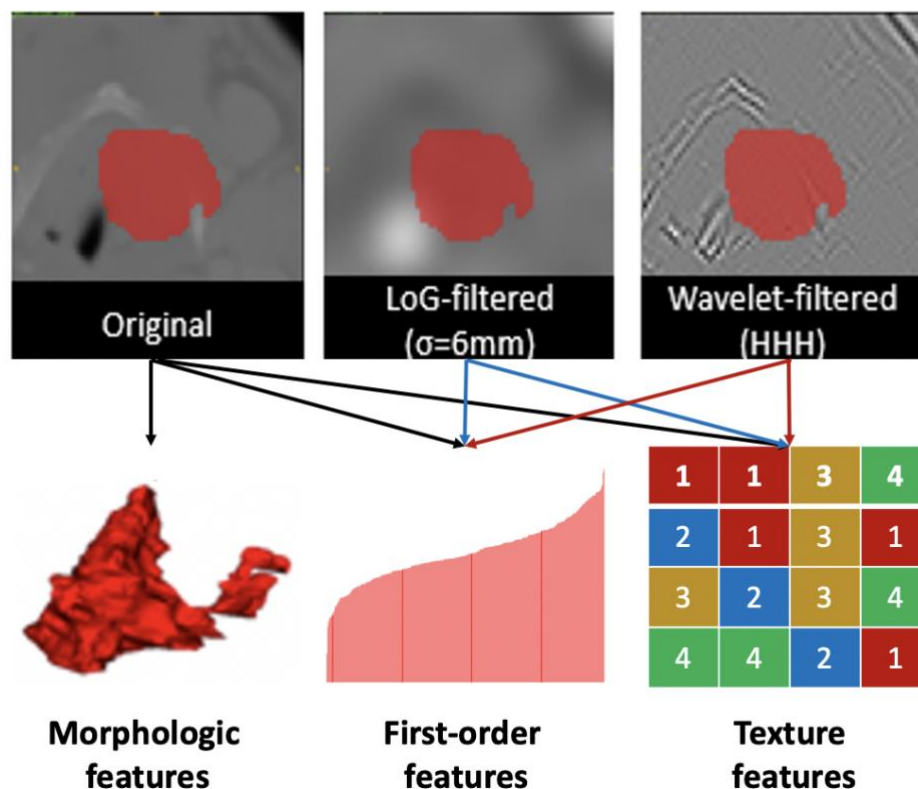
Figure 1. Simplified Diagram showing overview of different classes of radiomic features, with and without application of imaging filter (Laplacian-of-Gaussian or Wavelet) before feature extraction.

1.1.2 Dosiomics

Dosiomics is a natural extraction of Radiomics [6]. It characterizes spatial 3-dimensional (3D) distribution of radiation dose within a tissue. This has overcome the challenges of traditional dosimetric-volume histogram (DVH) parameters, where only 1-dimensional (1D) point dose or 2-dimensional (2D) dose-volume parameters can be obtained, therefore losing statistical and spatial information of the dose delivered to the patient. With the dosiomics features, one can comprehensive depict complex 3D dose distribution by shape-based, intensity-based and texture-based features as in Radiomics analysis. An illustration example of DVH and spatial dose heteogenity within a given volume-of-interest is provided in **Figure 2.** Notably, the area of Dosiomics in cancer predictive research is still in its infant stage. Rossi et al. investigated the potential of dosiomics for prediction of treatment-induced toxicity in 351 prostate cancer patients, and reporting addition of dosiomic features into traditional clinical model significantly improved model predictability for treatment-induced gastrointestinal toxicity [7]. Buizza et al. investigated the role of dosiomics, radiomics, and clinical predictors for predicting prognosis of local control in skull-based chordoma patients treated by carbon-ion RT, and reported that the combined dosiomics-shape based model were the best-performing model with AUC up to 0.80, demonstrating outstanding capability of dosiomic in prognostic risk stratification [8]. Besides, Adachi et al. examined potential of dosiomics in predicting radiation pneumonitis of lung cancer patients who were treated by using stereotactic body radiation therapy, by comparing the dosiomic models with radiomic model, and reported that the dosiomic model outperformed the traditional DVH-based model (AUC: 0.837 ± 0.054 vs 0.660 ± 0.054), and the combined dosiomics-DVH model was the best-performing one, with AUC of 0.846 ± 0.049, highlighting the capability of dosiomics for prediction of radiation pneumonitis [9].
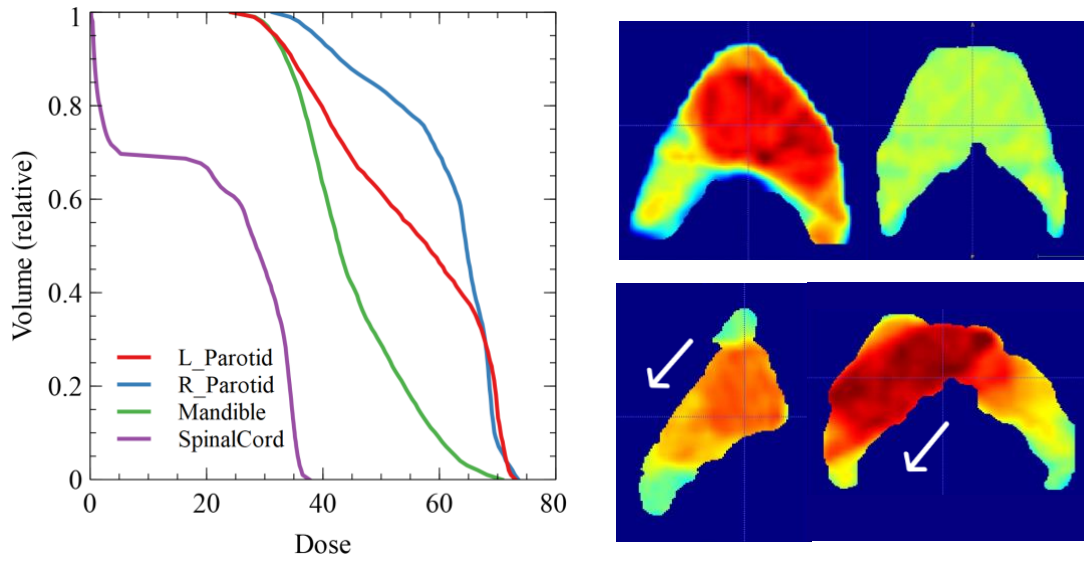
Figure 2. Illustration of Dose-Volume Histogram (left) of several organs-at-risk, and

Dosiomics features of dose heterogeneity within a given volume-of-interest.

1.2   *Epidemiology of Nasopharyngeal Carcinoma (NPC)*

1.2.1 Etiology of NPC

Nasopharyngeal Carcinoma (NPC) has a peculiarly skewed racial and geographic variation in distribution, being highly prevalent in Southeast Asia [10]. Annually, there are 129,000 new NPC cases globally, of which over 70% were found in east and southeast Asia, according to the International Agency for Research on Cancer [11, 12]. The distinct geographic distribution of NPC incidence between China and the rest of the countries around the globe can be seen in **Figure 3** [10]**.** In 2015, NPC incidence is approximately 2.5 times higher in males than that in females, with median age at the time of diagnosis being around 50-52 years old. Specifically, in Hong Kong, an average of 866 newly diagnosed were reported annually from 2007-2016 [11]. The latest statistics from Hong Kong Cancer Registry showed that NPC ranked 10th major cause of cancer death in 2016 [11].
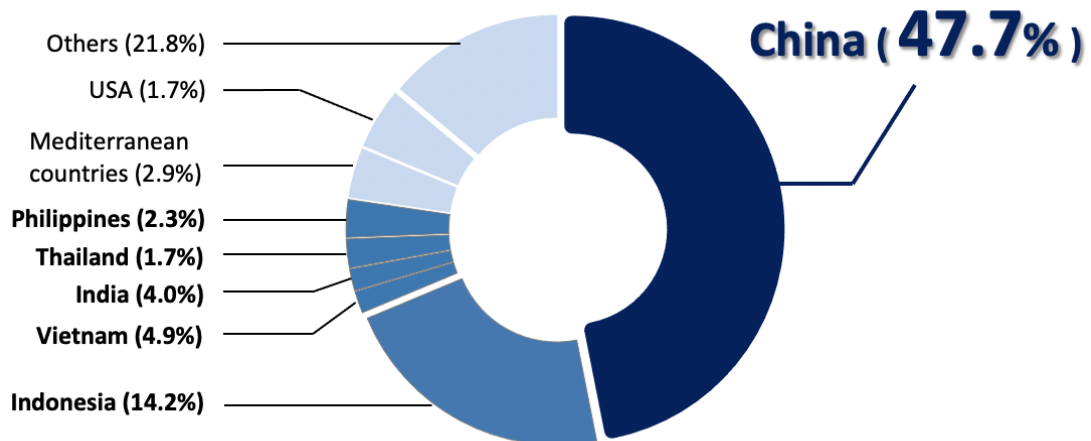
Figure 3. The distinct geographic distribution of NPC incidence between China and the rest of the countries around the globe [10] .

30

Development of NPC is multifactorial in nature. Risk factors of NPC include Epstein-Barr virus (EBV) infection, genetic predisposition, lifestyle and environmental factors [10]. Over 90% of the global population were infected by EBV virus [13], which is prevalently associated with non-keratinising subtype of NPC irrespective of ethnic origin [14]. An escalated level of antigen against the EBV viral capsid antigens is associated with development of NPC and the burden of the disease, and hence is used for NPC screening and treatment response assessment. Apart from this typical risk factor, multiple epidemiological research works have identified lifestyle, environmental factors and family history [11]. In Hong Kong, over 50% reduction in the incidence of NPC has been found compared to three decades ago [15], which could be ascribed to the increased public awareness of this disease and behavioural changes. For instance, over-consumption of salted fish has been identified as a prominent triggering factor for NPC development in endemic regions [13], however, profound changes in lifestyle and consumption pattern of salted fish over the past 30 years has been found in these population, accounting for the concurrent decreasing trend of NPC [13, 15]. Besides, a person who has a family history of NPC has been found to have up to 10-fold higher susceptibility to the development of NPC, compared to those who has no family history of NPC [13]. Other potential risk factors include passive cigarette consumption, over-use of alcohol, and poor oral hygiene [16-20].

1.2.2 Pathology of NPC

NPC can be categorized into three histological types according to the World Health Organisation (WHO). They are keratinising squamous (Type I), non-keratinising carcinoma (Type II) and basaloid squamous (Type III), where the Type II non-keratinising carcinoma can be further divided in to two sub-categories: differentiated and undifferentiated tumor. The incidence of these three histological subtypes demonstrated a distinct geographic distribution, with the Type II being most prevalent (>95%) in endemic regions and Type I being less (<20%) [21-23]. In particular the Type II non-keratinising carcinoma is strongly associated with infection of EBV [10].

1.2.3 Diagnosis of NPC

Due to the deep-seated nature of tumor, NPC is mostly asymptomatic in early stage. As the tumor growth progress and compress nearby functioning tissues and/or organs, several symptoms may be presented in NPC patients, though non-specific. These include headache, nasal obstruction or epistaxis, and tinnitus. In view of this, plasma EBV Deoxyribonucleic acid (DNA) testing is commonly deployed as screening tool for asymptomatic patients. Subjects with elevated level of plasma EBV DNA levels are suspicious for NPC development and a comprehensive head and neck evaluation will follow. This includes nasopharyngoscopy and endoscopic biopsy to assess the pathology of potential lesions for confirming the underlying histologic subtype [24, 25]. In addition, radiographic imaging, including computed tomography (CT), magnetic resonance imaging (MRI) and 18F-fluorodeoxyglucose (18F-FDG) positron emission tomography (PET)/CT can be deployed to further confirm the presence and extension (hence the stage) of the NPC tumor.

1.2.4 Staging of NPC

Stage of NPC is determined in accordance with tumour-node-metastasis (TNM) cancer staging protocol, formulated by the Union for International Cancer Control/the American Joint Committee on Cancer (UICC/AJCC). This staging system mainly take into account anatomic (size) and extension (numbers and invasion) of the primary NPC tumor and metastatic lymph node lesions to evaluate the tumor burden of an individual [10].

Multiple imaging modalities can be adopted to comprehensive evaluate stage, and biologic activity of tumor and lymph node lesions. In particular, MRI is the most commonly used modality for NPC staging. It provides superior soft-tissue contrast, and is ionizing radiation-free in comparison to CT. The high spatial resolution of MRI allows for a better evaluation of disease extension in three dimensions. While 18F-FDG PET/CT offers a desirable workup for assessing presence of distant metastases, though is comparatively less affordable and also accessible in some centers.

1.3 Treatment of NPC

Owing to the deep-seated anatomical position of NPC, operational removal of tumor induces a high risk of traumatic injury to surrounding critical normal tissues, including but not limited to brainstem, spinal cord, and optical nerves. Hence, surgical excision is rarely performed for NPC treatment. Thanks to the nature of high radiosensitivity, indeed, NPC is typically managed using non-invasive external radiation treatment for all stages of diseases. Notably, intensity modulated radiation therapy (IMRT) is currently the standard of care worldwide for local eradication of NPC tumor. On the other hand, chemotherapy presents a promising remedy for eliminating occult cancer cells throughout the entire body and is considered necessary for advanced-stage diseases. Chemotherapy can be administrated in combination of RT, referred to as chemo-radiotherapy or CRT, and chemotherapy can be given before (induction CRT), during (concurrent CRT) and after (adjuvant CRT) CRT treatment. They are described as follows.

1.3.1 Radiotherapy

Radiotherapy (RT) is currently the mainstay treatment for NPC. Over the past few decades, technological advancement, in both hardware and software perspectives, have enabled transformation of radiation delivery from two-dimensional (2D) and 3D conformal radiotherapy (3DCRT) to its current form of intensity-modulated radiation therapy (IMRT). IMRT is superior in terms of its dosimetric properties, enabling drastic enhancement in treatment outcome and reduction in radiation-induced toxicities [26-30]. By adopting the static or dynamic multi-lead collimator (MLC) during dose delivery, IMRT can yield a highly conformal radiation beam to the tumor shape with sharp dose fall-off that allows for sparing of nearby normal tissues [31], enabling higher

dose to be given to the tumor for eradication [32]. In a meta-analysis of over 3,500 patients, IMRT was demonstrated to yield significantly improved prognosis, in aspects of 5-year locoregional rate and overall survival, as compared to 2DRT and 3DCRT. Meanwhile, there were significant declines in incidences of various radiation-induced toxicities in NPC patients, including late xerostomia and temporal lobe damage [30].

1.3.2 Chemoradiotherapy

Local and distant failure are not uncommon in patients with locally-advanced NPC, different forms of chemotherapy in combination of RT is recommended to improve tumor control. Chemotherapy combined with RT (Chemo-RT, CRT) before (induction CRT), during (concurrent CRT, CCRT) and after (adjuvant CRT) CRT treatment is considered essential in treatment locally-advanced NPC patients. According to the National Comprehensive Cancer Network (NCCN) guidelines, concurrent CRT with/without induction or adjuvant chemotherapy is advised for stage II-IVA NPC patients.

Concurrent chemo-RT, referred to as CCRT, has been demonstrated to yield superior prognosis, compared to RT-alone in locally-advanced NPC patients in a copious amount of literatures [33-39]. Cisplatin is the most commonly employed chemotherapeutic agent in NPC treatment [10], which is a platinum-based chemical that prevents DNA synthesis from happening [40]. Adjuvant chemotherapy following CRT aims to eliminate the remnant cancerous cells throughout the blood circulation system. Cisplatin and fluorouracil are considered the typical chemo-agents. A metronomic administration of oral chemotherapeutic drugs following CRT allows regular and tolerated chemotherapy workup that renders less undesired toxicities to patients, this was supported by a number of reasonings regarding inhibition of blood

vessels formation, activation of anti-tumor immunity, etc. [41, 42]. Several retrospective studies have shown that this metronomic administration following CRT significantly reduced risk of post-treatment distant metastasis in high-risk patients [43, 44]. On the other hand, induction chemotherapy followed by CRT presents an appealing and promising treatment strategy in IMRT era for locally-advanced NPC patients. It not only aims to eliminate occult cancerous cells before administration of CRT, but also reduce the burden of initial tumor bed, such that a more intense radiation dose can be prescribed to tumor bed during the CRT treatment. Its efficacy in terms of improved survival has been well documented in the body of literature [45-48].

1.4   Adaptive Radiotherapy (ART)

Typically, radiation treatment plan is customized for individual patients according to their initial pre-treatment anatomy for the sake of maximizing radiation dose delivered to the tumor and metastatic lymph node lesion, meanwhile preventing the surrounding health tissues from excessive radiation exposure. Therefore, any deviations occurring during the course of treatment, either geometrical shift and/or volumetric shrinkage of tumor and/or nearby tissues, would lead to unfavorable dosimetric consequences to the patients, if not adjusted accordingly.

Unfortunately, a large body of literatures has confirmed that prominent volumetric and geometric variations, for instance, shrinkage of tumor or parotid glands, change of body or neck contour due to body weight loss, are not uncommon during the course of radiation treatment for head and neck cancer (HNC) patients [49-55]. Such variations often lead to unanticipated overdose to critical organs or underdose to tumor volume and neck lymph nodes. Modifications of a treatment plan during a course of RT, i.e., replanning or adaptive RT (ART), can compensate for these variations and maintain

satisfactory therapeutic index. The clinical and dosimetric advantages of ART for HNC cancer patients have been widely accepted, and the use of ART worldwide has become commonplace in today's clinical setting, reflected by the increasing number of relevant publications in the past three decades.

1.4.1 Multi-factorial Nature of ART

A copious amount of literature has attempted to identify risk factors for ART implementation. There is accumulating evidence demonstrating significant shrinkage of both primary tumor and neck nodal lesions occur during the course of treatment. A retrospective analysis of 40 pairs of planning CT and re-CT images of 40 NPC re-planned patients reported a 35% of shrinkage of clinical-target-volume of primary tumor [51]. Another studied looked into percentage change in gross-tumor-volume of primary tumor, nodal lesions and combined lesions in HNC patients, and demonstrated median change of 27%, 43% and 31% in respective volumes [56]. Further, a retrospective study investigated average absolute volume change of primary tumor in 159 NPC patients between pre-treatment CT and third week post treatment commencement, and they reported an average change from 45.9 cm^3 to 26.7 cm^3, accounting for over 40% volume reduction. Indeed, when significant tumor shrinkage occurs, the surrounding critical organs might move into the original high-dose region, leading to deleterious dosimetric impact on the surrounding organs [52, 55, 57] and/or insufficient dose delivery to targets [55, 58].

In addition to the response of tumor and nodal lesions to the radiation treatment, normal tissue could react drastically and dynamically during treatment course, causing undesirable risk of radiation safety. Due to the proximity of parotid glands (PG) to the primary tumor, damage to the PG organ is unavoidable, incurring a

series of downstream side effects that necessitate the need for ART implementation. Radiation-induced mucositis is one of the most debilitating radiation-induced toxicities in HNC patients, causing a severe sense of pain and difficulty in swallowing, which tremendously affects patient's ability to intake adequate nutrition and results in significant body weight (BW) loss. A prospective analysis of change in body weight of HNC patients following radiation treatment reported an average of almost 40% of BW loss > 5 kg by treatment completion [59]. Following such BW loss, patients tend to have a reduced skin separation at various levels of cervical spine and neck [60]. This would potentially lead to two major consequences. First, it would cause positional variability during daily setup of treatment position due to possible head movement inside the thermoplastic cast customized prior to treatment commencement. Second, the BW loss is often accompanied with alterations in body and neck contour of patients making the originally planned isodose lines appear deviated from its original position. As a result, it is highly possible that the deviations in contour induce significant dose deviations in target, leading to insufficient dose coverage for disease control, or healthy critical organs, such as spinal cord and brainstem structures, giving rise to deleterious, if not life-threatening, consequences to patients.

1.4.2 Therapeutic Role of ART

Numerous studies have attempted to investigate potential dosimetric impacts, either on tumor and organs at risk (OARs), due to variations in their volume and position during the course of radiation therapy [50, 61]. For instance, Cheng et al. conducted a prospective study of 19 NPC patients for the purpose of revealing volumetric and dosimetric variability during the course of intensity modulated radiation therapy [50]. The 19 NPC patients underwent CT scans when they received 30 Gy and

50 Gy of radiation dose for generating hybrid plans that adapted to the "new" patient anatomy at the corresponding time-points. Cheng et al. reported that there was an average body weight loss of 5.4% and 9.3%, shrinkage of the primary gross tumor volume (GTV) of 9% and 16%, volume reduction of the contra-lateral parotid gland of 0.7 and 3.4 cm3, volume shrinkage of the ipsilateral parotid gland of 5.3 and 8.4 cm3, at the time points of 30 Gy and 50 Gy, respectively. When the two hybrid plans were compared to the original pre-treatment plan, they found that the hybrid plan yielded a significantly higher dose and improved greater inhomogeneity in most target volume of most patients. In addition, radiation damage to several critical OARs, such as maximum dose to the spinal cord and brainstem, and mean dose to the parotid glands, were also found to be larger in the hybrid plans [50]. In another study, Hu et al. conducted a retrospective study of 40 NPC patients for the purpose of identifying potential benefits of ART in NPC patients undergoing volumetric modulated arc radiotherapy [51]. Forty NPC patients received two-phase treatment planning and a second CT scan at the 22nd fraction of radiation for the phase II ART. They reported volume reduction of multiple organs when comparing the adapted plan and the hybrid plan. For instance, there were significant reduction of volume of the ipsilateral parotid gland (23 vs 19 cc, $p<0.001$), contra-lateral parotid gland (23 vs 18 cc, $p<0.001$), clinical target volume (32 vs 21 cc, $p<0.001$), planning target volume (126 vs 107, $p<0.001$). They concluded that ART has a dosimteric advantages for NPC patients who are with higher initial body weight, larger body mass index, profound body weight loss during treatment, treated by concurrent chemo-radiotherapy, and with advanced tumor stage (stage $>=$ III) [51]. The dosimetric and clinical benefits of ART in NPC have been well-demonstrated in several prospective and retrospective studies [50, 55, 61-69].

Mounting evidence have also demonstrated the superiority of ART in enhancing tumor coverage and strengthening normal tissue sparing [63-65, 67, 68].

Apart from the dosimetric perspectives, ART can also bring about clinical benefits to NPC patients. The clinical benefits can generally be observed in two aspects: tumor locoregional control and patient's quality of life. For tumor locoregional control, Luo et al. attempted to compare the long-term outcomes in patients with locally-advanced NPC disease with and without ART implementation [66]. Two hundred NPC patients who were diagnosed with stage T3/T4 NPC were analyzed after a propensity score matching at a ratio of 1:1. They reported that patients who received ART had a prolonged 5-year local-regional recurrence-free survival than those without (96.7 vs 88.1%, p<0.05), while there were no significant differences in terms of distant metastasis-free survival, progression-free survival and overall survival between the two groups. They further reported that the ART implementation was determined to be an independent prognosticator for local-regional control in their study [66]. Other studies shared similar findings [55, 61, 69]. For patient's quality of life, Yang et al. carried out a non-randomized prospective controlled cohort for investigating impact of ART on patient's quality of life in the IMRT era [69]. They analyzed a total of 129 NPC patients (ART: 86, non-ART: 43). Results of their study demonstrated that the ART led to a profound impact on the quality of life of NPC patients, in both global QoL and other QoL scales. Similarly, they also reported that significantly improved local-regional control in patients who received ART, compared to those who did not. While no significant improvement was observed in terms of overall survival [69].

1.4.3   Clinical Practice of ART

Currently, there is no universal protocol for ART implementation for NPC patients in clinic, and the implementation depends of a number of factors. In a local hospital at Hong Kong [70], a daily megavoltage CT (MVCT) or weekly cone beam CT (CBCT) or planar orthogonal X-rays is often taken for all patients to correct for positional variations and to assess anatomic or geometric changes throughout the entire treatment chain. Additionally, weekly records of body weight are also made to assess whether significant body weight loss (BWL) occurred. The Radiation Oncology team will review the above scans on a regular basis, considering BWL of individual patients. When significant BWL occurs, possibly accompanied with noted change in body or neck contour, significant lymph nodes regression and/or loss of neck tissue, an adaptive review process will be initiated, where the original plan will be re-calculated on the MVCT scan for initial dosimetric evaluation to determine whether further actions (re-CT and/or re-plan) or continuous monitoring is appropriate. Patients who do not receive any actions from the first review session will then be treated with the original plan until the next review session for another dosimetric evaluation. On a plan review, radiation oncologist will assess the geometric, volumetric and dosimetric variations of both target and organs at risk (OARs) structures through both visual inspection and dosimetric evaluation. Considerations influencing ART implementation included risks of insufficient primary and nodal targets coverage, overdose to critical organs (such as spinal cord, optic chiasm, and brainstem), increase of high skin dose areas over neck, and unfit of thermoplastic cast for patient immobilization.

1.4.4   Challenges of Current ART Implementation

At present, there is no widely accepted protocol to identify NPC patients for ART ahead of treatment commencement, manual screening process throughout the course of treatment is deemed necessary to determine ART eligibility for each individual patient.

There is a number of limitations regarding the implementation of ART. First, the process is resource intensive and time-consuming due to the demand for repeated imaging, re-delineation, re-planning, and analyzing potential dosimetric impacts as a result of anatomic variations during treatment. All these incur huge clinical burden and elevated cost of patient care to an oncology department. Therefore, implementing ART for each individual patient is considered clinically impractical, particularly in busy centers. Second, there is no existing universal selection protocol for ART due to the multifactorial nature of ART eligibility (e.g. BW loss, change of body and neck contour, tumor shrinkage, etc.). Hence, the current ART practice in most oncology centers is not efficient. Further, the demand for ART can solely be assessed during the RT treatment; when a patient is determined to be ART candidate, the department would need to ad-hoc re-allocate manpower and resources for that particular patient, adding significant burden to the clinic. Without question, pre-therapeutic identification of high-risk NPC patients for ART is clinically favorable in the era of personalized oncology, enabling physicians to prescribe ART for NPC patients more effectively and facilitate resources allocation in clinical settings.

1.5   *Previous Research Works on ART Risk Prediction*

Despite the large body of literatures confirming significant tumor shrinkage of tumor and nodal lesions during the course of treatment, serving as a favorable ART

criterion, there is severe lack of studies that developed selection strategies for ART on the basis of tumor volume reduction.

Radiation dose has long been regarded as a prime attribute for morphometric volume change of tumors, neck lesions and bilateral parotid glands throughout the treatment course. Bahl et al. prospectively analyzed volumetric alterations in 20 NPC patients between pre-treatment computed tomography (CT) and mid-treatment CT at the 17th fraction [62]. They reported approximately 30% shrinkage of high-risk gross-tumor-volume (GTV), which was accompanied with a significantly increased median dose of 7.2-7.7 Gy to and reduced volume of bilateral parotid glands. Another prospective study by Cheng et al. demonstrated that the anatomic tissue shrinkage was dependent on radiation dose received. They analyzed repeated planning CT and magnetic resonance images (MRI) at 30-Gy and 50-Gy intervals and reported that the shrinkage of both primary NPC tumor and nodal lesions against pre-treatment baselines were higher when 50-Gy was delivered (13% and 29%, respectively) than that when 30-Gy was given (9% and 16%, respectively) and a similar trend was also observed for bilateral parotid glands [50]. Further evidence was also observed by Hu et al. who analyzed 40 re-planned NPC patients and confirmed the significant shrinkage of 35% in clinical-target-volume [51], and by Murat et al. who reported a median reduction of 27% and 43% in primary and nodal GTV, respectively, in 48 re-planned head-and-neck cancer patients [56]. A small-sized retrospective study could be found in the literature. They attempted to devise a decision tree for tumor shrinkage using 48 HNC patients. They incorporated initial tumor volumes and a number of clinical factors, achieving satisfying accuracy of 88% [48]. However, there is a number of deficiencies in their work. First, due to the limited sample size of only 48 patients, their results remain to be further validated in a large patient cohort. Second, some of the clinical factors in their

model, such as tumor growth pattern (endophytic or exophytic), are not widely used in many other medical centers, thereby potentially impeding generalizability of their decision tree model. Lastly, if not most importantly, their work was not intended to predict the need of ART ahead of treatment commencement, rather they predicted the risk of tumor shrinkage.

Recently, Radiomics appears to be an emerging field for personalized medicine. Coupled with the advanced machine learning technologies, radiomics has gained tremendous attention in the field of medical oncology [1]. Radiomics involves transformation of digitally encrypted medical images into mineable high-dimensional features, including morphologic shape of the volume of interest, distribution of pixel intensity and the spatial relationship between adjacent pixel intensity. Such multifarious features can subsequently be processed for decoding concealed genetic and molecular traits [1]. The capability of radiomics in areas of cancer differentiation, prognosis, treatment response and toxicity prediction have been well-documented in the body of literature [4, 71-74].

In particular, there is mounting evidence in the literature showing the power of Radiomics in predicting treatment response on the ground of volume shrinkage in various cancer diseases [5, 75-81], which has laid great foundation for Radiomics prediction of ART demand in cancer patients. For instance, Hou et al. investigated CECT-based biomarkers for prediction of therapeutic response to chemo-radiotherapy in esophageal carcinoma and reported the discriminability of their model in AUC ranging from 0.686 to 0.727 [77]. Wang et al. developed a radiomic signature combining features from multi-modal MR imaging sequences for prediction of early treatment response to induction chemotherapy in NPC patients, achieving an AUC of

0.822 [5]. Piao et al. devised a MR-based radiomic model to distinguish sensitive and resistant tumors in NPC patients following induction chemotherapy, yielding an AUC of 0.905 [81]. In these studies, the tumor response was defined in accordance with the Response Evaluation Criteria in Solid Tumors (RECIST) via quantitative assessment of tumor shrinkage, which follows the same line of thought as in this present study. All the above evidence indicates the outstanding capability of Radiomics in divulging patient-specific intrinsic tissue biologic characteristics for discerning respondent and non-respondent cancer patients upon treatment perturbations, laying great foundation for predicting patient-specific anatomic variations for ART eligibility.

Nevertheless, severely limited effort has yet been made to determining ART eligibility for cancer patients. A radiomics study, for the first time, tried to predict tumor shrinkage in 91 lung cancer patients and therefore predict the ART on the ground that they intentionally performed ART solely on the basis of shrinkage of planning target volume (PTV) on weekly CT scans [56]. This, however, is not a routine clinical practice and not applicable in NPC in view of the multifactorial ART eligibility, including dose tolerance of surrounding critical organs, change in neck contours and variability in treatment setup position. Further, the radiomic features in their study were extracted in PTV of lung tumor, where a significant portion of air was included, potentially limiting inference of their radiomic signatures in relation soft-tissue tumor response.

2. Chapter 2: Research Gap and Study Aims

According to the literature review conducted in previous sections, a series of research gaps have been identified.

Research gap 1: The volumetric shrinkage of neck nodal lesions following radiotherapy is a key contributor for ill-fitted thermoplastic mask (IfTM), a routinely used immobilization device for safeguarding radiation delivery accuracy and safety. To deal with this, an ad hoc ART may be required to ensure accurate and safe radiation delivery and to maintain treatment efficacy. Presently, the entire procedure for evaluating an eligible ART candidate is time-consuming, resource-demanding, and highly inefficient. Further, the degree of nodal lesion shrinkage following treatment has been shown to be heterogenous between patients. Therefore, in the artificial intelligence paradigm, an effective personalized pre-treatment identification of NPC patients at risk for IfTMs has become greatly demanding for achieving efficient ART eligibility screening, while no relevant studies have been reported.

Study aim 1: To address the above research gap, I carried out an experiment to investigate the capability of CT-based neck nodal radiomics for predicting IfTM-triggered ART events in NPC patients via a multi-center setting. The developed radiomic model was compared with traditional clinical predictors. This research is presented in Chapter 3 of this thesis report.

Research gap 2: There is enormous amount of studies in the literature demonstrating significant volume reduction of primary NPC tumor throughout the radiotherapy course. Indeed, when significant tumor shrinkage occurs, those critical organs might move into the original high dose region, leading to deleterious dosimetric

impact on the surrounding organs and/or insufficient dose delivery to targets. In case of severe deviations, ART needs to be performed. In view of the tedious procedure for evaluating an eligible ART candidate, there is a pressing demand for an effective pre-treatment evaluation for NPC patients. Pre-treatment multi-parametric MR images are standard-of-care of radiotherapy workflow in NPC patients and their capability in highlighting tissue anatomic and physiologic information is superior than CT images, however, MR-based radiomics in ART prediction have not yet been investigated.

Study aim 2: To address the above research gap, I performed an experiment to develop contrast-enhanced T1-weighted (CET1-w) and T2-weighted (T2-w) MR-based risk prediction model for ART, separately, and in combination, to evaluate their efficacy in assessing ART eligibility in NPC patients. This research is presented in Chapter 4 of this thesis report.

Research gap 3: It has been discovered that not only primary tumors and neck nodal lesions may experience the above-mentioned volumetric changes following radiotherapy, surrounding healthy organs (such as parotid glands) may exhibit significant morphometric volume and/or geometric alterations, all of which may then individually and jointly alter patient anatomy and jeopardize the efficacy of the original treatment plan. Further, three-dimensional spatial dose distribution within the pertinent organ structures (reflected by Dosiomics), initial morphologic characteristics of pertinent organs (reflected by morphologic features), and initial geometric relationship between different internal organs (reflected by Contouromics), may also work in conjunction with radiomics attributes for determining the final patient anatomy that necessitates ART implementation. However, no relevant study has been reported before.

Study aim 3: To address the above research gap, I conducted a series of experiments to develop a variety of single-omics models (Radiomics, or Dosiomics, or Contouromics) for different pertinent organ structures, and various multi-omics models, with a hope to identify the role of different omics-based models in prediction of ART eligibility in NPC patients. This research is presented in Chapter 5 of this thesis report.

3. Chapter 3: A Multi-Center Study of CT-Based Neck Nodal Radiomics for Predicting an Adaptive Radiotherapy Trigger of Ill-Fitted Thermoplastic Masks in Patients with Nasopharyngeal Carcinoma

Introduction: Significant lymph node shrinkage is common in patients with nasopharyngeal carcinoma (NPC) throughout radiotherapy (RT) treatment, causing ill-fitted thermoplastic masks (IfTMs). To deal with this, an ad hoc adaptive radiotherapy (ART) may be required to ensure accurate and safe radiation delivery and to maintain treatment efficacy. Presently, the entire procedure for evaluating an eligible ART candidate is time-consuming, resource-demanding, and highly inefficient. In the artificial intelligence paradigm, the pre-treatment identification of NPC patients at risk for IfTMs has become greatly demanding for achieving efficient ART eligibility screening, while no relevant studies have been reported. Hence, we aimed to investigate the capability of computed tomography (CT)-based neck nodal radiomics for predicting IfTM-triggered ART events in NPC patients via a multi-center setting. **Methods:** Contrast-enhanced CT and the clinical data of 124 and 58 NPC patients from Queen Elizabeth Hospital (QEH) and Queen Mary Hospital (QMH), respectively, were retrospectively analyzed. Radiomic (R), clinical (C), and combined (RC) models were developed using the ridge algorithm in the QEH cohort and evaluated in the QMH cohort using the median area under the receiver operating characteristics curve (AUC). Delong's test was employed for model comparison. Model performance was further assessed on 1000 replicates in both cohorts separately via bootstrapping. **Results:** The R model yielded the highest "corrected" AUC of 0.784 (BCa 95%CI: 0.673–0.859) and 0.723 (BCa 95%CI: 0.534–0.859) in the QEH and QMH cohort following bootstrapping, respectively. Delong's test indicated that the R model performed

significantly better than the C model in the QMH cohort (p < 0.0001), while demonstrating no significant difference compared to the RC model (p = 0.5773). To conclude, CT-based neck nodal radiomics was capable of predicting IfTM-triggered ART events in NPC patients in this multi-center study, outperforming the traditional clinical model. **Conclusion:** The findings of this study provide valuable insights for future study into developing an effective screening strategy for ART eligibility in NPC patients in the long run, ultimately alleviating the workload of clinical practitioners, streamlining ART procedural efficiency in clinics, and achieving personalized RT for NPC patients in the future.

3.1   *Background*

Radiotherapy (RT) is a cornerstone modality for nasopharyngeal cancer (NPC) patients [10, 15], among which the involvement of neck lymph nodes (LNs) is of high prevalence [82]. Irradiation down to the cervical LNs, in addition to the primary NPC tumor, is essential for achieving thorough cancer eradication and mitigating the risk of cancer recurrence [83, 84]. Throughout the 6–7 weeks of a RT course, a thermoplastic mask (TM) immobilization device that provides full coverage of the head and bi-lateral shoulders is deployed for each NPC patient to ensure reproducible patient positioning between RT fractions in order to maintain treatment efficacy [85]. However, anatomic variations and body weight loss of NPC patients are not uncommon [49, 54, 70, 86-89], posing a risk of TM unfit. In cases of ill-fitted TMs (IfTMs), an ad hoc adaptive radiotherapy (ART) may be triggered to ensure accuracy and safe radiation delivery and to maintain treatment efficacy [55, 63, 64, 68]. Presently, clinical ART practice is still in its infant stage. The entire procedure for evaluating an eligible ART candidate

50

is time-consuming, resource-demanding, and requires multidisciplinary efforts [88, 90]. In the artificial intelligence paradigm, the pre-treatment identification of NPC patients at risk for IfTMs has become greatly demanding for the sake of improving medical resource allocation and achieving greater procedural efficiency in oncologic care delivery.

The volumetric shrinkage of neck nodal lesions is a key factor for IfTM. Since neck nodal lesions locate in a close proximity to the body skin surface of NPC patients, the significant shrinkage of neck LNs in response to treatment would cause a palpable change in the patient's neck contour, producing a TM-to-skin air gap which in turn elevates the risk of intra-fractional patient movement during RT delivery and hence jeopardizes treatment efficacy [85, 91-93]. A representative NPC patient showing remarkable neck lymph node shrinkage during the course of RT treatment, reflected by the reduced lateral diameter of the neck region can be seen in **Figure 4**. Indeed, there is mounting evidence indicating that substantial LN volume shrinkage occurs throughout the RT course in NPC and head-and-neck cancer (HNC) patients, potentially triggering ART implementation [49, 50, 54, 56, 94]. For example, Wing et al. quantified the anatomic changes of 30 NPC patients and reported that there was a significant regression of neck volumes over time with a mean loss rate of 0.39 ± 0.15%/day and a mean volume loss of 11.91 ± 5.57% upon treatment completion [94]. Murat et al. reported that there was a 43% reduction of neck nodal target volumes in HNC patients undergoing RT [56]. Similarly, Cheng et al. re-viewed both mid-treatment computed tomographic (CT) and magnetic resonance (MR) scans of NPC patients and showed that there were up to 30% reductions in the volume of nodal lesions

[50]. Despite the above evidence, inter-patient heterogeneity in treatment response has impeded the accurate individualized prediction of tumor shrinkage for decades.
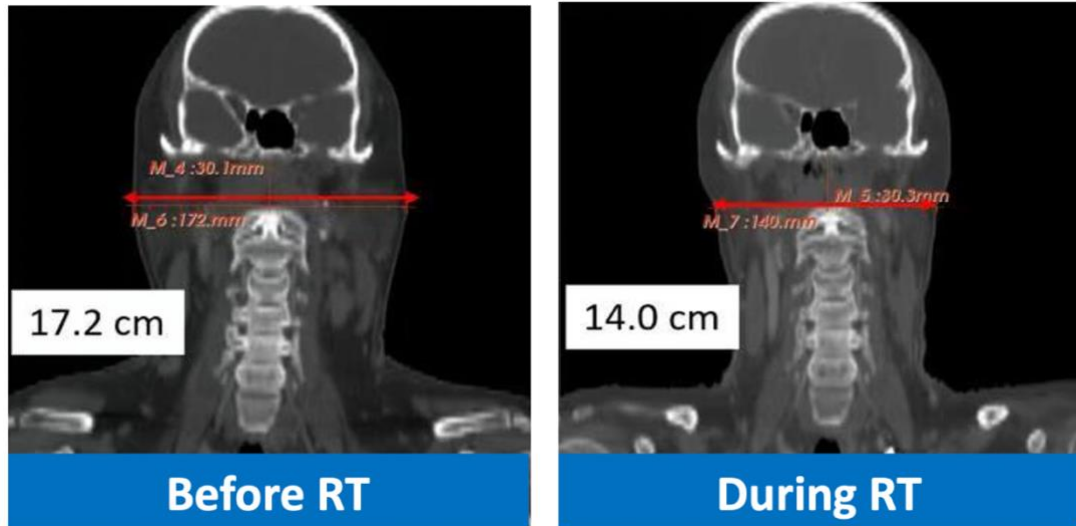
Figure 4. A representative NPC patient showing remarkable neck lymph node shrinkage during the course of RT treatment, reflected by the reduced lateral diameter of the neck region from "Before RT" to "During RT".

Recently, radiomics, which involves the extraction of high-throughput quantitative features from medical images, has become an emerging area for divulging the intrinsic biologic and genetic characteristics of tissue for individual cancer patients [1, 95-97]. Radiomics has been extensively studied for treatment response prediction in various cancer types on the basis of Response Evaluation Criteria in Solid Tumors (RECIST), where criteria are determined by the extent of tumor shrinkage following treatment [5, 75-81]. For instance, Hou et al. investigated CT-based biomarkers for the prediction of the therapeutic response to chemoradiotherapy in esophageal carcinoma and reported that the discriminability of their model achieved area under the receiver operating characteristics curves (AUC) ranging from 0.686 to 0.727 [77]. Wang et al. developed

a radiomic signature combining features from multi-modal MR imaging sequences for the prediction of an early treatment response to induction chemotherapy in NPC patients, achieving an AUC of 0.822 [5]. Piao et al. devised an MR-based radiomic model to distinguish sensitive and resistant tumors in NPC patients following induction chemotherapy, yielding an AUC of 0.905 [81]. These research efforts have laid a great foundation for the radiomics prediction of ART eligibility in cancer patients. Ramella et al. performed a radiomic analysis on pre-treatment CT images of replanned non-small cell lung cancer patients and generated a radiomic signature for the prediction of tumor shrinkage during chemoradiotherapy, yielding an AUC of 0.82 [56]. Yu et al. investigated MR-based radiomics from primary NPC tumors for predicting ART eligibility in a single cohort, achieving AUCs ranging from 0.75 to 0.93 [70].

Unlike MR imaging, CT is often the first-line modality for the neck nodal imaging of NPC patients. In this study, we aimed to investigate the capability of CT-based neck nodal radiomics for predicting IfTM-triggered ART events in NPC patients via a multi-center setting. The main contributions of this study can be presented in three aspects: First, the application of CT-based neck nodal radiomics for developing a prediction model for IfTM-triggered ART events in NPC patients is proposed for the first time. Second, the multi-center setting of this study allows for the assessment of model generalizability across medical institutions. Third, the use of radiomics renders the possibility for the pre-treatment identification of NPC patients who are at a greater risk of experiencing IfTM-triggered ART events, potentially alleviating the workload of clinical practitioners, streamlining ART procedural efficiency in clinic, and achieving personalized RT for NPC patients in the future.

3.2   Methods

3.2.1 Patient Data

A total of 261 NPC patients who received RT at Hong Kong Queen Elizabeth Hospital (QEH) between 2012 and 2015 and 160 NPC patients who received RT at Hong Kong Queen Mary Hospital (QMH) between 2012 and 2020 were retrospectively screened for study eligibility. Patient informed consent was waived due to the retrospective nature of this study. Patients who had biopsy-proven primary NPC without the existence of distant metastasis and co-existing tumors of other types at diagnosis and who received curative concurrent chemoradiotherapy (CCRT) were included in this study. Patients who were treated by induction chemotherapy or did not have a complete set of clinical/image data were excluded from this study. The clinical records of all the enrolled patients were input by the attending radiation oncologists and were carefully examined to determine the binary prediction outcome in this study. Patients who had clinical records regarded as ill-fitted with the TM, necessitating the implementation of ART, were labelled as 1, and were otherwise labeled as 0.

3.2.2 Image Acquisition and Volume-of-Interest (VOI) Definition

All the planning contrast-enhanced CT (CECT) images were retrospectively collected in the format of Digital Imaging and Communications in Medicine (DICOM) and archived using a picture archiving and communication system (PACs).

At QEH, intravenous (IV) CECT simulation was performed in a supine position with an immobilization thermoplastic cast. This was typically acquired at 3 mm intervals from the vertex to 5 cm below the sternoclavicular notch under a 16-slice

Brilliance Big Bore CT scanner (Philips Medical Systems, Cleveland, OH). CECT acquisition parameters were as follows: scan mode = helical, voltage = 120 kVp, X-ray tube current = 264 mA, exposure = 325 msec, pixel spacing = 1.152×1.152-mm, slice thickness = 3 mm, and matrix = 512×512 pixels. Two types of IV contrast agents were available: (i) OMNIPAQUE TM 350 mg I/mL and (ii) VISIPAQUE TM 320 mg I/mL; either one of them was prescribed to each eligible patient and was injected at a rate of 2 mL/sec for 70 mL, followed by scanning after a 30 sec delay.

At QMH, IV CECT simulation was performed in a supine position with an immobilization thermoplastic cast. This was typically acquired at 3 mm intervals from the vertex to 5 cm below the sternoclavicular notch under a 16-slice GE Discovery RT CT scanner (General Electric). CECT acquisition parameters were as follows: scan mode = helical, voltage = 120 kVp, X-ray tube current = 40 mA, exposure = 325 msec, pixel spacing = 1.152×1.152 mm, slice thickness = 2.5 mm, and matrix = 512×512 pixels. IV contrast agents OMNIPAQUE TM 300 mg I/mL was prescribed to patient and the injection rate was 2 mL/sec for 100 mL, followed by scanning after a 20 sec delay.

In both hospitals, the gross-tumor-volume of NPC nodal lesions (GTVn) was chosen as the VOI for the extraction of radiomic features in this study, which was manually delineated on axial CT slices by experienced radiation oncologists specializing in head-and-neck cancers with accreditations, using Eclipse ARIA (Varian Medical System, Inc.) version 13 at QEH and version 13.6 at QMH.

3.2.3 Image Pre-processing

Given that variations in image acquisition and reconstruction parameters within and between medical centres exist, image pre-processing prior to radiomic feature extraction is of paramount importance for optimizing the consistency and hence reproducibility and validity of radiomics studies. In this study, four key image pre-processing steps were involved, including voxel size resampling, VOI re-segmentation, image filtering, and the quantization of grey levels. All these steps were performed in accordance with well-accepted recommendations from the Image Biomarker Standardization Initiative (IBSI) guidelines [98], using an in-house developed pipeline tool based on Python v3.7.3.

First of all, the CECT images were resampled to a voxel size of $1 \times 1 \times 1$ mm3 using linear interpolation to correct for different imaging voxel spacing and slice thicknesses among the different institutions. VOI re-segmentation was then performed to confine the Hounsfield unit (HU) to the range of (-150,180) within the VOI for eliminating non-tumor components, such as air cavities and bony structures. Subsequently, Laplacian-of-Gaussian (LoG) filters with various Gaussian radius parameters of 1 mm, 3 mm, and 6 mm were deployed to produce filtered images for obtaining multi-scale texture features, from fine to coarse. The quantization of image grey levels was applied to normalize the image signal intensities. Grey-level intensities of the images were discretized with various settings of fixed bin counts, ranging from 50 to 350 with an incremental interval of 50.

3.2.4 Feature Extraction

The extraction of radiomic features was performed using the publicly available PyRadiomics v2.2.0 and SimpleITK v1.2.4 package, which was embedded in the in-house developed Python-based v3.7.3 pipeline. Radiomic features were calculated from GTVn on CECT images, with and without LoG filters applied. They can be divided into three major families: shape, first-order statistics, and texture features, which can be further categorized into gray level difference matrix (GLDM), gray level cooccurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), and neighboring gray tone difference matrix (NGTDM) classes. In this study, a total of 2130 radiomic features, including 14 shape features, 72 first-order statistics, and 2044 texture features, were extracted from raw and LoG-filtered images under the pre-defined bin count settings.

3.2.5 Model Development and Evaluation

The eligible patients from QEH were used for the development and internal validation of the prediction model, while those from QMH were used for an external independent evaluation of the trained model. For QEH, the eligible patients were randomly stratified into a training dataset and a validation dataset with 20 iterations, generating an ensemble feature set. For QMH, all the eligible patients were employed as an external testing set for assessing the model generalizability across medical centres. All the development and evaluation processes of the radiomic model were conducted by using R software v3.6.3.

For the radiomic (R) model, the entire QEH cohort was partitioned into a training set (~75%, n = 100) and an internal validation set (~25%, n = 35) with 20 iterations. In

each iteration, z-score normalization was first applied to the training set to scale the value of each of the extracted features to a mean of 0 and standard deviation of 1, which was then applied to the internal validation set. In the training set, Spearman's correlation (SCC) analysis was performed using the "caret" package to assess inter-feature correlation; features which had an SC coefficient equal to or larger than 0.8 and had the greater mean absolute value between a pair of features were excluded. Unpaired two-sided Mann–Whitney U analysis was carried out using the "wilcox.test" function to examine the clinical association of each individual feature; features which showed a clinical association with a p-value of less than or equal to 0.1 were retained. Hence, a set of remnant features that had strong a clinical association to the prediction outcome and were free of highly redundant features was formed under each of the 20 iterations. A combination of these 20 feature sets formed an ensemble feature set, in which each feature was ranked according to its frequency of occurrence under the 20 iterations. Features with a higher frequency were retained for downstream model development. The maximum number of features in the final model was set to 10% of the training sample size [99-101]. In the case of exceeding the amount of features, features that had the least frequency of occurrence were excluded from the final R model. Ridge regression was employed for model development using the "glmnet" package in the R software. Meanwhile, a 10-fold cross validation was performed within the training set of each iteration to minimize the risk of model overfitting; nine of them were used for model training, followed by internal validation on the remaining partition. The final R model was selected when the model predictability reached its maximum on the validation cohort. A simplified workflow of the radiomic model development is illustrated in **Figure 5**.
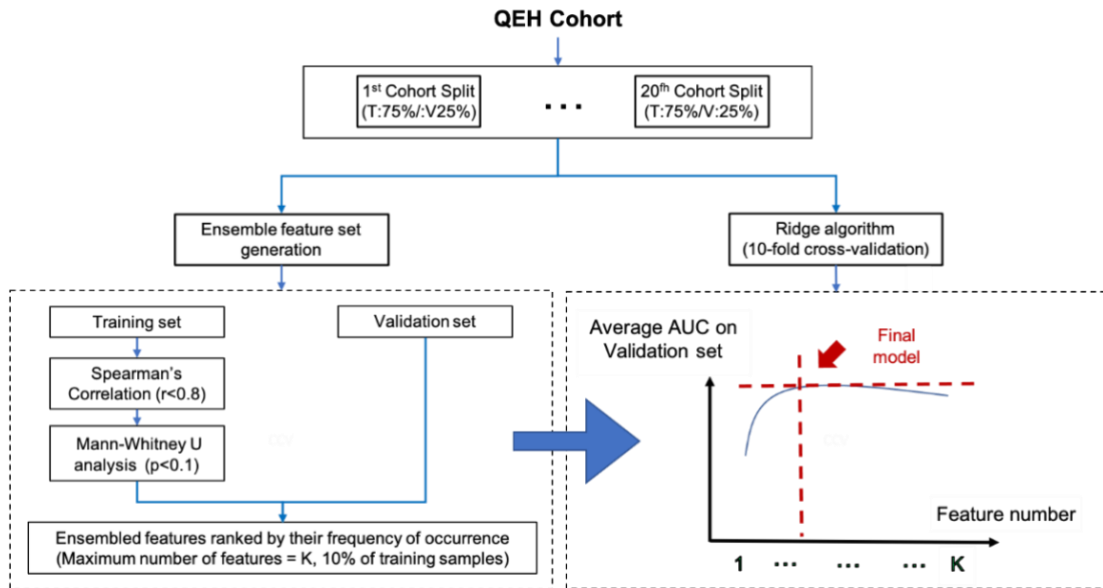
Figure 5. Workflow for radiomic model development.

For the clinical (C) model, a series of clinical data, including the patient's gender, age, volume of GTVn, T-stage, N-stage, whether the patient had T-stage ≥ 3, N-stage ≥ 2, T-stage ≥ 3 plus N-stage ≥ 2, their pre-treatment body weight, pre-treatment body mass index (BMI), whether the patient had pre-treatment BMI ≤ 18.5, $18.5 < BMI < 22.9$, $23 < BMI < 24.9$, $BMI > 23$, and $BMI > 25$, were analyzed in the QEH cohort using an unpaired two-sided Mann–Whitney U test. Clinical data with a p-value of less than or equal to 0.01 were selected as the predictive features for the subsequent development of the C model. The finally selected radiomic features and clinical parameters were integrated for building the combined (RC) model.

The predictive performance of the models was evaluated using median area under receiver operating characteristic (AUC) curve using the "ROCR" package. Further, a bootstrap resampling technique with 1000 replications was applied to the entire QEH cohort for obtaining the 95%CI of the AUC estimates of the model. In each sub-sample,

the model was trained using the selected features, resulting in an individual predictability in terms of the AUC. The results over all the 1000 replicates were then reported for the mean "corrected" AUC with a bias-corrected and accelerated 95%CI (BCa 95%CI). Furthermore, the developed models were independently evaluated on an external testing dataset of the QMH cohort for assessing the model generalizability across medical institutions.

With regard to the use of the bootstrapping technique, it employs random sampling with replacement to mimic the sampling process, and it is a statistical procedure that resamples a single dataset for generation of multiple simulated samples for efficient and accurate cross-validation of a prediction model [102]. Unlike traditional methods, bootstrapping uses samples to draw inferences about populations without assumption of any parametric form for the distribution of a population (dataset). Consequently, bootstrapping is commonly applied for a wider variety of distributions, unknown distributions, and smaller sample sizes. The core idea of this technique is to resample with replacement from a single dataset multiple times (i.e., 1,000 as in this study), generating 1,000 resampled datasets, each with individual statistical estimate of AUC. Therefore, not one but all possible configurations were studied on bootstrap datasets, with the model performance being estimated as average loss on the out-of-sample predictions. In this dissertation project, the model was first developed and determined under nested cross-validation using the QEH dataset, the bootstrapping technique was then applied to both QEH and QMH cohorts to generate multiple bootstrap samples from the two respective cohorts for estimating overall model performance in QEH dataset (where the model was developed) and QMH dataset (where the data was not involved in any of the model development processes). Each bootstrap dataset has its

own set of sample statistics of AUC, such as the mean, median, and standard deviation. Specifically, the AUC was first computed on the full data of QEH and QMH datasets, the AUC was then computed on each bootstrap sample of the two datasets. The average difference between the full data-trained model and the bootstrap-trained model was computed to estimate the bias in the full-data-estimated AUC. The final estimate of AUC was given by the difference in the full-data AUC, named as "corrected" AUC, and the estimated bias, named as "bias-corrected and accelerated 95%CI" or "BCa 95%CI".

With regard to the use of external testing, Ramspek et al. provided a comprehensive overview on the use of external validation in machine learning models [103]. The external dataset are naïve data to the developed model, which has not been involved in any of the model development processes, so they serve as a new data to assess the reliability and generalizability performance of the developed model. As the model performance is generally poorer in new patient population than that in the population where the model was developed, it is essential to assess model generalizability before its clinical implementation. Because the data characteristics between centers may vary to different extent, and may influence the model performance, the model external validation will enable the clinicians and researchers to gain more understanding on how the model performs on their own patient population [103].

3.2.6 Statistical Analysis

The discriminability of the R model, in terms of the median AUC scores across the 20 iterations, was compared against the C and RC models in the training, validation, and testing datasets using Delong's test. After bootstrapping, the "corrected" AUC and its BCa 95%CI were recorded and analyzed between the entire QEH and QMH cohort. On the other hand, a Chi-square test was employed to assess the statistical difference of the categorical patient clinical factors between the QEH and QMH cohorts, while an unpaired two-sided student t-test was applied for the continuous clinical factors. In all the above analyses, unless specified, a p-value of ≤ 0.05 was considered statistically significant.

3.3 Results

3.3.1 Patient Characteristics

A total of 124 and 58 NPC patients from QEH and QMH, respectively, were considered eligible in this study. There were 24 (~20%) and 16 (~28%) patients labelled as 1 in the QEH and QMH cohort, respectively.

Table 1 summarizes the major characteristics of the patients in each studied cohort. It indicates that there was no statistically significant difference in the distribution of age, gender, histologic subtype, and volume of GTVn between the QEH and QMH cohort ($p > 0.05$), while there were significant differences in the distribution of a tumor's T-/N-stage and the pre-treatment BMI of patients between the two cohorts ($p < 0.05$). However, further analyses of the clinical association of these features showed that neither the T-stage nor the N-stage of tumors were significantly associated with IfTMs in the QEH ($p = 0.163$ and $p = 0.215$, respectively) and QMH ($p = 0.576$, $p = 0.443$, respectively) cohort; pre-treatment BMI was found to be statistically significant in the QMH cohort ($p = 0.014$), but not in the QEH cohort ($p = 0.600$).

Table 1. Distribution of patient characteristics in both QEH and QMH cohorts.

Patient Characteristics	QEH Cohort		QMH Cohort		*p*-Value
Age (average, range)	54.3 (27–81)		50.8 (32–81)		0.0591
Gender					0.329
Male (no.,%)	93	75	48	83	
Female (no.,%)	31	25	10	17	

63

WHO histologic subtype *					0.705
Type-1 (no., %)	3	2	1	2	
Type-2 (no., %)	2	2	2	3	
Type-3 (no., %)	119	96	55	95	
Tumor stage (7th AJCC)					
T-stage					<0.05
T1–T2 (no., %)	15	12	30	52	
T3–T4 (no., %)	109	88	28	48	
N-stage					<0.05
N0–1 (no., %)	19	15	23	40	
N2–3 (no., %)	105	85	35	60	
Pre-treatment BMI (average, range)	23.3 (14.3–35.5)		25.6 (17.9–35.8)		<0.05
Volume of GTVn (average, range)	30,025.7 (501.0–330,143.0)		22,808.6 (942.0–95,606.0)		0.233

* WHO histologic subtype of NPC: Type 1, keratinizing squamous cell carcinoma; Type 2, non-keratinizing differentiated carcinoma; Type 3, non-keratinizing undifferentiated carcinoma. Abbreviation: AJCC, American Joint Committee on Cancer.

3.3.2 Model Development

Figure 6 illustrates the change in the AUC of the R model against the number of selected features in the model. The best-performing R model in the QEH internal validation set was determined when the number of selected features reached a value of 4, where the model predictability reached the maximum in the internal validation set. The selected features included LoG-6mm-glszm_Low Gray Level Zone Emphasis (Bin count = 100), LoG-6mm-glszm_Zone Entropy (Bin count = 50), Original_gldm_Large Dependence Low Gray Level Emphasis (Bin count = 300), and LoG-6mm_glcm_Inverse Variance (Bin count = 50). For the C model, the results indicated that only N-stage ≥ 2 and the volume of GTVn were found to be significantly different between patients who experienced IfTM-triggered ART events and those did not in the QEH cohort (both $p < 0.01$). The four selected radiomic features and two clinical parameters were combined to form an ensemble of six features for the combined (RC) models.
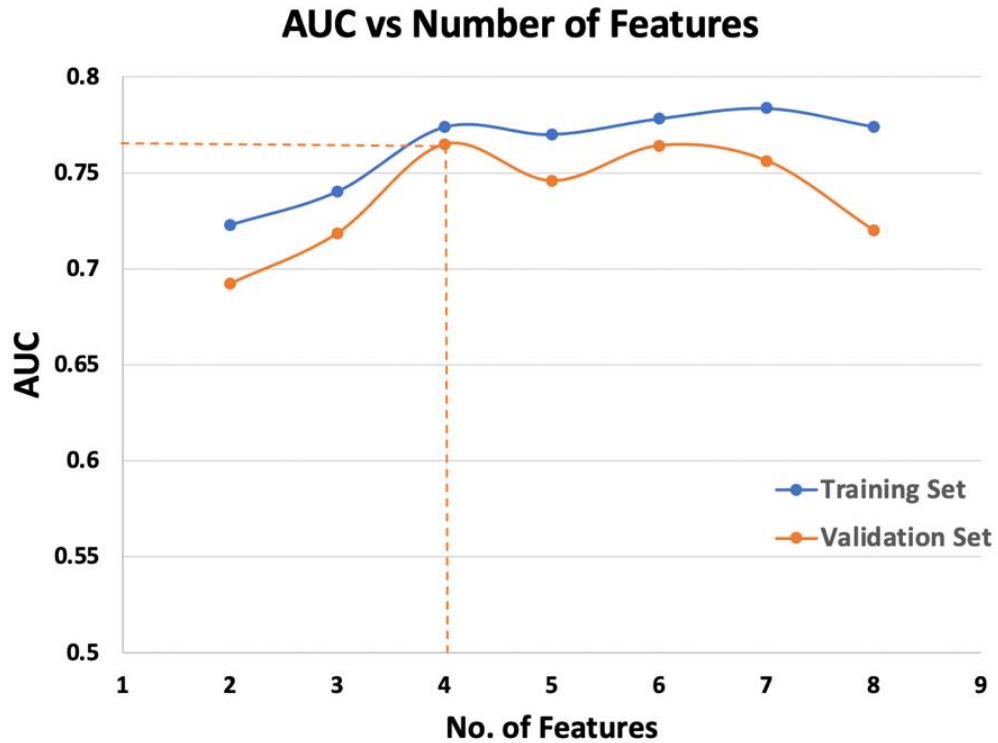
Figure 6. The change in AUC of the R model in both training and internal validation sets of the QEH cohort against the number of selected features.

3.3.3 Model Evaluation

Table 2 summarizes the performance of the different models (R, C, and RC) in the training and internal validation sets of the QEH cohort and the external testing set of the QMH cohort. The bootstrapped AUCs and the corresponding BCa 95%CI of the models in both the QEH and QMH cohorts were also calculated and reported.

From **Table 2**, it can be observed that the R model achieved the highest score of the "corrected" AUC at 0.784 (BCa 95%CI: 0.673, 0.859) in the QEH and 0.723 (BCa 95%CI: 0.534, 0.859) in the QMH cohort following the bootstrapping of 1000 replicates. The C model was the most under-performing model in both cohorts. Similarly, Delong's test showed that the R model was significantly superior in predictability, in terms of the median AUC, over the C model in the training (0.753 vs.

66

0.624, p < 0.001), internal validation (0.716, 0.570, p < 0.01), and external testing (0.637 vs. 0.593, p < 0.001) sets.

Apart from this, the addition of the two selected clinical features into the final R model (i.e., the RC model) did not yield better a predictive performance, with a "corrected" AUC of 0.782 (BCa 95%CI: 0.683, 0.862) in the QEH and 0.710 (BCa 95%CI: 0.474, 0.834) in the QMH cohort. Similarly, the external testing of the RC model demonstrated that there was no statistically significant difference in its predictability as compared to the R model (0.641 vs. 0.637, p = 0.816).

The reported AUC of the testing cohort was generally lower that the "corrected" AUC of the QMH cohort (i.e., the testing cohort) for the three types of models, as indicated in **Table 2**. This is probably due to the fact that the former one was reported as median AUC, while the latter one was reported as average AUC. They are not inter-comparable due to their difference in nature. In particular, a median AUC is required for inter-model comparisons by the Delong's test using the "roc.test" in R software. Nevertheless, the key messages from the **Table 2** are that the radiomic model outperformed the clinical model, and addition of clinical attributes into the radiomic model did not improve the model predictability. Of note, these key messages remain valid in both the "corrected" AUC and the median AUC.

Table 2. A summary of the performance of different studied models (R, C, and RC) in different studied cohorts.

Model	"Corrected" AUC (Average, BCa 95%CI)		Median AUC			p-Values		
	QEH Cohort	QMH Cohort	Training Cohort	Validation Cohort	Testing Cohort			
R	0.784 (0.673, 0.859)	0.723 (0.534, 0.859)	0.753	0.716	0.637	Reference		
C	0.648 (0.516, 0.747)	0.673 (0.499, 0.814)	0.624	0.570	0.593	**** [1]	*** [2]	**** [1]
RC	0.782 (0.683, 0.862)	0.710 (0.474, 0.834)	0.757	0.679	0.641	0.488	**** [1]	0.816

[1] ****: p-value < 0.001; [2] *** $0.001 < p$-value < 0.01.

3.4 Discussion

NPC patients often present the involvement of neck LNs at presentation. Significant neck LN shrinkage is not uncommon in NPC patients undergoing RT, causing a risk of IfTMs during daily RT setup. If this occurs, tremendous efforts are made to implement ad hoc ART to ensure accurate and safe radiation delivery. However, the entire ART procedure for a single patient is highly time-consuming, resource-intensive, and requires multi-disciplinary efforts. Hence, the pre-treatment identification of individual patients who are at a greater risk of having an IfTM is highly desirable to alleviate the clinical workload, facilitate ART practice in the clinic, and achieve personalized RT. For the first time, we attempted to reveal the capability of CT-based neck nodal radiomics in predicting IfTM-triggered ART events in NPC patients via a multi-center setting in this study.

The results of this study showed that radiomics plays a key role in predicting IfTM risk in NPC patients. The R model achieved the highest score of the "corrected" AUC

at 0.784 (BCa 95%CI: 0.673, 0.859) in the QEH and 0.723 (BCa 95%CI: 0.534, 0.859) in the QMH cohort following the bootstrapping of 1000 replicates, achieving a profoundly superior predictability over the C model, which was found to be the most under-performing model in both cohorts **(Table 2).** Moreover, the combined RC model did not result in a better predictive performance than the R model, with the "corrected" AUC of 0.782 (BCa 95%CI: 0.683, 0.862) in the QEH and 0.710 (BCa 95%CI: 0.474, 0.834) in the QMH cohort; it also demonstrated no statistically significant difference in its predictability as compared to the R model in the external testing set (0.641 vs. 0.637, p = 0.816) **(Table 2).** To a degree, the superiority of radiomics may be ascribed to its unique property of unravelling tissue biologic characteristics in response to treatment perturbations. Indeed, an enormous number of articles in the literature have demonstrated the capability of radiomics in predicting tumor responsiveness on the basis of Response Evaluation Criteria in Solid Tumors (RECIST) [5, 75-81], where the criteria are defined according to the degree of tumor volume shrinkage following treatment, which appears to follow the same line of thought as in this study. For example, Piao et al. investigated the potential of MR-based radiomics in differentiating NPC patients who are more likely to get a better treatment response from induction chemotherapy (IC) and those who are not; their radiomic model achieved an outstanding AUC of 0.905 [81]. Similarly, Wang et al. studied a diverse range of MR sequences for the prediction of early therapeutic response to IC in patients with esophageal carcinoma, yielding an AUC of 0.822 in their final model [5]. These studies have laid a great foundation for the radiomics prediction of neck tumor shrinkage leading to IfTM-triggered ART events in NPC patients, as in the present work.

Indeed, multiple research groups have reported a varying extent of neck lymph shrinkage during the course of RT treatment in head-and-neck cancer (HNC) patients, potentially triggering ART implementation. For example, Wing et al. quantified the anatomic changes of 30 NPC patients and reported that there was a significant regression of neck volumes over time with a mean loss rate of $0.39 \pm 0.15\%$/day and a mean volume loss of $11.91 \pm 5.57\%$ upon treatment completion [94]. Murat et al. reported that there was a 43% reduction of the neck nodal target volumes in HNC patients undergoing RT [56]. Similarly, Cheng et al. reviewed both mid-treatment CT and MR scans of NPC patients and found out there were up to 30% reductions in the volume of nodal lesions [50]. All these investigations have suggested that the shrinkage of neck nodal lesions may serve as a favorable criterion for selecting patients for ART. However, there are limited studies on developing an ART eligibility screening strategy based on nodal tumor shrinkage. Yu et al. were the first to demonstrate the capability of MRI-based radiomics from primary tumors in predicting the ART eligibility of NPC patients. The performance of the prediction models in terms of AUC ranged from 0.75 to 0.93 in the testing datasets [70], which appears to be considerably higher than that in this study. To account for this, we inferred that the discrepancy may largely lie in the superiority of MRI in capturing tissue contrast over CT imaging. Given that the nodal lesions of NPC patients were mostly scanned with a CT imaging modality, instead of MRI, in the majority of the data available to us, the development of MR-based radiomic models was not feasible in this work. Nevertheless, investigations on the potential of MR-based nodal radiomics in predicting the ART eligibility of NPC patients can be an interesting area in the future.

Presently, the study conducted by Yu et al. is the only publication in the literature that attempted to predict the ART eligibility of NPC patients though MRI-radiomics from primary tumors (rather than CT-radiomics from metastatic lymph nodes as in this study) [70]. Radiomic features can be broadly divided into three categories: shape, first-order, and texture features. Notably, four texture radiomic features were included in the final R model in this study, including LoG-6mm-glszm-Low Gray Level Zone Emphasis, LoG-6mm-glszm-Zone Entropy, Original_gldm-Large Dependence Low Gray Level Emphasis, and LoG-6mm-glcm_Inverse Variance. On the contrary, the results from Yu et al. indicated that the majority of the selected features in the final radiomic models be-longed to shape or first-order categories, with five out of eight in their contrast-enhanced T1-weighted model, and three out of six in both the T2-weighted and joint T1–T2 models [70]. Whether or not such a difference in the categorical distribution of the selected features between the two studies depends on the type of imaging modality (i.e., CT or MRI) or the source of features (i.e., primary NPC tumor or metastatic lymph nodes) or other factors remains to be fully elucidated; evidence from the body of literature on CT-based radiomics in predicting RECIST-defined treatment response (i.e., tumor shrinkage) may provide us with valuable insights for this.

First, GLCM remains the common feature category in the final CT-based radiomic models, no matter if the source of the features was primary tumors or lymph node lesions [75, 76, 78-80]. Second, multiple studies on CT-based radiomics prediction for RECIST-defined treatment response share similar findings to this study. For instance, Coroller et al. analyzed CT-based radiomics from the primary tumor and lymph nodes of non-small cell lung cancer patients for predicting the pathological response, and

reported that six out of the eight selected features from lymph nodes belonged to textural radiomic features [76]. Trebeschi et al. investigated the potential of CT-based radiomics from primary and lymph node lesions in predicting the cancer immunotherapy response, and found that 7 out of the 10 selected features were texture features [80]. Similarly, Santiago et al. assessed the nodal response of diffuse large B-cell lymphoma to treatment using CT-based radiomics, and reported that 6 out of the 10 selected features were texture features [79]. Moreover, Colen et al. developed a series of CT-based radiomic models for predicting the response to pembrolizumab in patients with advanced rare cancers, and all the selected features belonged to the texture category [75]. Third, the texture features of GLCM-Inverse Variance and GLSZM-Zone Entropy, selected in this current work, were also reported in previous studies on CT-based radiomics for treatment response prediction [75, 78-80]. GLCM-Inverse Variance measures local homogeneity within the tissue volume and GLSZM-Zone Entropy measures texture irregularity or randomness quantified by the amount of homogeneous connected areas within the tissue volume of a certain size and intensity, describing the regional heterogeneity of the tissue [98]. Although intra-tumoral heterogeneity has been regarded to reflect tumor aggressiveness and hence its responsiveness (i.e., shrinkage) to treatment perturbations, the explicit correlation between these features and nodal tumor shrinkage remains unclear and deserves further exploration in the future. Based on findings from the above literature, it can be observed that the dominance of texture features in the R model of this study may be partly ascribed to the use of CT imaging and/or the feature source of lymph node lesions for treatment response prediction. Nevertheless, further investigations in this aspect are warranted.

This study has several limitations. First, the retrospective nature of this study might account for the potential bias; there were significant differences of the tumor T- and N-stage and the pre-treatment BMI of patients between the QEH and QMH cohort (**Table 1**). Nevertheless, these variables were in general not significantly associated with the IfTM-triggered ART events in both cohorts. Therefore, we speculated that its overall impact on the model development should be minimal.

Second, the sample size involved in the model development and evaluation was relatively small, potentially causing model overfitting. To deal with this, a 10-fold cross-validation with 20 iterations and bootstrapping with 1000 replicates were applied in this work with an attempt to minimize the potential prediction bias. Moreover, an independent external test was performed to assess the model generalizability (i.e., degree of model overfitting) between medical centers. Nonetheless, a larger study cohort is warranted in the future to achieve a higher statistical inference.

Third, the predictive performance of the radiomic model was limited by the use of CT images, which may impede its widespread clinical utility. However, this was limited by the dataset available to us, where the nodal lesions of NPC patients were mostly scanned with a CT imaging modality; thus, the development of MR-based radiomic models was not feasible in this work.

Notably, the ART eligibility screening for NPC patients is still in its infant stage; this study is the first of its kind in investigating the potential of CT-based neck nodal radiomics in a multi-institutional setting for predicting IfTM-triggered ART demand in NPC patients. Hence, the findings of this study should provide valuable insights into

developing a more effective screening for ART eligibility in NPC patients in the long run.

To summarize, in this study, we attempted to investigate potential of CT-based neck nodal radiomics in a multi-institutional setting for predicting IfTM-triggered ART demand in NPC patients. The results of this work revealed that CT-based neck nodal radiomics was capable of predicting IfTM-triggered ART events in NPC patients undergoing RT. The R model consisted of four texture radiomic features, achieving a "corrected" AUC of 0.784 in the QEH cohort and 0.723 in the external QMH cohort. The findings of this study provide valuable insights for future study into developing an effective screening strategy for ART eligibility in NPC patients in the long run, ultimately alleviating the workload of clinical practitioners, streamlining ART procedural efficiency in clinics, and achieving personalized RT for NPC patients in the future. Future work on a larger cohort with MR nodal radiomics is highly warranted for strengthening the model predictability and statistical inference.

4. Chapter 4: Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients

Introduction: Adaptive radiotherapy (ART) can compensate for the dosimetric impacts induced by anatomic and geometric variations in patients with nasopharyngeal carcinoma (NPC). Yet, the need for ART can only be assessed during the radiation treatment and the implementation of ART is resource intensive. Therefore, we aimed to determine tumoral biomarkers using pre-treatment MR images for predicting ART eligibility in NPC patients prior to the start of treatment. **Methods:** Seventy patients with biopsy-proven NPC (Stage II-IVB) in 2015 were enrolled into this retrospective study. Pre-treatment contrast-enhanced T1-w (CET1-w), T2-w MR images were processed and filtered using Laplacian of Gaussian (LoG) filter before radiomic features extraction. A total of 479 radiomics features, including the first-order (n = 90), shape (n = 14), and texture features (n = 375), were initially extracted from Gross-Tumor-Volume of primary tumor (GTVnp) using CET1-w, T2-w MR images. Patients were randomly divided into a training set (n = 51) and testing set (n = 19). The least absolute shrinkage and selection operator (LASSO) logistic regression model was applied for radiomic model construction in training set to select the most predictive features to predict patients who were replanned and assessed in the testing set. A double cross-validation approach of 100 resampled iterations with 3-fold nested cross-validation was employed in LASSO during model construction. The predictive performance of each model was evaluated using the area under the receiver operator characteristic (ROC) curve (AUC). **Results:** In the present cohort, 13 of 70 patients (18.6%) underwent ART. Average AUCs in training and testing sets were 0.962

75

(95%CI: 0.961–0.963) and 0.852 (95%CI: 0.847–0.857) with 8 selected features for CET1-w model; 0.895 (95%CI: 0.893–0.896) and 0.750 (95%CI: 0.745–0.755) with 6 selected features for T2-w model; and 0.984 (95%CI: 0.983–0.984) and 0.930 (95%CI: 0.928–0.933) with 6 selected features for joint T1-T2 model, respectively. In general, the joint T1-T2 model outperformed either CET1-w or T2-w model alone. **Conclusions:** Our study successfully showed promising capability of MRI-based radiomics features for pre-treatment identification of ART eligibility in NPC patients.

4.1 Background

Due to the high proximity of the primary NPC tumor to the surrounding critical organs (spinal cord, brainstem, parotid glands) and metastatic neck lymph nodes, NPC is rarely treated surgically; radiation therapy (RT) remains the mainstay of NPC treatment [104]. Intensity-modulated radiation therapy (IMRT) with/without induction chemotherapy (IC) or adjuvant chemotherapy (AC) is currently the standard of care for NPC patients [104]. In clinical practice, RT treatment plans are tailor-made based on anatomic information of individual patients from their pre-treatment planning computed tomography (CT) images to maximize the radiation dose to tumor while protecting nearby critical structures and maintaining sufficiently high dose coverage to surrounding nodal targets.

However, an abundance of research has shown that significant anatomic and geometric variations are not uncommon throughout the course of RT for NPC due to either body weight loss (BW loss) or tumor regression [49-55]. Radiation- induced mucositis is a common and debilitating complication for RT to HNC patients, which can lead to severe pain and difficulty in eating, largely affecting one's nutritional intake

and resulting in significant BW loss. A prospective study reported a 37% of BW loss > 5kg by the end of treatment [59]. Patients having significant BW loss tends to accompany with reduced skin separation at various levels of cervical spine and neck [60], causing positional variability due to possible head movement inside the thermoplastic cast. Consequently, such variations would leave the issue of whether the contour deviations induced significant dose deviations in target or organs at risk. For tumor regression, Hu et al. conducted a retrospective study and reviewed the planning CT and re-CT images of 40 re-planned NPC patients and confirmed the significant clinical-target-volume shrinkage of 35.1% [51]. Murat et al. also reported median percentage change in GTV of HNC patients for primary (26.8%), nodal (43.0%), and total (31.2%) GTVs [56]. Indeed, when significant tumor shrinkage occurs, those critical organs might move into the original high dose region, leading to deleterious dosimetric impact on the surrounding organs [52, 55, 57] and/or insufficient dose delivery to targets [55, 58]. ART can compensate for these dosimetric impact and maintain desirable therapeutic index. The clinical and dosimetric benefits of ART for HNC and NPC cancer patients have been widely reported [69, 105-107]. Yet, the implementation of ART is limited by several reasons. First, the choice to ART can be resource intensive and time-consuming for repeat imaging, re-contouring, re-planning, and analyzing dosimetric impacts between previous and new treatment plans, adding significant clinical burden and cost of patient care to an oncology center. Hence, performing ART on a patient basis is clinically impractical, especially for some busy units. Second, due to the nature of multifactorial ART eligibility, there is no universal selection protocol for ART that can be applied to all hospitals. In this regard, a huge amount of efforts has been constantly made to identify possible ART criteria for HNC

and NPC cancer patients [51, 53, 54, 56, 86, 94, 108, 109] to facilitate the clinical application of ART. Despite that, the current ART practice in most oncology centers, particularly for those busy units, is not efficient. The need for ART of each patient can now be only assessed during the RT treatment. Therefore, pre-treatment identification of high-risk NPC patients for ART is crucially favorable to achieve optimal personalized RT treatment, enabling radiation oncologists to more effectively and accurately prescribe ART for NPC patients and streamline resources management in clinical settings.

Recently, the field of radiomics together with rapid machine learning paradigms have increasingly gained popularity in the community of medical research, paving the way toward precision and personalized medicine [1]. Radiomics, first introduced by Lambin et al. is now shifting the role of medical imaging beyond the traditional diagnostic purposes [1]. It allows for transformation of digitally encrypted medical images into mineable high-dimensional data, which can then be quantitatively analyzed to decode concealed genetic and molecular traits for decision making in oncology [110]. While the predictive powers of radiomics in both cancer diagnosis and disease progression have been widely investigated [4, 71-74], an extremely limited effort has yet been made to identify cancer patients for ART. Given the evidence of significant tumor shrinkage between two CT scans along RT treatment for re-planned NPC patients, we hypothesize that radiomic features extracted from 3-dimensional tumor volume contain predictive biomarkers for tumor shrinkage following cancer treatment—an implication for ART.

Currently, there is no research to include radiomics in predicting ART eligibility for NPC patients and its tumoral predictive biomarkers for ART has not been explored before. The objective of our study was to identify tumoral radiomic features using multi-parametric MR images, which are capable of predicting the ART eligibility for NPC patients. A study flow of current study is shown in **Figure 7**.
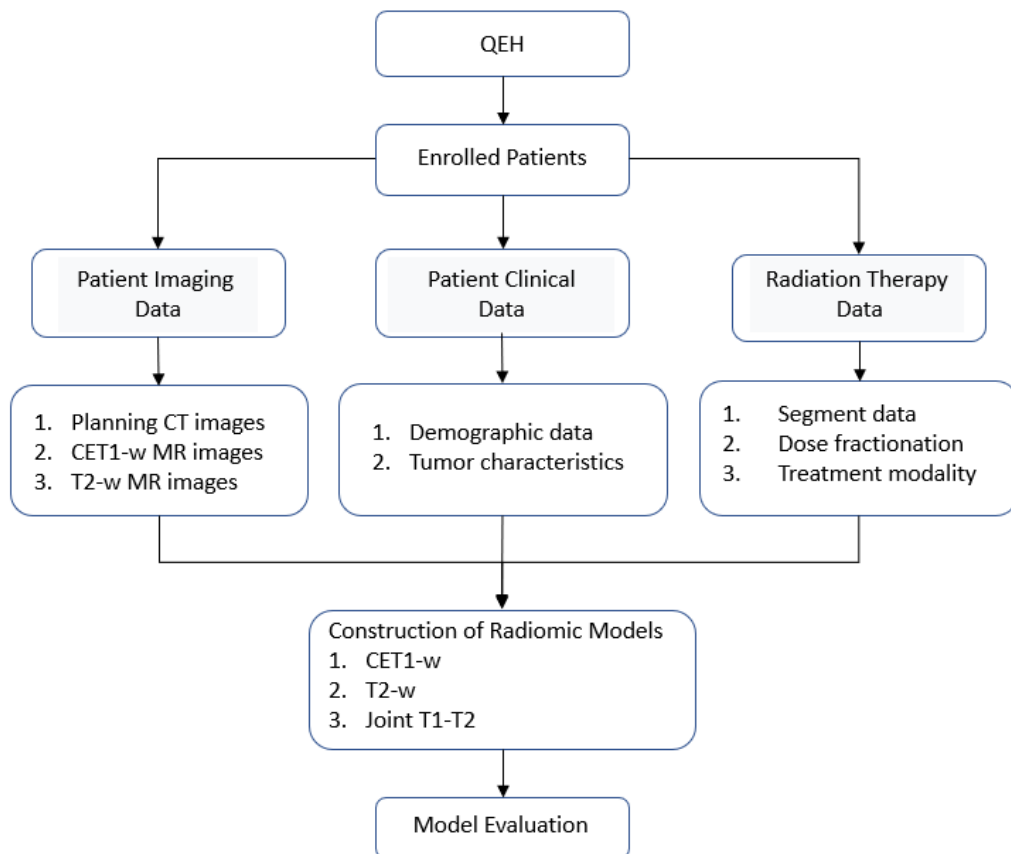
Figure 7. An illustration of study flow of current study.

4.2   Methods

4.2.1   Patient Source

The current research was approved by the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University and Kowloon Central/Kowloon East Cluster Research Ethics Committee of the Hospital Authority. This is a retrospective study, based on analyses of anonymized radiographic data and clinical data, the requirement for individual informed consent was waived. A total of 100 newly diagnosed patients with biopsy- proven (II-IVB) NPC (According to 7th edition of American Joint Committee on Cancer/Union for International Cancer Control TNM staging system) who received primary radiation therapy with/without chemotherapy at the Department of Clinical Oncology of Queen Elizabeth Hospital (QEH) between April 2015 and February 2016 were retrospectively reviewed. Based on the inclusion and exclusion criteria, 70 eligible patients were enrolled in the current study and randomly stratified into training (n = 51) and testing (n = 19) sets, as illustrated in **Figure 8**.
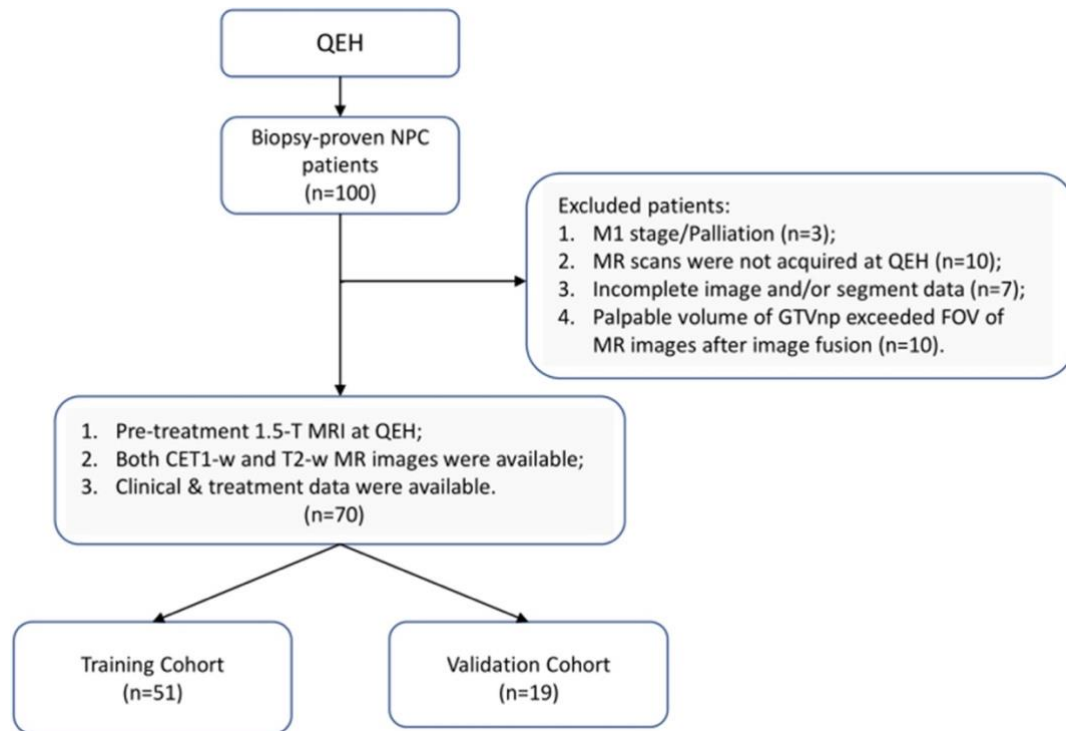
Figure 8. Inclusion and exclusion criteria used in the current study.

4.2.2 Inclusion and Exclusion Criteria

Patients treated at QEH were included in this study if the following inclusion criteria were met: (a) evidence of biopsy-proven NPC; (b) availability of pre-treatment 1.5 Telsa (1.5 T) MRI data acquired following the imaging protocol used at QEH; (c) availability of both Contrast-enhanced T1- weighted (CET1-w) and T2-weighted (T2-w) MR images; (d) availability of clinical (e.g. age, gender, TNM stage, body weight data), treatment and outcome (replan status) data. Patients were excluded if any of the following exclusion criteria were met: (a) any evidence of distant metastasis at diagnosis prior to the initial treatment; (b) palliative treatment was intended; (c) MRI data acquired at hospitals/clinics other than QEH; (d) incomplete image data or segment data; (d) palpable volume of primary tumor (GTVnp) exceeded the field-of-view of MR

images after image registration with planning CT images. Based on these criteria, a total of 70 patients were enrolled in this research. Eligible patients were randomly stratified into training (n = 51) and testing (n = 19) cohorts **(Figure 8).**

4.2.3 Patient Data

Patient clinical data, including demographic information (age, gender) and tumor characteristics (T stage, N stage, histological subtype); imaging data (planning CT images, pre- treatment CET1-w and T2-w MR images); treatment-related data (contouring data, treatment machine, treatment strategies, dose fractionation scheme); outcome data (re-plan status and any replan-related medical records) were retrospectively collected.

4.2.4 Treatment

In general, patients with early-stage (I-II, n = 3) tumors were treated with curative RT alone, while those with advanced- stage (III-IVB, n = 67) were treated with radical concurrent chemoradiotherapy (CCRT), with/without IC or AC. Pre-treatment MRI and planning CT scans were performed within a week prior to the start of IC treatment for target delineation and during the last cycle of IC treatment, respectively. In our dataset, 7 out of 70 patients received IC, while only one underwent ART procedures, who subsequently refused further IC after completion of the first cycle due to repeated vomiting.

The concurrent chemoradiotherapy regimen comprised cisplatin (30 mg/m2 on a weekly basis, 3- 5 cycles). Induction chemotherapy and adjuvant chemotherapy consisted of 3-week cycles of gemcitabine and cisplatin (GP) for 3 cycles and 4-week

82

cycles of cisplatin and fluorouracil (PF4) for 3 cycles, respectively. Reasons for not receiving chemotherapy included age, organ dysfunction indicating intolerance to treatment, and an individual patient's refusal.

Intensity-modulated radiation therapy (IMRT) or TomoTherapy was administrated with a standard dose-fractionation schedule of 66 Gy in 33 fractions in 5 – 6 fractions/week to high-risk clinical target volume (CTV) with 3-mm margins in both nasopharynx (Planning target volume, PTVnp66) and neck region (PTVn66) with fractional dose of 2.18 Gy; Whereas, 60 Gy was prescribed to the low-risk CTV with 3-mm margins in both PTVnp60 and PTVn60 with 1.82 Gy/fraction; A simultaneous integrated boost (SIB) of 70 Gy was given to GTVnp with 3-mm margins (PTVnp70) to achieve optimal local tumor control with 2.12 Gy/fraction.

4.2.5 Clinical Endpoint

The clinical endpoint of this study was defined as the re-plan status of patients: whether or not a patient received ART during RT treatment at the discretion of radiation oncologist.

4.2.6 Multifactorial ART Eligibility

In the dataset, 39 (of 100) patients were initially enrolled into the adaptive review processes, while only 16 ultimately received re-planned procedures. Among the 16 patients, 13 were enrolled in the study, the replans were mostly done during week 4–5 and after the 20th fraction on average A diagram of leading causes for ART implementation are illustrated in **Figure 9**.

83
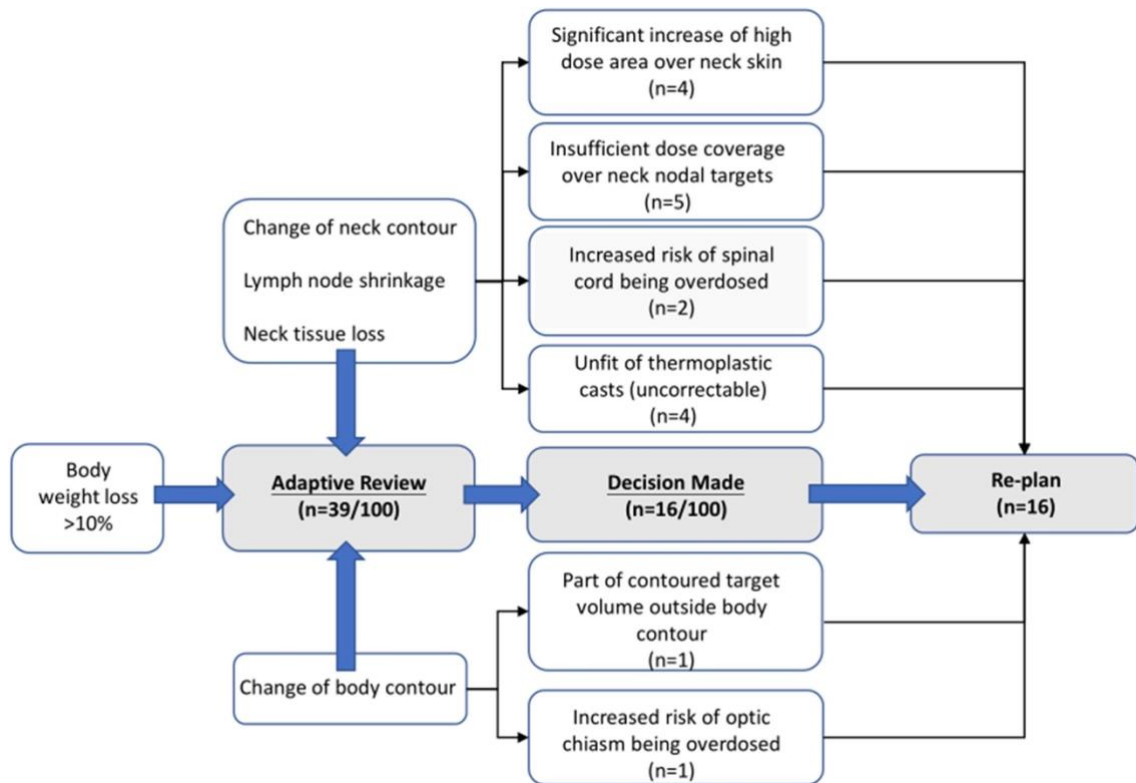
Figure 9. An illustrative example of clinical decision on ART implementation.

Among the 39 patients who were enrolled into a review of need for ART, the majority of patients (29/39, among which 14/29 patients subsequently underwent ART) were reviewed for the need of ART because of significant body weight (BW) loss (>10% drop from the initial body weight measured one day before receiving radiation treatment); other reasons for reviewing the need for ART included significant loss of neck tissue (6/39, among which 4/6 subsequently underwent ART); significant shrinkage of lymph nodes (8/39, among which 4/8 subsequently underwent ART), noted change of body contour or neck contour (5/39, all of which subsequently underwent ART), significant change in neck position/MVCT scan showed twisting of neck (3/39, among which 1/3 subsequently underwent ART); cord displacement larger than 3mm found on 2D X-ray film when Linear Accelerator was used for treatment

(1/39, among which no ART was performed subsequently); teeth removal (1/39, among which no ART was performed subsequently). Some patients exhibited more than one of the above reasons for reviewing the need for ART.

Among the 16 patients who subsequently underwent ART after being reviewed, 4 (out of 16) patients received ART due to unfit of their thermoplastic casts, probably due to significant BW loss and shrinkage of neck lymph nodes; 5 (out of 16) were found to have insufficient dose coverage of their neck node regions, probably due to significant change in neck contour and/or shrinkage of neck lymph node and/or significant loss of neck tissue and/or change of body contour; 4 (out of 16) were found to have significant increase in high dose area over neck skin, probably due to significant decrease in neck lymph node volume/ neck tissue volume; 2 (out of 16) were found to be risky in the dose tolerance of spinal cord (a portion of spinal cord was going to receive >45Gy, the 45Gy isodose line was in close proximity to the spinal cord), probably due to significant change in neck contour; 1 (out of 16) was found to have slightly overdosed optical chiasm, possibly due to change of body contour caused by significant BW loss; 1 (out of 16) received ART because a part of the contoured target volume was found to be outside the body contour. Some patients exhibited more than one of the above reasons for receiving ART.

4.2.7 MRI Acquisition and Segmentation

All 70 patients were scanned with 1.5-T MRI (Avanto, Siemens, Germany) at QEH. We acquired T2-w and CET1-w Digital Imaging and Communications in Medicine (DICOM) images archived using Picture Archiving and Communication System (PACs).

The MR images acquisition parameters were as follows: axial T2-w using short-tau-inversion- recovery (STIR) MR sequence (repetition time [TR]/ echo time [TE]: 7640/97 ms, field-of-view [FOV] = 24 x 24 cm, number of acquisition = 1, slice thickness = 4 mm x 25 slices, spacing: 0.75mm x 0.75mm x 4.4mm, matrix: 320) and axial CET1-weighted spin-echo MR sequence (repetition time [TR]/ echo time [TE]: 739/17 ms, field-of-view [FOV] = 24 x 24 cm, number of acquisition = 1, slice thickness = 3 mm x 48 slices, spacing: 0.938mm x 0.938mm x 3.3mm, matrix: 256).

Intravenous contrast enhanced computed tomography (CT) simulation was performed at 3 mm intervals from the vertex to 5 cm below the sternoclavicular notch with a 16-slice Brilliance Big Bore CT (Philips Medical Systems, Cleveland, OH). All segmentations (tumor, nodal volume and other organs-at-risk) were manually delineated on axial CT slices by an experienced radiation oncologist (with >20 years of experience), which was then fused with MR images for further processing.

4.2.8 MRI Image Pre-processing

Before extracting radiomic features, all MR images were processed using 3DSlicer (version 4.11.0). Isotropic resampling was performed by linear interpolation to obtain a voxel size of $1 \times 1 \times 1$ mm to account for variations in scanning parameters between studied MR series. MRI inhomogeneity correction was applied to account for the locally varying intensity using N4ITK algorithm. To ensure meaning comparison of the extracted features values across all patients, intensity normalization was conducted using brainstem as a reference region-of-interest (ROI), which was chosen because its signal intensity is comparatively homogeneous. The existing contour of the brainstem structure for RT planning purpose was modified to exclude air. Image discretization

with a fixed bin width of 5 to maintain constant intensity resolution across resampled images. Apart from the original images, image reconstructions were performed using Laplacian of Gaussian (LoG) filter with sigma values of 2, 3, 4, 5mm to extract features at multiple scales of resolution, from fine, medium to coarse.

4.2.9   Feature Extraction and Pre-processing

A total of 479 radiomic features were extracted from GTVnp on CET1-w and T2-w MR images, respectively, using SlicerRadiomics in 3D Slicer (version 4.11.0). A representative example of axial pre-treatment MR images with GTVnp contour is shown in **Figure 10**.
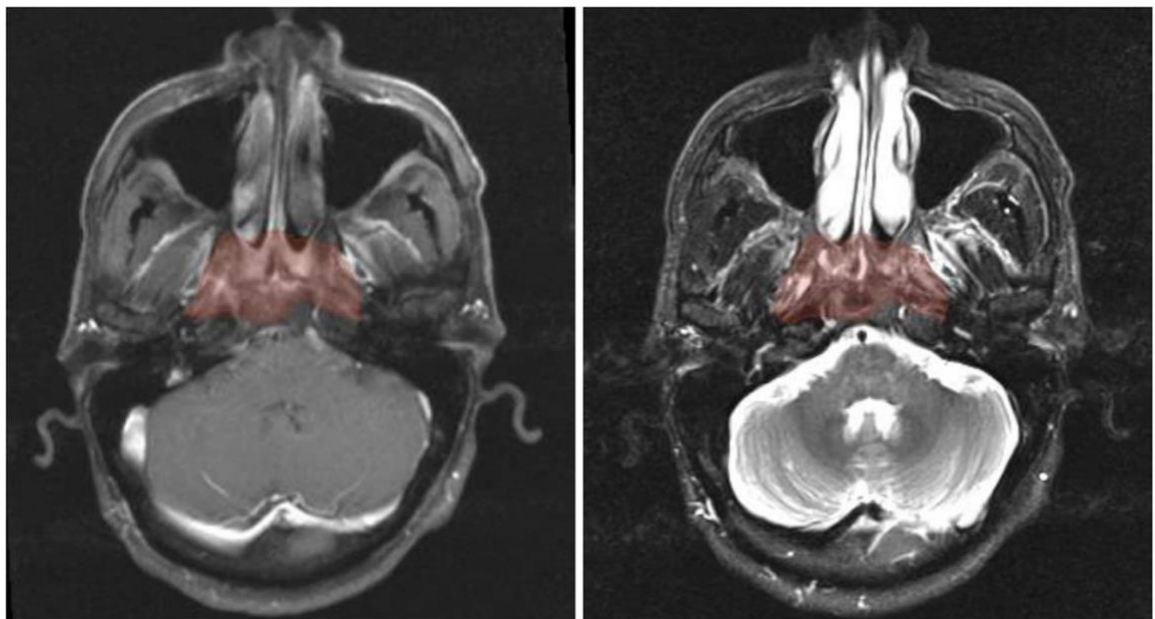
Figure 10. Axial pre-treatment morphological MR images of a 44-year-old man with undifferentiated carcinoma of NPC (T3N2M0). Features of radiomics were extracted from the primary tumor area - GTVnp (red overlay). From left to right: CET1-w and T2-w MR image, respectively.

Extracted features included shape features (n = 14), first-order intensity features (n = 90), and texture features (n = 375). Detailed listing of types of extracted features is indicated in **Figure 11**. All extracted radiomics features were centered and scaled to a value with a mean of 0 and a standard deviation of 1 (z-score transformation) before further analysis using R software (version 3.5.2).

Shape Features:
1.Voxel Volume	8.Major Axis Length
2.Mesh Volume	9.Least Axis Length
3.Surface Area	10.Minor Axis Length
4.Surface Volume Ratio	11.Max. 2D Diameter Column
5.Sphericity	12.Max. 2D Diameter Row
6.Elongation	13.Max. 2D Diameter Slice
7.Flatness	14.Max. 3D Diameter

First-order Features:
1.Skewness	10.Range
2.Kurtosis	11.Mean
3.Entropy	12..Median
4.Uniformity	13. Variance
5.Maximum	14.Interquartile Range
6.Minimum	15.10 Percentile
7.Energy	16.90 Percentile
8.Total Energy	17.Robust Mean Absolute Deviation
9.Root Mean Square	18.Mean Absolute Deviation

GLRLM Features:
1.Gray Level Variance
2.Gray Level Non-Uniformity
3.Gray Level Non-Uniformity Normalized
4.High Gray Level Run Emphasis
5.Low Gray Level Run Emphasis
6.Long Run Emphasis
7.Long Run High Gray Level Emphasis
8.Long Run Low Gray Level Emphasis
9.Short Run Emphasis
10.Short Run High Gray Level Emphasis
11.Short Run Low Gray Level Emphasis
12.Run Variance
13.Run Percentage
14.Run Entropy
15.Run Length Non-Uniformity
16.Run Length Non-Uniformity Normalized

GLCM Features:
1.Joint Average
2.Sum Average
3.Difference Average
4.Joint Entropy
5.Sum Entropy
6.Difference Entropy
7.Inverse Variance
8.Difference Variance
9.Joint Energy
10.Sum Squares
11.Max. Probability
12.Cluster Tendency
13.Cluster Prominence
14.Cluster Shade
15.Contrast
16.Correlation
17.Auto-correlation
18.MCC
19.Id
20.Idm
21.Idn
22.Idmn
23.Imc2
24.Imc1

GLDM Features :
1.Gray Level Variance
2.Gray Level Non-uniformity
3.High Gray Level Emphasis
4.Low Gray Level Emphasis
5.Dependence Entropy
6.Dependence Variance
7.Dependence Non-uniformity
8.Dependence Non-Uniformity Normalized
9.Small Dependence Emphasis
10.Small Dependence High Gray Level Emphasis
11.Small Dependence Low Gray Level Emphasis
12.Large Dependence Emphasis
13.Large Dependence High Gray Level Emphasis
14.Large Dependence Low Gray Level Emphasis

NGTDM Features:
1.Coarseness	3.Contrast	5.Strength
2.Complexity	4.Busyness	

GLSZM Features:
1.Gray Level Variance	9.High Gray Level Zone Emphasis
2.Gray Level Non-Uniformity	10.Low Gray Level Zone Emphasis
3.Gray Level Non-Uniformity Normalized	11.Large Area Emphasis
4.Zone Variance	12.Large Area High Gray Level Emphasis
5.Zone Percentage	13.Large Area Low Gray Level Emphasis
6.Zone Entropy	14.Small Area Emphasis
7.Size Zone Non-Uniformity	15.Small Area High Gray Level Emphasis
8.Size Zone Non-Uniformity Normalized	16.Small Area Low Gray Level Emphasis

LoG-derived Features:
2-mm kernel size category
3-mm kernel size category
4-mm kernel size category
5-mm kernel size category

Figure 11. Types of radiomic features extracted in this study.

The distribution and names of extracted features, including gray level co-occurrence matrices (GLCM), gray level dependence matrices (GLDM), gray level run-length matrices (GLRLM), gray level size zone matrix (GLSZM), and neighboring gray tone difference matrices (NGTDM) categories. Distribution of the different classes of radiomic features is summarized in **Table 3.**

Table 3. Distribution of extracted features in current study.

	Distribution of Extracted Features (n=479)				
	Original Features	LoG-Features (Kernel size: 2mm)	LoG-Features (Kernel size: 3mm)	LoG-Features (Kernel size: 4mm)	LoG-Features (Kernel size: 5mm)
Shape	14	0	0	0	0
First-order	18	18	18	18	18
GLCM	24	24	24	24	24
GLDM	14	14	14	14	14
GLRLM	16	16	16	16	16
GLSZM	16	16	16	16	16
NGTDH	5	5	5	5	5
Sub-Total	**107**	**93**	**93**	**93**	**93**

4.2.10 Feature Selection and Model Optimization Methodology

To avoid over sensitive model, we removed highly inter- correlated radiomics features. By using the R package "caret," we computed Pearson correlation coefficient (PCC) based on a correlation matrix to quantify the pair-wise correlations. If two radiomic features appeared a strong correlation with an absolute correlation coefficient $(r) \geq 0.9$, we removed the feature with the largest mean absolute correlation. As a result, we obtained a primary feature set of 53 from 479.

Following this, we applied Least Absolute Shrinkage and Selection Operator (LASSO) algorithm in R package "glmet" to select the most predictive radiomic features based on the ART status of patients in the training set. The LASSO is typically applied to select high-dimensional biomarkers, and coefficients of the regression variables were penalized in the process of regularization to minimize the prediction error. The ratio of patients who did not receive ART ($n = 57$) to those who did ($n = 13$) was 4, approximately. Considering the imbalance data, we adopted the three-step feature screening strategy, as illustrated in **Figure 12**, to construct CET1-w, T2-w, and joint T1-T2 based radiomic models. The first two steps aimed to further eliminate less/least predictive features in terms of their frequency of occurrence among hundreds of generated models. With the reduced features, we performed PCC with $r \geq 0.8$ to avoid highly correlated features in the final models. Lastly, model trainings were performed with reduced number of input features using a double cross-validation approach, similar to the one adopted by Xu et al. [111]. In short, 100 random sampling was conducted to balance the class distribution within the cross- validation partitions, which would result in a distribution of AUC values across the generated models and

hence allow us to assess the model performance. A 3-fold nested cross-validation was performed with 20 repetition to determine the optimal value for the model tuning parameter (λ). As a result, a total of 2,000 models were generated for each input set of features. In total, 8 sets of radiomic features with number of variables ranging from 3 to 10 were analyzed for the prediction capability in terms of AUCs using box and whisker plots and 95 percent confidence interval (CI).
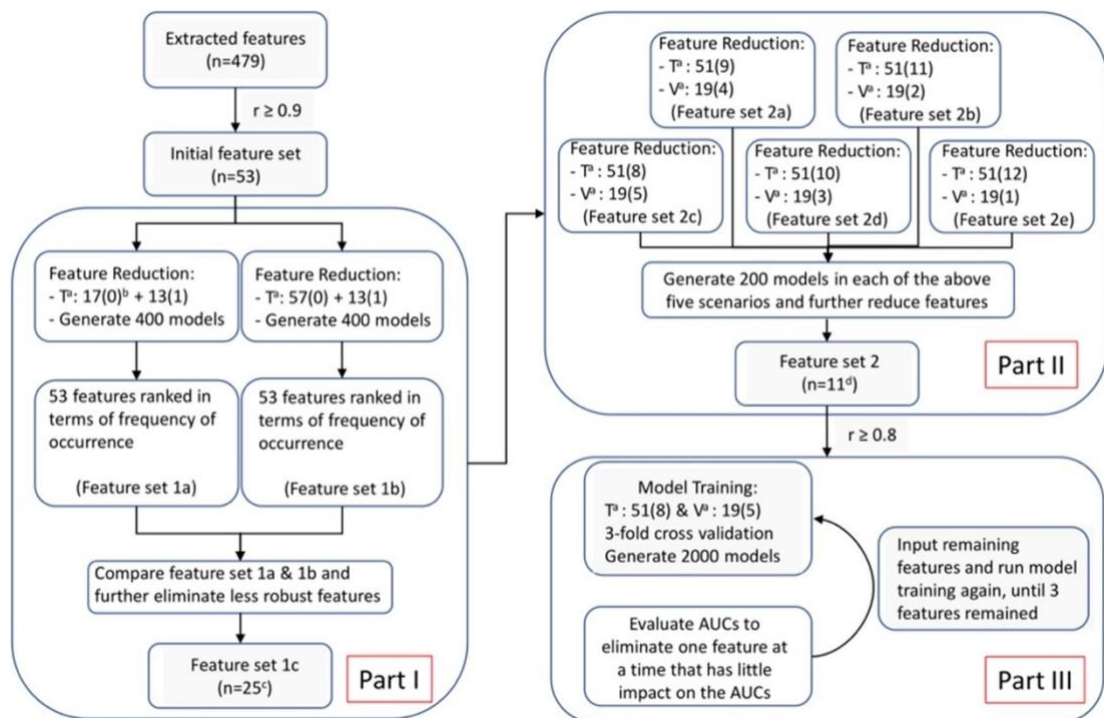
Figure 12. Feature selection and model optimization methodology. Superscript "a": T for training cohort; V for validation cohort. "b": The number inside the parentheses is either "1" or "0," representing "re-planned" and "not re-planned" patients; Numbers in front of the parentheses indicate number of patients. "c": 25 features remained in feature set 1c for CET1-w-based model; while 28 and 39 for T2-w-based and Joint T1-T2-based models, respectively. "d": 16 features remained in feature set 2 for

CET1-w-beased model; while 13 and 22 for T2-w-based and Joint T1-T2-based

models, respectively.

4.2.11 Feature Screening Methodology

As illustrated in **Figure 12**, the feature screening methodology were divided in

three parts. Part I: The goal of part I was solely to eliminate the radiomic features (out

of the 53 features) that were less, if not least, robust. In part I (a), we put all the re-

planned patients (n=13) plus other 17 non-re-planned patients into training cohort

(n=30) with a three-fold cross-validation, and performed multiple rounds of training to

generate a total of 400 sub-models. Then, we ranked the 53 features in a descending

order according to their frequency of occurrence (from a maximum of 400 to a

minimum of zero) to obtain the first set of features. In part I (b), we put all 70 patients

into the training cohort with a ten-fold cross-validation, and generated 400 sub-models

for the same goal. Then, we ranked the 53 features as described to obtain the second set

of features. Following this, we then eliminated the features that did not appear in any

of the 400 sub-models, as they were considered as the least robust to the outcome

prediction. After all these, we compared both sets of features and reduced to 28 features

according to their frequency of occurrence in all the models.

Part II: The goal was to further eliminate the less predictive features by considering

five different possible distributions of re-planned patients (n=13) in training and

validation cohorts. We put 51 patients (including 8 re-planned patients) into training

cohort, and remaining 19 patients (including 5 re-planned patients) into validation

cohort. We employed three-fold cross-validation to generate 200 sub-models, we then

ranked the 28 features in a descending order according to their frequency of occurrence (from a maximum of 200 to a minimum of zero). We repeated the above procedures with other 4 ratios of number of re-planned patients in training cohort to that in validation cohort (i.e. 9:4, 10:3, 11:2, and 12:1). Subsequently, by comparing these 5 sets of features, we further reduced the features number to 16.

Part III: The goal was to create an optimized radiomic model with reduced number of remaining features. We put 51 patients (including 8 re-planned patients) into training cohort, and remaining 19 patients (including 5 re-planned patients) into validation cohort with three-fold cross-validation. Two thousand sub-models were generated, each of them might contain different amounts of features. We then categorized the 2,000 sub-models according to the numbers of remaining features in these models and evaluated these categories one-by-one. By assessing the consistency and stability of area under the receiver operator characteristic (ROC) curve (AUC) in training cohort among different sub-models in specific category, we further removed features that appeared to be less influential to the AUC values and hence less predictive to the outcome.

4.2.12 Statistical Analysis

The statistical correlations between available clinical data and replan status were assessed using univariate logistic regression. All statistical analyses were performed using R software (version 3.5.2). The following R packages were used: The glmnet package was used for LASSO logistic regression. The caret package was used to perform Pearson correlation study. The ROCR package was employed to perform ROC

analysis. All statistical tests were two-sided, and P-values of <0.05 were considered significant.

4.3 Results

The demographic and tumor characteristics of 70 NPC patients are summarized in **Table 4**. Thirteen (18.6%) patients who underwent ART procedure were included. There is no statistical association between the available clinical data and re-plan incidence.

Table 4. Patient characteristics in the present cohort.

Clinical Factor	Category	Number (Percent / SD)	P-values
Gender	Male	50 (71.4%)	0.256
	Female	20 (28.6%)	
Age in years	<51	42 (60%)	0.386
	51-70	7 (10%)	
	>70	2 (2.9%)	
T stage	T1	2 (2.9%)	0.554
	T2	2 (2.9%)	

	T3	50 (71.4%)	
	T4	16 (22.8%)	
N stage	N1	5 (7.1%)	0.859
	N2	56 (80%)	
	N3	9 (12.9%)	
Overall stage	Stage II	3 (4.3%)	0.535
	Stage III	43 (61.4%)	
	Stage IV	24 (34.3%)	
Histology	Type I	3 (4.3%)	0.827
	Type II	1 (1.4%)	
	Type III	66 (94.3%)	
Treatment	EBRT-alone	14 (20%)	0.841
	CCRT	37 (52.9%)	
	CCRT + AC	11 (15.7%)	

	IC + CCRT	7 (10%)	
	Others	1 (1.4%)	
Initial weight (kg) (average, SD)	Replan Group	61.6 (15.5)	0.929
	Non-replan Group	61.9 (12.2)	
Abbreviations: EBRT, External Beam Radiation Therapy; CCRT, Concurrent Chemotherapy Radiation Treatment; IC, Induction Chemotherapy; AC, Adjuvant chemotherapy; Type 1, Keratinizing squamous cell carcinoma; Type II, Non-keratinizing differentiated carcinoma; Type III, Non-keratinizing undifferentiated carcinoma.			

Figure 13 displays the AUC distributions for each feature set (from 3 to 10 features). Figures 5A–C shows the box and whisker plots of the three types of models (CET1-w, T2-w, and joint T1-T2) for training set; Figures 5D–F are for testing set; Figures 5G–I visualizes the range of 95% CI of AUCs in both training and testing sets for the three types of models. The optimal feature sets for each type of models were determined considering the overall distribution of AUC values and its stability. When adding one more feature to the current feature set made no/less difference to the AUC values, the current feature set was considered as the optimal feature set that would give optimal predictive performance of the models. Selected features for each model are listed in **Table 5**.
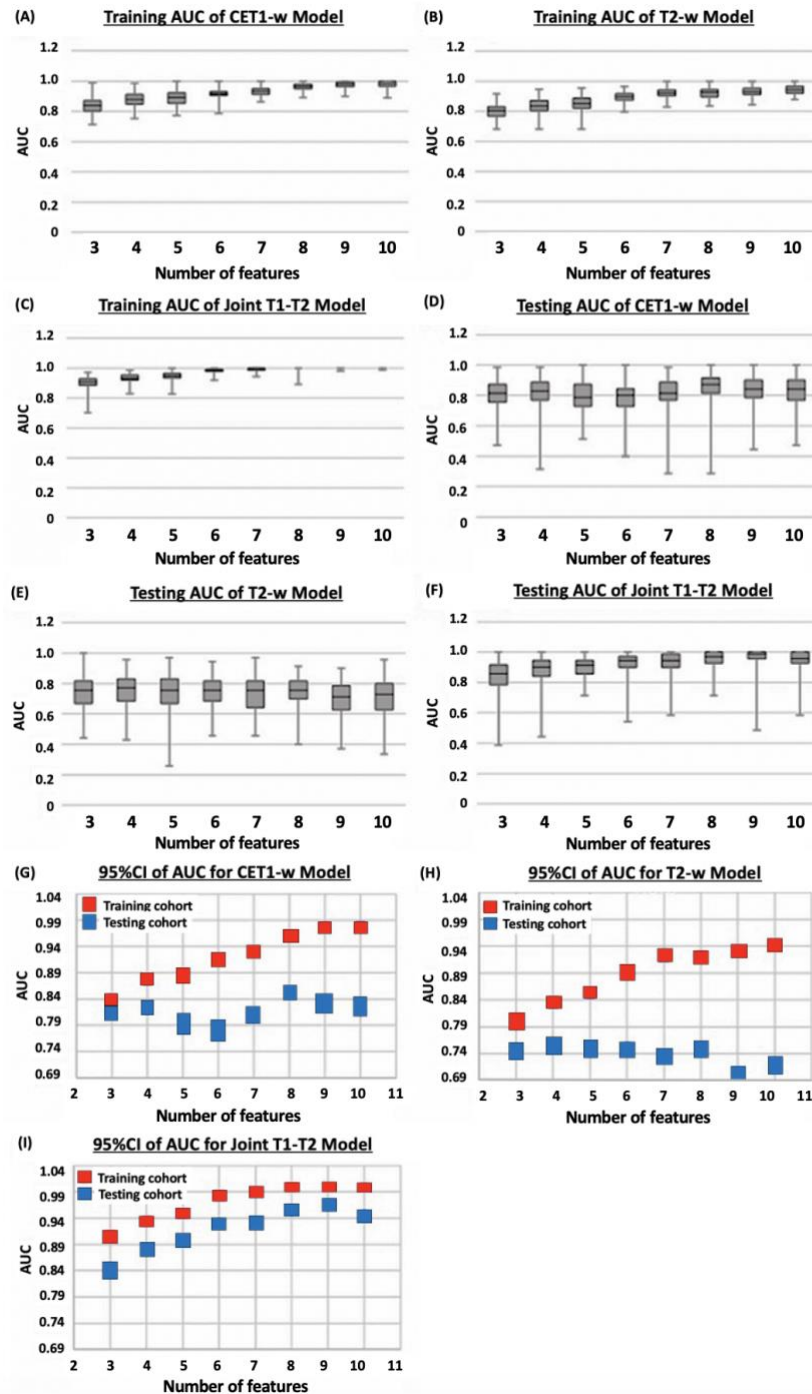
Figure 13. A-F: Distribution of AUC values against number of features in the models (CET1-w model, T2-w model, and Joint T1-T2 model) for training (A-C) and testing (D-F) chorts. G-I: 95% CI of AUC for both cohorts against number of selected features in the models.

Table 5. Table of selected features in CET1-w, T2-w, and joint T1-T2 radiomics models.

MRI Series	Selected Radiomic Features	CET1-w Model	T2-w Model	Joint T1-T2 Model
CET1-w	Original shape sphericity	Yes		
CET1-w	Original shape maximum 2D diameter slice	Yes		
CET1-w	Log-sigma-2-0-mm-3D glcm MCC	Yes		Yes
CET1-w	Log-sigma-2-0-mm-3D first-order Kurtosis	Yes		Yes
CET1-w	Log-sigma-3-0-mm-3D first-order Skewness	Yes		Yes
CET1-w	Log-sigma-4-0-mm-3D first-order Kurtosis	Yes		
CET1-w	Log-sigma-5-0-mm-3D gldm Dependence Entropy	Yes		

CET1-w	Log-sigma-5-0-mm-3D gldm Small Dependence Low Gray Level Emphasis	Yes		Yes
CET1-w	Original first-order Kurtosis			Yes
T2-w	Original shape sphericity		Yes	
T2-w	Original shape elongation		Yes	
T2-w	Log-sigma-2-0-mm-3D gldm Large Dependence High Gray Level Emphasis		Yes	
T2-w	Log-sigma-2-0-mm-3D glcm lmc1		Yes	
T2-w	Log-sigma-3-0-mm-3D ngtdm strength		Yes	
T2-w	Log-sigma-5-0-mm-3D first-order Kurtosis		Yes	
T2-w	Log-sigma-3-0-mm-3D glcm ldn			Yes

Average AUC values in training and testing sets were 0.962 (95%CI: 0.961–0.963) and 0.852 (95%CI: 0.847–0.857) with 8 selected features for CET1-w model; 0.895 (95%CI: 0.893– 0.896) and 0.750 (95%CI: 0.745–0.755) with 6 selected features for T2-w model; and 0.984 (95%CI: 0.983–0.984) and 0.930 (95%CI: 0.928–0.933) with 6 selected features for joint T1-T2 model, respectively.

4.4 Discussion

We successfully revealed the predictive capability of MRI-based radiomics in ART eligibility using the dataset. Eight features were identified for CET1-w model, including 2 shape features (sphericity, maximum 2D diameter slice) and 6 LoG-based features which include 3 first-order features (kurtosis, skewness) and 3 texture features (GLCM and GLDM). Six features were selected for T2-w model, including 2 shape features (sphericity, elongation) and 4 LoG-based features which include 1 first-order feature (kurtosis) and 3 texture features (GLDM, NGTDM). Six features were chosen for joint T1-T2 model, including 1 first-order feature (kurtosis) and 5 LoG-based features which consist of 2 first-order features (kurtosis, skewness) and 3 texture features (GLCM, GLDM), as shown in **Table 5**. With these selected features, we achieved average AUCs of 0.962 (0.852), 0.895 (0.750), 0.904 (0.930) in training (testing) set for CET1- w, Tw-2 and joint T1-T2 models, respectively, representing a promising result for pre-treatment prediction of ART eligibility in NPC patients.

Multiple groups have confirmed that significant tumor shrinkage occurs during RT, triggering the need for ART. Hu et al. reviewed the planning CT and re-CT images of 40 re-planned NPC patients and confirmed the significant clinical-target-volume shrinkage of 35.1% [51]. Murat et al. reported median percentage change in GTV of

HNC patients for primary (26.8%), nodal (43.0%), and total (31.2%) GTVs [56]. Lee H et al. confirmed average volume reduction of GTVnp of 45.9 cm3 (pre- RT) to 26.7 cm3 (third week of RT) in 159 NPC patients. All these studies have suggested that tumor shrinkage serves as a favorable ART criterion [53]. However, only a few studies have developed ART selection strategies based on the tumor volume reduction. Murat et al. developed a decision tree for tumor shrinkage for HNC patients, incorporating initial target volumes and other clinical factors; although an accuracy of 88% was reported in predicting the tumor shrinkage in 48 patients, the validity was not tested and some of the clinical factors used may not be available in other clinics, such as tumor growth pattern (endophytic or exophytic), hindering the generalizability of the decision tree [56]. Recently, Ramella et al. explored the radiomic capability for ART in lung cancer patients and reported that radiomic features extracted from planning target volume (PTV) of lung cancer on CT images were capable of distinguishing patients between ART and non-ART group with AUC of 0.82, on the ground of tumor shrinkage during treatment [112]. To our best knowledge, this study is the first to include radiomics in predicting ART eligibility for NPC patients and its tumoral predictive biomarkers for ART has not been explored before. The promising results are also in line with the work done by Ramella et al. [112].

From the results, we observed that the joint T1-T2 radiomic model outperformed either CET1-w or T2-w alone model in terms of AUCs in both training and testing sets. From **Figures 13 (G–I)**, it can be observed that the joint T1-T2 model gives a more consistent variation in 95% CI of AUCs against different feature sets in both training and testing sets, suggesting that joint T1-T2 model might be the preferable predictive system among the others. Another interesting observation was that the majority (5 of

6) of the selected features in the joint T1-T2 model were from CET1-w images, suggesting that features from CET1- w images might be more predictive than those from T2-w images. A possible reason could be attributed to the inherent limitation of LASSO; when pairwise correlations exist between predictors, the LASSO picks one correlated predictor and ignores the rest. To account for this, we performed another PCC with r ≥ 0.8 prior to part III in the feature selection methodology **(Figure 12)** to avoid highly correlated features in the final models. Further investigations on the feature selection methodology will be part of future studies.

On the other hand, NPC radiomics studies on MR images have been widely studied, focusing mainly on prediction of prognosis (disease progression) and treatment response to either induction chemotherapy (IC) or chemo-radiotherapy, while prediction of the need for replanning has not been previously reported. Besides, each study developed a unique radiomic signature for the same outcome prediction, which limits the feasibility to directly compare all the resultant features between studies. However, interestingly, categories of resultant features might be different depending on prediction outcomes, which might explain the results to some extent. For prognostic prediction, texture features were obviously dominant in their final radiomic signatures relative to first-order and shape features, and GLCM (Gray-Level Co-occurrence Matrix) was the only shared-feature category between studies. A possible rationale might be that the texture features were considered to reflect intra-tumor heterogeneity by depicting the spatial arrangement of voxels (regularity) and variability of local intensity within tumor, which was acknowledged as a characteristic of malignancy. For prediction of treatment response, while GLCM were still the only common resultant feature category between studies, however, first-order features were dominant in final

radiomics signature. Wang et al. investigated the capability of MRI-based radiomic signatures to predict early response to IC for NPC patients using T1-w, CET1-w, and T2-w MR images [5]. Among the 15 features selected in their joint-T1-CET1-T2-w model, 7 were first-order features, three were GLCM features, and the rest were Gabor and wavelet features. Another radiomic study by Hou et al. exploring feasibility of CECT-based biomarkers to predict therapeutic response of esophageal carcinoma to chemo-radiotherapy reported that first-order features (skewness and/or kurtosis) were identified as significant parameters for differentiating SDs (stable disease) from PRs (partial response) and SDs from CRs (complete response) [77]. In both studies, the tumor response was assessed according to the Response Evaluation Criteria in Solid Tumors (RECIST), which takes into account the reduction of tumor size following treatment. Similar to this study, we hypothesized that the image-based tumoral biomarkers are predictive to tumor shrinkage.

From the results, shape features (e.g., Sphericity, Elongation, Maximum 2D diameter slice) and/or first-order features (e.g., kurtosis and skewness) were generally dominant relative to texture features in the models, which is consistent with results from abovementioned radiomic studies for treatment response prediction. Interestingly, it is also worth noting that tumor sphericity was found to be predictive in both CET1-w and T2-w models in this work. Tumor sphericity reflects tumor compactness, with smaller value of sphericity representing a less compact tumor, implicating a higher infiltrative nature of NPC tumor. Similarly, Du et al. investigated prognostic value of pre-treatment MR-based radiomic features for 3-year disease progression following intensity-modulated radiotherapy in NPC patients [113]. They found that the tumor sphericity was the most predictive factor, with lower sphericity indicative of an elevated risk of

disease progression. From the Shapley additive explanations (SHAP) analysis, Du et al. discovered a distinct relationship between tumor sphericity and overall stage of NPC patients, with low sphericity corresponding to lower risk for early-stage tumor (Stage I + II) but higher risk for advanced-stage tumor (Stage III + IV) [113]. It would be of great interest to investigate the potential interaction between tumor sphericity and disease stage in relation to ART prediction in the future. Apart from this, kurtosis and/or skewness and GLCM-based features are the common features shared in all three models. Kurtosis and skewness measure the peakiness and asymmetry of the histogram, respectively, while GLCM features quantify the spatial gray-level variation within local neighbors on a pixel basis. Nevertheless, the understanding of the meaningfulness of these features, especially in relation to the prediction outcome, is still largely unknown and deserves further investigations.

This study has several limitations. Firstly, the heterogeneity of image acquisition and reconstruction protocols and ART strategies in different medical centers limit the generalizability of the identified models and reproducibility of the selected features. In future study, we will perform testing on different datasets obtained from other oncology departments with patients undergoing MRIs on different scanners. Secondly, the rate of adaptive re-planning in the small cohort is relatively low. A more convincible conclusion could be drawn by recruiting larger cohorts with more balanced dataset between patients who underwent replan and those did not, which will be part of future efforts. Lastly, the retrospective nature of this study might account for the potential bias. However, the novelty of this study was to highlight the capability of using pre-treatment MRI radiomic features to predict which patients undergoing radiotherapy for NPC were selected for ART. In future study, radiomics features from other ROIs and

other pertinent non- radiomic clinical data, such as volumetric and dosimetric data of tumor and nearby organs (e.g., lymph nodes and parotid glands), and geometric relations among these structures, will be incorporated into the radiomics models in future to yield a more comprehensive prediction.

In summary, the present study successfully demonstrated promising capability of MRI-based radiomics for pre-treatment identification of ART eligibility in NPC patients. In particular, the joint T1-T2 model with 6 selected radiomic features appears to be the preferable predictive system over other studied models. This would allow radiation oncologists to more effectively and accurately prescribe ART on individual patient basis to achieve true personalized radiotherapy for NPC patients, meanwhile streamlining resources management in clinical settings. In future work, multi-institution prospective studies with larger patient sample are warranted to improve the clinical efficacy of the prediction models.

5. Chapter 5: Multi-organ Omics-based Prediction for Adaptive Radiation Therapy Eligibility in Nasopharyngeal Carcinoma Patients undergoing Concurrent Chemoradiotherapy

Introduction: To investigate the role of different multi-organ omics-based prediction models for pre-treatment prediction of Adaptive Radiotherapy (ART) eligibility in patients with nasopharyngeal carcinoma (NPC). **Methods:** Pre-treatment contrast-enhanced computed tomographic and magnetic resonance images, radiotherapy dose and contour data of 135 NPC patients treated at Hong Kong Queen Elizabeth Hospital were retrospectively analyzed for extraction of multi-omics features, namely Radiomics (R), Morphology (M), Dosiomics (D), and Contouromics (C), from a total of eight organ structures. During model development, patient cohort was divided into a training set and a hold-out test set in a ratio of 7 to 3 via 20 iterations. Four single-omics models (R, M, D, C) and four multi-omics models (RD, RC, RM, RMDC) were developed on the training data using Ridge and Multi-Kernel Learning (MKL) algorithm, respectively, under 10-fold cross validation, and evaluated on hold-out test data using average area under the receiver-operator-characteristics curve (AUC). The best-performing single-omics model was first determined by comparing the AUC distribution across the 20 iterations among the four single-omics models using two-sided student t-test, which was then retrained using MKL algorithm for a fair comparison with the four multi-omics models. **Results:** The R model significantly outperformed all other three single-omics models (all p-value<0.0001), achieving an average AUC of 0.942 (95%CI: 0.938-0.946) and 0.918 (95%CI: 0.903-0.933) in training and hold-out test set, respectively. When trained with MKL, the R model (R_MKL) yielded an increased AUC of 0.984 (95%CI: 0.981-0.988) and 0.927

(95%CI: 0.905-0.948) in training and hold-out test set respectively, while demonstrating no significant difference as compared to all studied multi-omics models in the hold-out test sets. Intriguingly, Radiomic features accounted for the majority of the final selected features, ranging from 64% to 94%, in all the studied multi-omics models. **Conclusions:** Among all the studied models, the Radiomic model was found to play a dominant role for ART eligibility in NPC patients, and Radiomic features accounted for the largest proportion of features in all the multi-omics models.

5.1 Background

Nasopharyngeal carcinoma (NPC) presents immediate proximity to a variety of surrounding critical healthy organs such as spinal cord and brainstem within an intricated nose-pharynx ministry, dysfunction of which can incur catastrophic complications. At present, concurrent chemo-radiotherapy (CCRT) is a standard-of-care remedy for advanced NPC patients; adoption of Intensity-modulated Radiotherapy (IMRT) allows for highly conformal and precise dose delivery to the treatment targets, meanwhile protecting the adjacent healthy tissues. Notably, the success of treatment relies on an assumption that the patient anatomy remains throughout the 6-7 weeks of IMRT course. In response to treatment perturbations, however, tumors and surrounding healthy organs may exhibit significant morphometric volume and/or geometric alterations, which may jointly alter patient anatomy and jeopardize the efficacy of the original treatment plan. The issue of these variabilities can be more detrimental in the IMRT era, where slight anatomic deviations may deleteriously lead to significant dosimetric consequences due to the sharp dose falloff beyond the target lesions. Confronted with this, Adaptive Radiotherapy (ART), a modification of the original

treatment plan, has been introduced to compensate for these patient-specific variations. The dosimetric and clinical benefits of ART for NPC patients have been well-documented in the literature [55, 61, 63-65, 67, 68].

Notwithstanding, ART generally involves re-imaging, re-segmentations of tumor and organs-at-risk (OARs), and re-planning, requiring a highly specialized multidisciplinary team. This labor-intensive and time-consuming nature of ART procedures preclude the feasibility of routine ART practice on a patient basis in clinic. In light of this, tremendous effort has been constantly made to evaluating the underlying morphometric and geometric variations of patient anatomy amid the radiotherapy course, in the hope of streamlining clinical implementation of ART [49-52, 56-58, 60, 62, 70, 108, 114, 115].

Radiation dose has long been regarded as a prime attribute for morphometric volume change of tumors, neck lesions and bilateral parotid glands throughout the treatment course. Bahl et al. prospectively analyzed volumetric alterations in 20 NPC patients between pre-treatment computed tomography (CT) and mid-treatment CT at the 17th fraction. They reported approximately 30% shrinkage of high-risk gross-tumor-volume (GTV), which was accompanied with a significantly increased median dose of 7.2-7.7 Gy to and reduced volume of bilateral parotid glands [62]. Another prospective study by Cheng et al. demonstrated that the anatomic tissue shrinkage was dependent on radiation dose received. They analyzed repeated planning CT and magnetic resonance images (MRI) at 30-Gy and 50-Gy intervals and reported that the shrinkage of both primary NPC tumor and nodal lesions against pre-treatment baselines were higher when 50-Gy was delivered (13% and 29%, respectively) than that when

30-Gy was given (9% and 16%, respectively) and a similar trend was also observed for bilateral parotid glands [50]. Further evidence was also observed by Hu et al. who analyzed 40 re-planned NPC patients and confirmed the significant shrinkage of 35% in clinical-target-volume [51], and by Murat et al. who reported a median reduction of 27% and 43% in primary and nodal GTV, respectively, in 48 re-planned head-and-neck cancer patients [56].

Notably, volumetric shrinkages of these organ structures are often accompanied with geometric shifts of internal structures [58, 60] and/or body contour modification [60, 70], which may in concert contribute to an elevated risk of ill-fitted immobilization cast during daily setup [60, 70], and/or detrimental consequences following treatment (e.g., overdosing to OARs [52, 57, 61], underdosing to targets [58, 61], triggering the demand for ART. In view of this, research community has introduced numerous criteria as ART triggers [56, 58, 108, 114, 115], mainly on dosimetric aspects. Nevertheless, most of these factors require close monitoring throughout the radiotherapy course for each patient, pre-treatment prediction of ART eligibility is greatly demanding. Further, these factors are deficient in capturing inter-patient disparity in intrinsic biologic response of tissue upon receiving treatment perturbation.

Until more recently, emerging Radiomics has opened up opportunities for divulging concealed biologic traits and genetic association of tumor and organ structures [96, 116, 117]. There is mounting evidence in the literature showing the power of Radiomics in predicting treatment response on the ground of volume shrinkage in various cancer diseases [5, 75, 77, 78, 81, 112], which has laid great foundation for Radiomics prediction of ART demand in cancer patients. Ramella et al.

performed radiomic analysis on pre-treatment CT images of replanned non-small cell lung cancer patients and generated a radiomic signature for prediction of tumor shrinkage during chemo-radiotherapy, yielding an Area Under the Receiver Characteristics Curves (AUC) of 0.82 [112]. For the first time, Yu et al. generated several radiomic models for ART eligibility in NPC patients using tumoral radiomic features from multi-parametric pre-treatment MRI, achieving AUCs ranging from 0.75 to 0.93 [70]. It is worth noting that ART eligibility is multifactorial in nature. Joint response of multiple organ structures upon treatment perturbations, treatment aggressiveness, and pre-treatment geometric and morphologic condition of patient anatomy, may all come into play for triggering ART.

Therefore, it is pertinent to investigate the role of these attributes, in the form of -omics features, from multiple relevant organ structures within head-and-neck regions using pre-treatment CT, MRI, contours, and three-dimensional dose map for prediction of ART eligibility in NPC patients, which constituted the main objective of this present study. The success of this study may provide the community with valuable insights into developing ART screening strategies in future, particularly in view of the soaring demand of ART in this vulnerable subgroup of cancer sufferers in the IMRT era.

5.2   Methods

5.2.1 Patient Data

This study is a retrospective analysis of 261 NPC patients who received radiotherapy at Hong Kong Queen Elizabeth Hospital between 2012 and 2015. Patient informed consent was waived due to the retrospective nature of this study. Patients were

included if they (1) were diagnosed with biopsy-proven primary NPC without presence of distant metastasis and co-existing tumors of other types at presentation, (2) underwent curative concurrent chemo-RT (CCRT) or CCRT plus adjuvant chemotherapy (AC), and (3) were treated with Helical Tomotherapy. Patients were excluded if they (1) received induction chemotherapy before CCRT treatment, or (2) received RT-alone without concurrent chemotherapy, or (3) did not receive injection of contrast agent for obtaining planning contrast-enhanced CT (CECT) images or planning contrast-enhanced T1-w (CET1-w) MR images, or (4) did not have complete set of clinical/image data. The binary status of whether or not an individual patient has undergone ART treatment during their main course of RT at the discretion of radiation oncologist was chosen as the clinical endpoint for this study. Patients were labelled as 1 if he/she has received ART treatment, otherwise were labelled as 0.

5.2.2 Image Acquisition

All the enrolled patients underwent pre-treatment planning CECT and MRI scans, which were retrospectively retrieved in Digital Imaging and Communications in Medicine (DICOM) format, archived using Picture Archiving and Communication System (PACs).

For CECT image acquisition, patients were immobilized with thermoplastic cast in a supine position. The scan range covered from the vertex to 5-cm below the sternoclavicular notch and was acquired at 3-mm intervals using a 16-slice Brilliance Big Bore CT (Philips Medical Systems, Cleveland, OH). Details of acquisition parameters: scan mode = Helical, voltage = 120-kVp, X-ray tube current = 264-mA,

exposure = 325-msec, pixel spacing = 1.152x1.152-mm, slice thickness = 3-mm, matrix = 512x512 pixels. Two types of intravenous contrast agents were available: (i) OMNIPAQUE TM 350 mg I/ml and (ii) VISIPAQUE TM 320 mg I/ml; either one of them was prescribed and injected at a rate of 2-ml/sec for 70-ml, followed by scanning after a 30-sec delay.

MRI scans were acquired under a 1.5-Tesla MR scanner (Siemens Avanto, Germany). Details of MRI acquisition parameters: axial T2-weighted (T2-w) using short-tau-inversion- recovery MR sequence (repetition time [TR]/ echo time [TE]: 7640/97-ms, field-of-view [FOV] = 24x24-cm2, number of acquisition = 1, slice thickness = 4-mm x 25 slices, spacing: 0.75x0.75x4.4-mm3, matrix: 320) and axial CET1-w spin-echo MR sequence TR/TE: 739/17-ms, FOV = 24x24-cm2, number of acquisition = 1, slice thickness = 3-mm x 48 slices, spacing: 0.938x0.938x3.3-mm3, matrix: 256).

5.2.3 Volume-of-Interest (VOI) Definition

There were a total of 8 different VOIs of organ structures involved in this study, including gross-tumor-volume of primary NPC tumor (GTVnp) and metastatic lymph nodes (GTVn), ipsi-lateral parotid gland (IpsiPG), contra-lateral parotid gland (ContraPG), brainstem (BS), spinal cord (SC), high-dose and low-dose regions of nodal planning target volume (PTVn_high_dose for the PTVn with the prescribed dose level of 70-Gy, PTVn_low_dose for the PTVn with the prescribed dose level of 60-Gy, respectively). **Figure 14** illustrates location of each of the VOIs involved in this study.
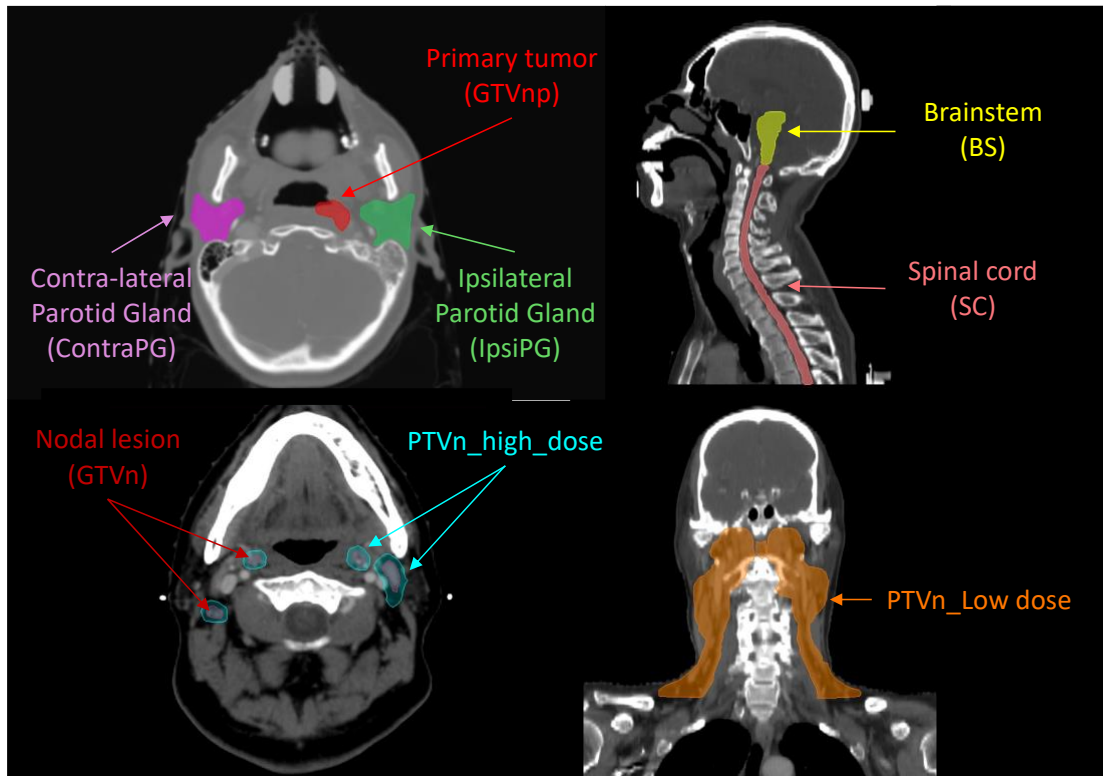
Figure 14. Illustration of the eight VOIs involved.

GTVnp was manually delineated on axial CT slices after registration with planning MR images, and GTVn was delineated on CECT images by an experienced radiation oncologist specializing in head-and-neck cancers with accreditations, in accordance with International Consensus Guidelines for the CT-based delineation of neck levels [118]. To determine whether each of the segmented parotid glands (PG) belongs to IpsiPG or ContraPG for each patient, the minimum geometric distance between a particular voxel point on the PG volume and all voxel points on the GTVnp surface was first determined. This procedure was repeated for another voxel point on the PG volume until the minimum distances between each of all the voxel points on the PG volume and the GTVnp surface were determined. Lastly, a median value of these calculated minimum distances was obtained to determine the overall proximity of that PG to the

GTVnp for each patient.   The PG with smaller median value of the minimum distances was denoted as IpsiPG, otherwise it was denoted as ContraPG. All segmentations were carried out using Varian ARIA and Eclipse treatment planning system v13 (Varian Medical Systems Inc, Palo Alto, CA).

5.2.4   Multi-omics Feature Extraction

Radiomics (R) and Morphologic (M) Features: Prior to radiomic feature extraction, a series of image pre-processing steps were performed on CECT and MR images according to well-recognized recommendations from the Image Biomarker Standardisation Initiative (IBSI) guidelines [98], using the in-house developed Python-based (v3.7.3) platform. Details of the image pre-processing procedures are indicated as follows:

All CECT and MR images were resampled to a voxel size of $1 \times 1 \times 1$-mm^3 to mitigate impacts of difference in image acquisition parameters among different patients. Quantization of grey levels was applied to normalize image signal intensities in both types of images. Grey-level intensities of the images were discretized to a range of fixed bin counts, ranging from 50 to 350 with an incremental interval of 50. Besides, all images were convolved with Laplacian of Gaussian (LoG) filter under three levels of Gaussian radius parameters to produce filtered images which highlight specific texture radiomic features at multiple scales from fine (1 mm), medium (3 mm) to coarse (6 mm).   Wavelet filters (HHH, HLL, LHL, LLH, LHH, HLH, HHL, LLL) were also applied to both CECT and MR images for yielding radiomic features with multiple resolutions. Re-segmentation was performed on CECT images to confine the Hounsfield Unit (HU) to the range of [-150, 180] for eliminating non-soft tissue

114

components within the studied VOI, such as air cavities and bony structures. In particular to MR images, inhomogeneity correction of image pixel value was implemented using N4B bias correction provided in the "N4 Bias Field Correction Image Filter" in SimpleITK (v1.2.4).

In this study, 4 different VOIs of organ structures (GTVnp, GTVn, IpsiPG and ContraPG) were involved in radiomic feature calculations. Extraction of radiomic features was performed using the publicly available SimpleITK (v1.2.4) and PyRadiomics (v2.2.0) packages embedded in the platform in accordance with the IBSI guidelines [98]. Radiomic features can be generally divided into three major families: morphologic features, first-order statistics, and texture features which can be further categorized into Gray Level Difference Matrix (GLDM), Gray Level Cooccurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighboring Gray Tone Difference Matrix (NGTDM) classes. Radiomic feature calculations were performed on CECT, CET1-w and T2-w MR images, with and without being filtered by Laplacian of Gaussian (LoG) filter (kernel size: 1-mm, 3-mm, 6-mm) and wavelet filters (HHH, HLL, LHL, LLH, LHH, HLH, HHL, LLL). In this study, morphologic features of all the 4 VOIs were separated from the radiomic feature set, resulting in a total of 6,348 radiomic features for each studied VOI. The morphologic features of all the studied organ structures were combined to form a set of 56 features.

Dosiomics (D) Features: All the 8 different VOIs of organ structures were employed for dosiomic feature calculation using RT dose data. Conventional dose-volume histogram (DVH) does not contain information on spatial dose distribution

within irradiated organs. By contrast, dosiomics is capable of characterizing spatial pattern of local radiation dose distributions within the 8 studied VOIs. It has been extensively studied in various predictive modelling for cancer prognosis and treatment responses [119, 120]. In this study, dosiomic features of DVH curve points for the 8 VOIs were calculated based on the method adopted by Gabryś et al., examples include but not limited to maximum dose, minimum dose, mean dose, volume of the VOI receiving at least certain dose levels, and minimum dose received by certain volume of the VOI [121]. Besides, spatial dose distribution within each studied VOI was extracted to comprehensively depict the heterogeneity of deposited dose, such as dose gradients along the three imaging axes (x-, y- and z-directions). The definitions of these features were described in a previous publication by Buettner et al. [122]. Further, the three-dimensional (3D) dose distribution within each studied VOI was transformed into a 3D image, such that radiomics-alike dosiomics features were subsequently calculated using the PyRadiomics package; examples include first-order dose statistics, GLDM, GLCM, GLRLM, GLSZM and NGTDM. A total of 1608 dosiomic features were extracted from the 8 VOIs in this study.

Contouromics (C) Features: In this work, we extracted features that depict complex geometric relationships between 4 pairs of VOIs of organ structures (GTVnp and IpsiPG, GTVnp and ContraPG, GTVnp and SC, and PTVn_low_dose and SC), on the ground that the implementation of ART is triggered by change of geometric relationship of different internal organs within head and neck regions. These features were extracted from the RT contour data. For the first time, they were termed as "Contouromics" in this study. For each of the VOI pairs, a series of contouromic features were calculated from a distance descriptor overlap-volume histogram (OVH), as adopted in a previous

publication [123]; for instance, the maximum and minimum distances between SC and PTVn_low_dose during the treatment planning stage were calculated as the distances on the OVH at zero and full volume, respectively. In this study, the calculation of OVH was implemented using the algorithm employed in a previous publication [123]. Besides, an angle descriptor projection-overlap-volume (POV), defined as one VOI that overlaps with the parallel projection of another VOI at specific projection angle, was used for further divulging potential contouromic features from the VOI pairs. A total of 132 contouromic features were extracted from the 4 pairs of VOIs in this study. **Table 6** summarizes the sources of VOIs involved in calculation of the four types of -omics features studied.

Table 6. A summary of the sources of VOIs involved in calculation of the four types

of -omics features studied.

Radiomics (R)	Morphology (M)	Dosiomics (D)	Contouromics (C)
CECT-GTVnp	GTVnp	GTVnp	PTVn_low_dose-SC
CECT-GTVn	GTVn	GTVn	GTVnp-IpsiPG
CECT-IpsiPG	IpsiPG	IpsiPG	GTVnp-ContraPG
CECT-ContraPG	ContraPG	ContraPG	GTVnp-SC
CET1w-GTVnp		BS	
CET1w-IpsiPG		SC	
CET1w-ContraPG		PTVn_high_dose	
T2w-GTVnp		PTVn_low_dose	
T2w-IpsiPG			
T2w-ContraPG			

5.2.5 Determination of Optimal Feature Selection (FS) Algorithms for Each -omics Dataset

Feature dimensionality reduction is considered essential in machine learning when it comes to minimizing the risk of model overfitting. Although there are a multitude of unsupervised and supervised FS algorithms currently available for assessing redundancy and outcome relevance of the studied features, an optimal combination of both kinds of FS algorithms remains unclear. In this study, a total of 6 unsupervised and 4 supervised FS algorithms that have been commonly adopted in machine learning were studied [124] and are publicly available (https://jundongl.github.io/scikit-feature/algorithms.html), giving rise to a resultant amount of 24 FS combinations.

Figure 15 outlines 24 combinations of the 4 supervised and 6 unsupervised FS algorithms studied. Each of the studied FS combinations was applied to training data of each iteration, followed by a 10-fold cross validation (CV) during Ridge modeling. More details of the studied FS algorithms can also be found in the literature [125-129].
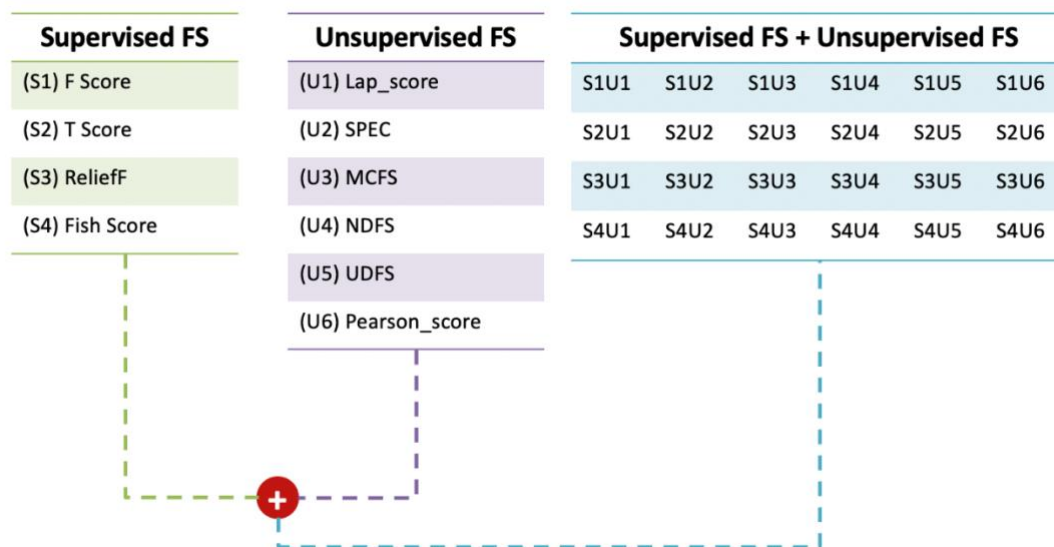
Figure 15. A summary of all the studied supervised and unsupervised FS algorithms and the studied FS combinations. Abbreviations: Lap score: Laplacian score, SPEC: Spectral, MCFS: multi-cluster feature selection, NDFS: Nonnegative discriminative feature selection, UDFS, Unsupervised discriminative feature selection

A proper selection of FS combination for a particular feature set is crucial to ensure that the final selected features of a prediction model are of high discriminability (i.e., high score of Area Under the Receiver Operating Characteristics Curve, AUC score) and high reproducibility under multiple train/test splits of the dataset (i.e., high feature output stability score). To this end, we adopted a strategic workflow (**Figure 16**) to calculate both scores and determined the optimal FS combination using a decision graph (**Figure 16**) for a particular -omics dataset.
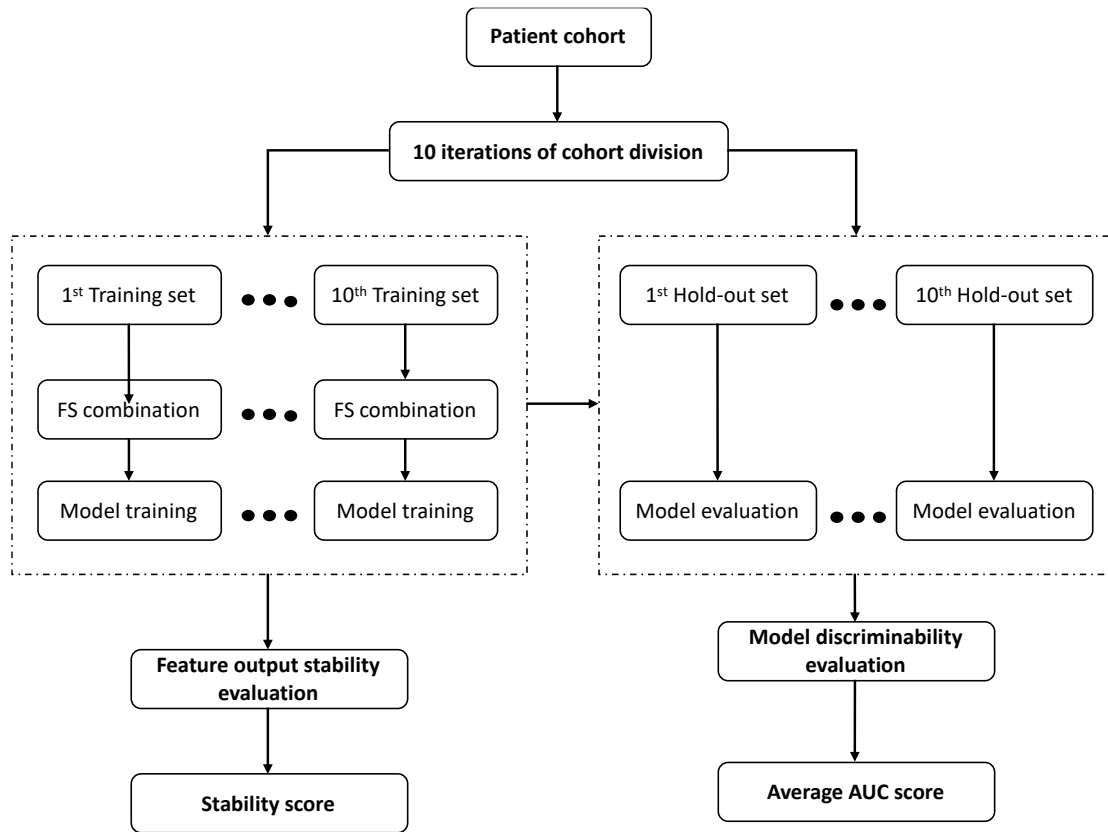
Figure 16. A simplified workflow for optimal FS combination determination of a given feature set.

Figure 16 illustrates a simplified workflow for optimal FS combination determination of a given feature set. The patient cohort was divided into a training and a hold-out dataset with 10 iterations. The hold-out test dataset of each iteration was employed to assess the model discriminability in terms AUC score to avoid introducing bias. Feature output stability of the models was assessed through the 10 iterations and quantified to a value ranging from 0 to 1 using a frequency-based criterion. A value of 1 for stability indicates that the model has completely consistent feature selection results under the 10 iterations. More details on the frequency-based criterion approach are explained as follows:

After implementing the workflow of **Figure 16** with 10 iterations, a binary matrix Z which indicates results of the feature selection of each iteration can be obtained. The matrix Z is defined by **Equation 1**:

$$\mathcal{Z} = \begin{pmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,d} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ z_{M,1} & z_{M,2} & \cdots & z_{M,d} \end{pmatrix}$$

(Eq. 1)

where each of the 10 rows represents results of the feature selection in each iteration. For example, $z_{1,d}=0$ means the dth feature was removed during the 1st iteration. As such, feature output stability of the models was assessed through the 10 iterations and quantified to a value ranging from 0 to 1 using a frequency-based criterion and formulated in **Equation 2 – 5** below:

$$\hat{\Phi}(\mathcal{Z}) = 1 - \frac{\frac{1}{d}\sum_{f=1}^{d} s_f^2}{\mathbb{E}\left[\frac{1}{d}\sum_{f=1}^{d} s_f^2 | H_0\right]} = 1 - \frac{\frac{1}{d}\sum_{f=1}^{d} s_f^2}{\frac{\bar{k}}{d}\left(1 - \frac{\bar{k}}{d}\right)}$$

(Eq. 2)

where

$$s_f^2 = \frac{M}{M-1}\hat{p}_f(1-\hat{p}_f)$$ **(Eq. 3)**

$$\bar{k} = \frac{1}{M}\sum_{i=1}^{M}\sum_{f=1}^{d} z_{i,f}$$ **(Eq. 4)**

$$\hat{p}_f = \frac{1}{M}\sum_{i=1}^{M} z_{i,f}$$ **(Eq. 5)**

As such, scores of feature output stability (i.e., stability score) and model discriminability (i.e., average AUC score on the hold-out test datasets) for each FS combination can be determined for a particular feature dataset. In this study, a product of both scores was calculated and plotted in a decision graph **(Figure 17)** for all the studied FS combinations. The FS combination with the highest value of the score product was determined to be the optimal FS combination for each of the studied feature datasets, which was used in subsequent model development.
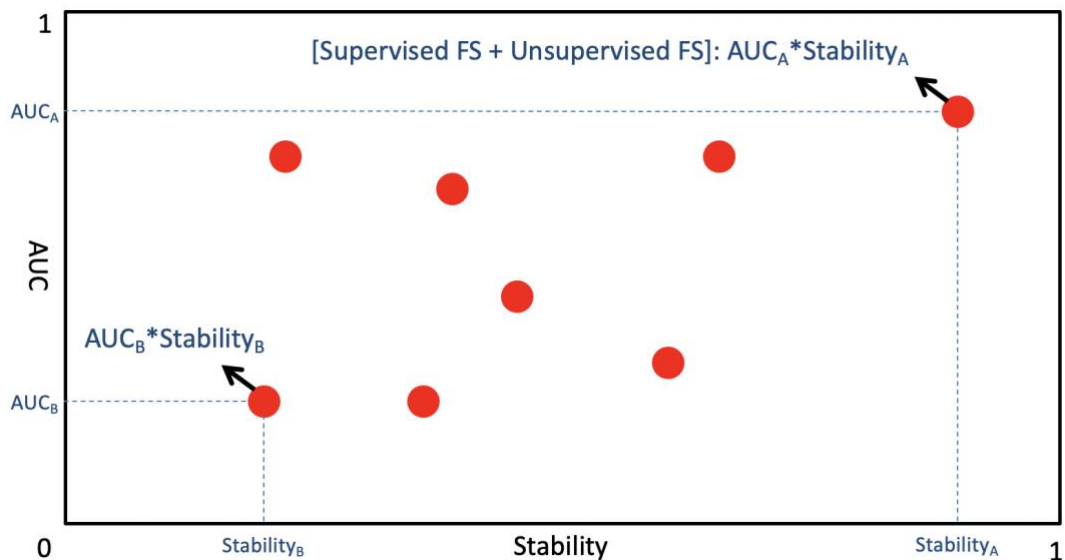

Figure 17. A decision graph showing results of the product of stability score and average AUC score for 8 different FS combinations of a given feature set. Point A has the highest product score of stability and AUC, and hence refers to the optimal FS combination while Point B represents the worst FS combination.

5.2.6 Development and Evaluation of ART Prediction Models

In this study, a total of 4 single-omics models (R, M, D, C) and 4 multi-omics models (RM from R+M, RD from R+D, RC from R+C, RMDC from R+M+D+C) were developed using the corresponding -omics features from multiple VOIs of organ structures.

Figure 18 shows a schematic diagram for model development. The patient cohort was divided into a training dataset and a hold-out test dataset in a ratio of 7 to 3 via 20 iterations. The optimal supervised FS algorithm was applied only to the training dataset of each iteration to maintain clinical relevance of the remnant features. The optimal unsupervised FS algorithm was subsequently applied to remove highly redundant features, leading to a reduced feature set of K features. Development of prediction models was conducted with the initial K features using the Ridge algorithm (for single-omics model) or Multi-Kernel Learning (MKL) algorithm (for multi-omics model) via a 10-fold cross-validation (CV) within the training set to mitigate the risk of model overfitting. Evaluation of model discriminability, in aspects of AUC, was performed on the hold-out test set of each iteration. The model development process was repeated on (K-1) features after removing the feature of the lowest ranking of frequency of occurrence across the 20 iterations until one feature remained in the feature set. An optimal prediction model was finally determined when the average AUC on the hold-out test datasets reached its maximum.
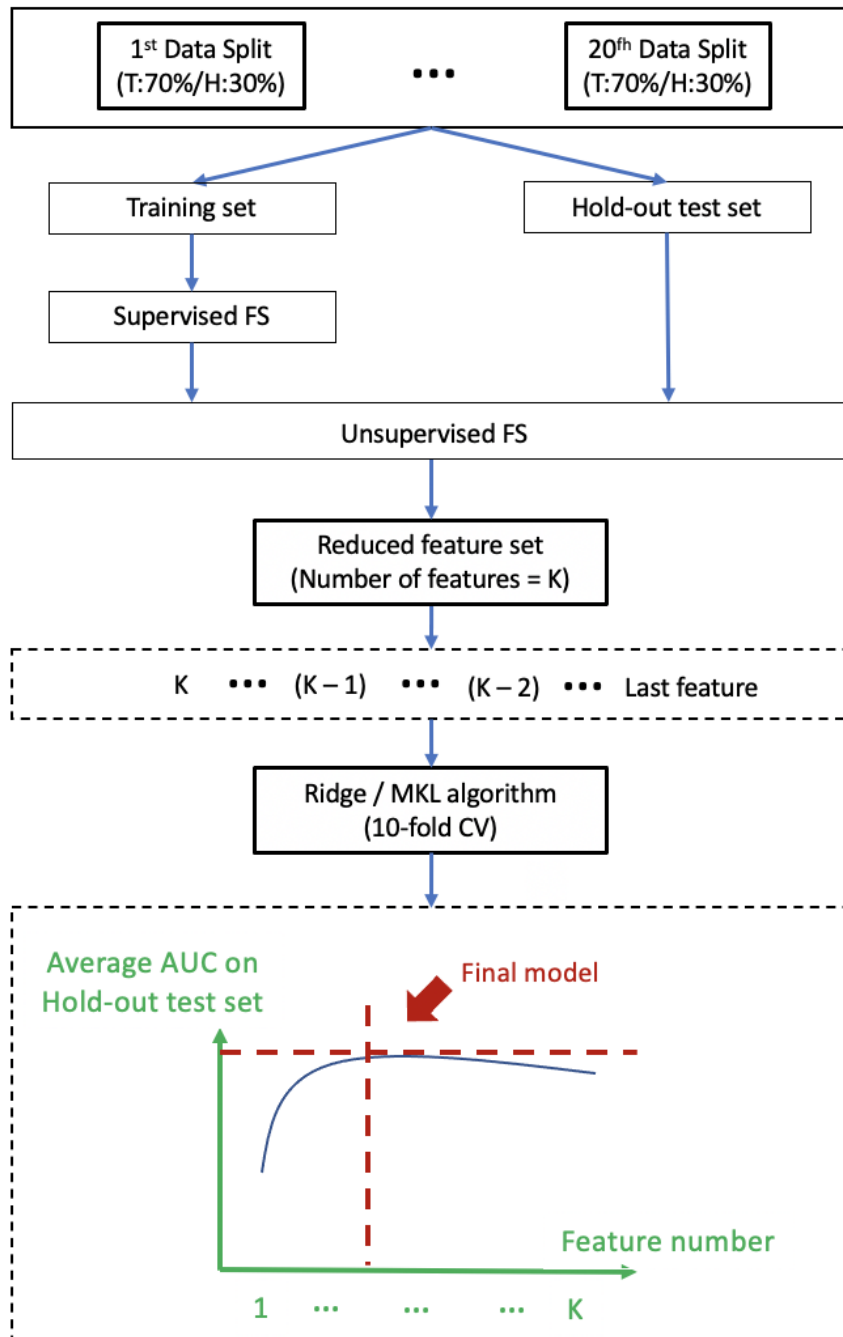
Figure 18. shows a schematic diagram for model development. Abbreviations: T:

Training set, H: Hold-out test set, FS: feature selection, MKL: Multi-Kernel Learning,

CV: Cross-Validation, AUC: Area Under the Receiver Operating Characteristics

Curves.

With regard to the model training algorithm, Ridge classifier was adopted for generation of the 4 single-omics models. It is a typical statistical approach for resolving bias-variance trade-off with the use of a linear function; the principles and advantages of Ridge algorithm have been well-documented [130]. On the other hand, MKL algorithm was applied for development of multi-omics models in this study. Unlike single-omics features, different types of multi-omics data may contain distinctly different data representations. Ridge algorithm is deficient in capturing the difference in representations of multi-omics data and non-linear relationship between predictors and prediction outcome. Therefore, MKL was adopted in this study with an attempt to divulging complementary (non-linear) relationship between different types of -omics features and prediction outcomes. Specifically, two types of kernels (Gaussian and Polynomial) with a range of kernel parameters were applied. Each kernel was embedded into the feature space of a given multi-omics feature set for subsequent multi-omics fusion. **Figure 19** illuminates the multi-omics fusion framework in the study. More details of the MKL algorithms can be a previous publication [131] and is described as follows:

MKL algorithm was applied for development of multi-omics models in this study. Unlike single-omics features, different types of multi-omics data may contain distinctly different data representations. In kernel learning, such data representations are implicitly selected through kernel. The use of kernel allows machine learning practitioners to define similarity between two types of -omics data and a proper regularization term for a learning task. Lanckriet et al. has demonstrated that adoption of multiple kernels improved model interpretability and classification performance

when compared to single kernel application and introduced MKL for binary classification [132-134].

From mathematical perspective, the MKL was adopted by reconstructing the kernel representation in Kernel Ridge Regression to the following **Equation 6**:

$$K = \sum_{i=1}^{P} \alpha_i K_i \quad \textbf{(Eq. 6)}$$

where K_i represents different kernel matrices constructed by different types of kernels, α_i is the kernel weight of K_i, P is the number of kernel matrices. More information on KRR can be found in a previous literature [135].

Figure 19 illuminates the multi-omics fusion framework in the study. Considering a multi-omics feature dataset containing "s" types of -omics data and a total number of "t" features, M1, M2, Ms refer to all the features in the 1st, 2nd and all types of -omics data, respectively; f1, f2, ft refer to the 1st, 2nd, tth individual features of the given multi-omics feature dataset. KM refers to kernel matrix. Gaussian [0.5, 1, 2, 5, 7, 10, 12, 15, 17, 20] corresponds to kernel parameters used in the Gaussian kernel function; Polynomial [1, 2, 3] corresponds to kernel parameters used in the Polynomial kernel function. As such, each of the kernel matrices can either use the combined set of multi-omics features or individual set of single-omics features under specific kernel parameters, achieving multi-omics fusion.
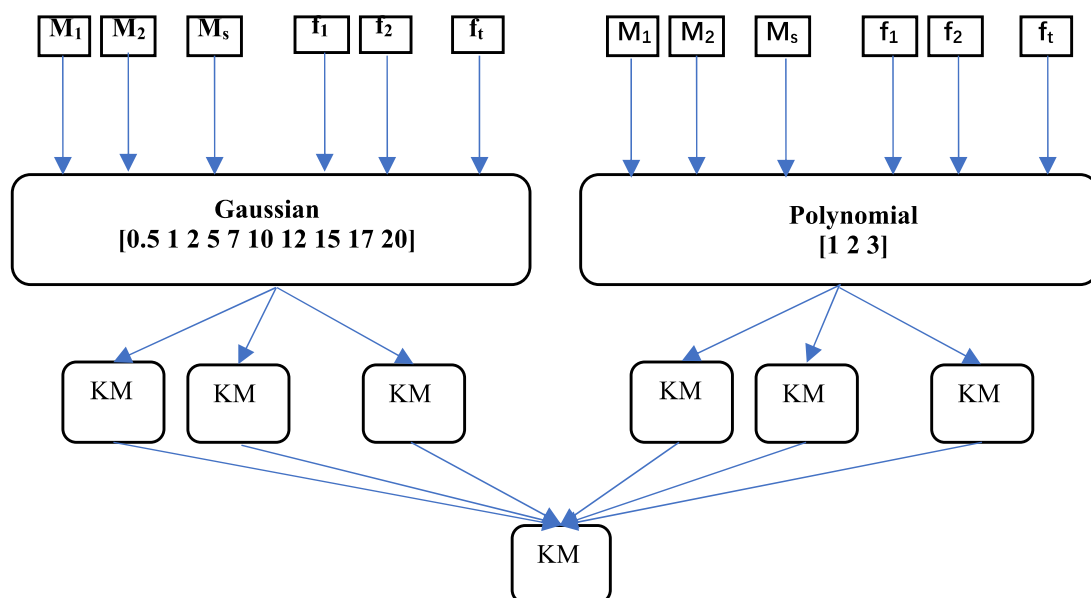
Figure 19. Schematic diagram of the multi-omics fusion framework. Considering a multi-omics feature dataset containing "s" types of -omics data and a total number of "t" features, M1, M2, Ms refer to all the features in the 1st, 2nd and all types of -omics data, respectively; f_1, f_2, f_t refer to the 1st, 2nd, t^{th} individual features of the given multi-omics feature dataset. KM refers to kernel matrix. Gaussian [0.5, 1, 2, 5, 7, 10, 12, 15, 17, 20] corresponds to kernel parameters used in the Gaussian kernel function; Polynomial [1, 2, 3] corresponds to kernel parameters used in the Polynomial kernel function.

5.2.7 Model Comparison and Statistical Analysis

For single-omics models, discriminability of the final radiomic model (R), in terms of distribution of the AUC scores across the 20 iterations, was compared against the other 3 single-omics models (M, D, and C) in both training and hold-out test datasets. For multi-omics models, discriminability of the final RMDC model was compared against the other 3 multi-omics models (RM, RD, and RC) in both training and hold-

out test datasets. Further, we also compared the best-performing single-omics model against all the 4 studied multi-omics models (RM, RD, RC, and RMDC). With this regard, the selected single-omics model was firstly re-trained using MKL algorithm for achieving a fair comparison with multi-omics models.

Statistical estimates of model discriminability in terms of average AUC, its standard deviation (STD) and 95% confident interval (95%CI) across the 20 iterations for all the studied prediction models were reported in this study. Two-sided paired student t-test was employed for the abovementioned comparisons. On the other hand, Chi-square test was employed to assess statistical difference of categorical patient clinical factors between patients who received ART and those who did not, while two-sided student t-test was applied for continuous clinical factors. A p-value of ≤ 0.05 was considered statistical significant.

5.3   Results

5.3.1 Patient Characteristics

A total of 135 NPC patients (35 experienced ART, approximately 26%) were finally considered eligible for this study. **Table 7** summarizes major characteristics of the patients. There were no statistically significant differences in the studied clinical factors between patients who experienced ART and those who did not.

Table 7. Patient clinical characteristics. *WHO histologic subtype of NPC: Type 1: Keratinizing squamous cell carcinoma; Type 2: Non-keratinizing differentiated

carcinoma; Type 3: Non-keratinizing undifferentiated carcinoma. Abbreviation:

AJCC: American Joint Committee on Cancer.

Clinical factor	Data		p-value
Age			
Average, Range	54	27 - 81	0.142
Gender			
Male (no.,%)	101	75	0.348
Female (no.,%)	34	25	
WHO Histologic subtype*			
Type-1 (no., %)	4	3	0.544
Type-2 (no., %)	3	2	
Type-3 (no., %)	128	95	
T-Stage			
T1 (no., %)	9	7	0.133
T2 (no., %)	9	7	
T3 (no., %)	94	70	
T4 (no., %)	23	17	
N-Stage			
N0 (no., %)	1	1	0.146

N1 (no., %)	22	16	
N2 (no., %)	98	73	
N3 (no., %)	14	10	
Overall stage (7th AJCC)			
Stage-I (no., %)	1	1	
Stage-II (no., %)	7	5	
Stage-III (no., %)	92	68	0.077
Stage-IVA (no., %)	23	17	
Stage-IVB (no., %)	12	9	
Initial size of primary tumor (mm^3)			
Average, range	43,482	4,537 - 184,333	0.341
Initial size of nodal lesion (mm^3)			
Average, range	31,078	501 - 330,143	0.202
Initial total tumor burden (primary + nodal lesion) (mm^3)			
Average, range	74,560	7,886 - 438,998	0.153
Pre-treatment body weight (kg)			
Average, range	63	37-102	0.265

5.3.2 Optimal FS Combination Determination and Model Development

Optimal combinations of FS algorithms for the 4 single-omics datasets (R, M, D, C) and the 4 multi-omics datasets (RM, RD, RC, RMDC) were determined using the decision graphs (**Figure 20 – Figure 27**). The optimal FS algorithms and the product scores were determined to be S4U6 and 0.80 for the R model (**Figure 20**), S2U1 or S3U6 and 0.56 for the D model (**Figure 21**), S2U6 and 0.48 for the C model (**Figure 22**), S3U3 and 0.47 for the M model (**Figure 23**), S1U6 and 0.75 for the RD model (**Figure 24**), S4U6 and 0.72 for the RC model (**Figure 25**), S4U6 and 0.76 for the RM model (**Figure 26**), S4U6 and 0.69 for the RMDC model (**Figure 27**), as summarized in **Table 8**.

Figure 20. Results of decision graphs for the Radiomic model, with the corresponding optimal FS algorithms and the product scores being S4U6 and 0.80, respectively (as indicated by Red arrow).
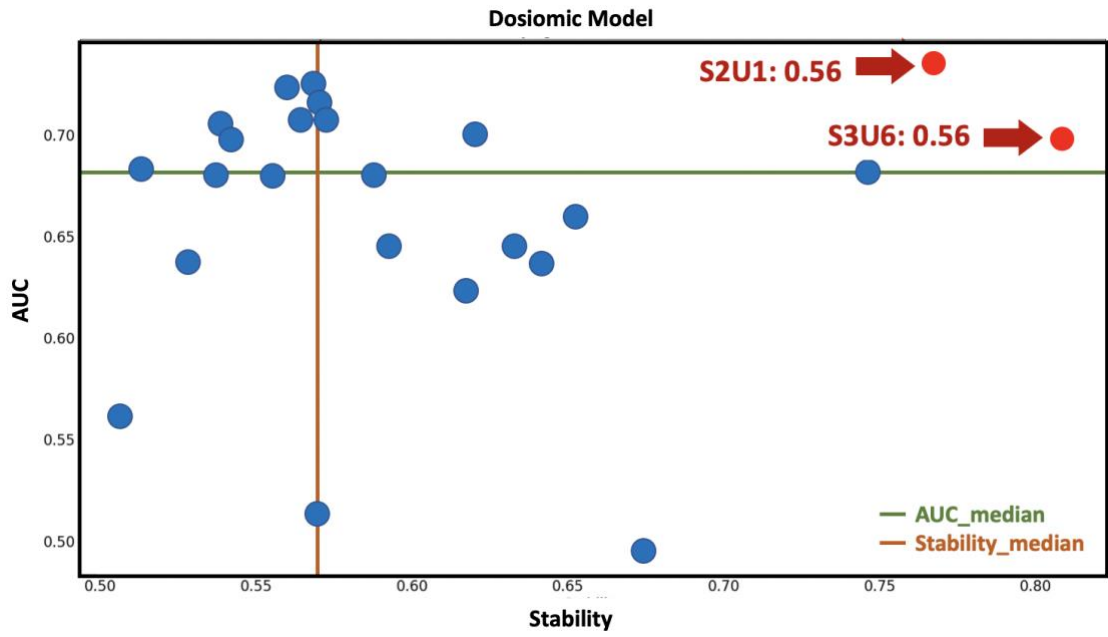
Figure 21. Results of decision graphs for the Dosiomic model, with the corresponding optimal FS algorithms and the product scores being S2U1 or S3U6 and 0.56, respectively (as indicated by Red arrow).
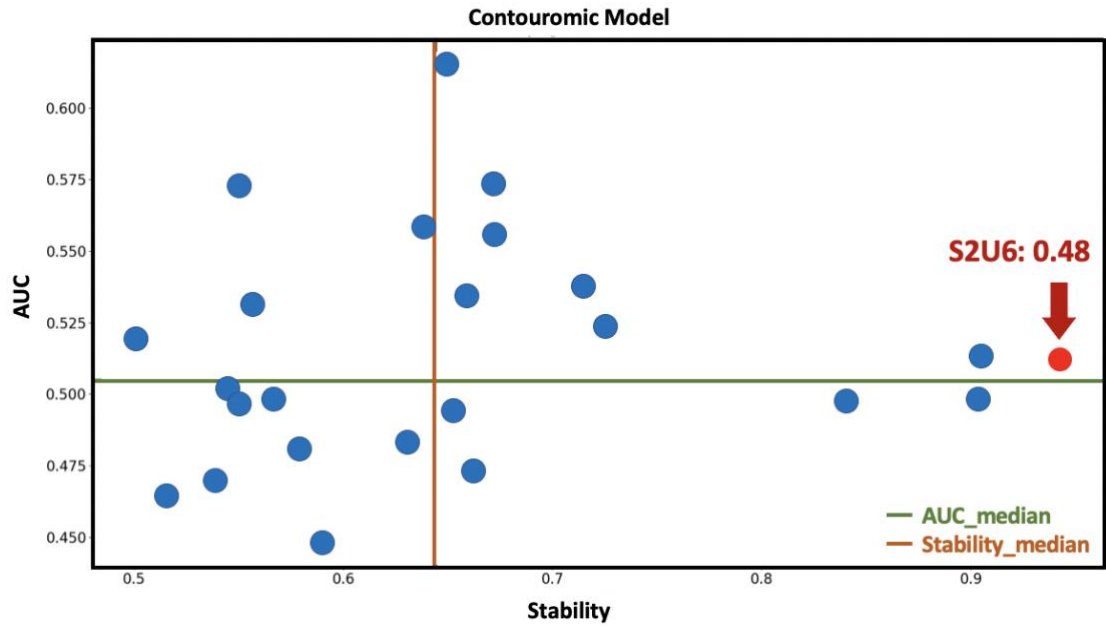
Figure 22. Results of decision graphs for the Contouromic model, with the corresponding optimal FS algorithms and the product scores being S2U6 and 0.48, respectively (as indicated by Red arrow).
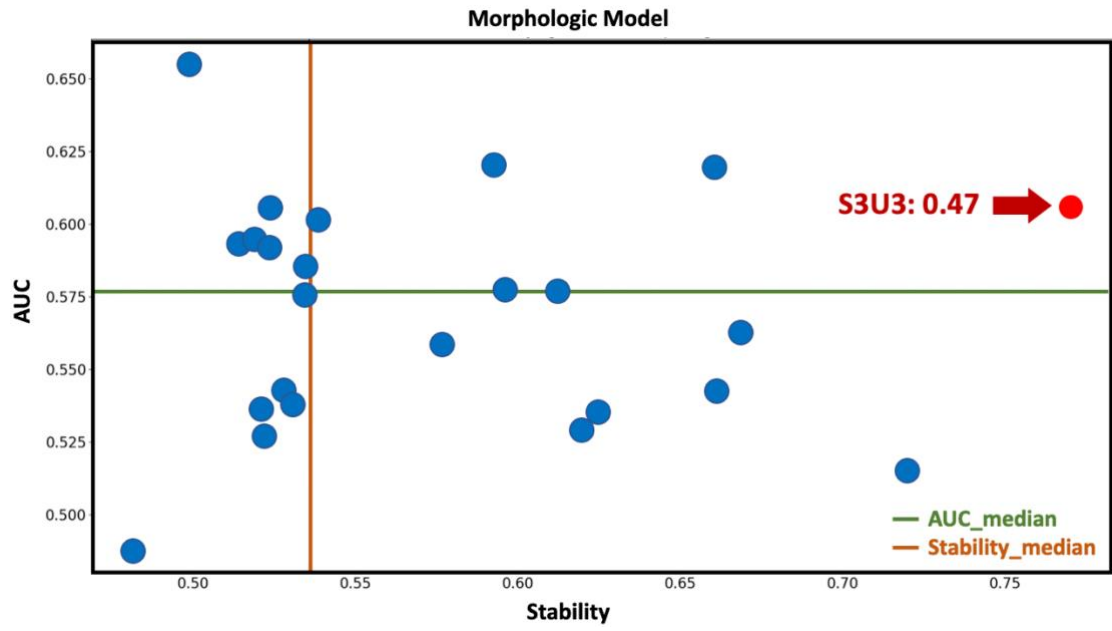
Figure 23. Results of decision graphs for the Morphologic model, with the corresponding optimal FS algorithms and the product scores being S3U3 and 0.47, respectively (as indicated by Red arrow).
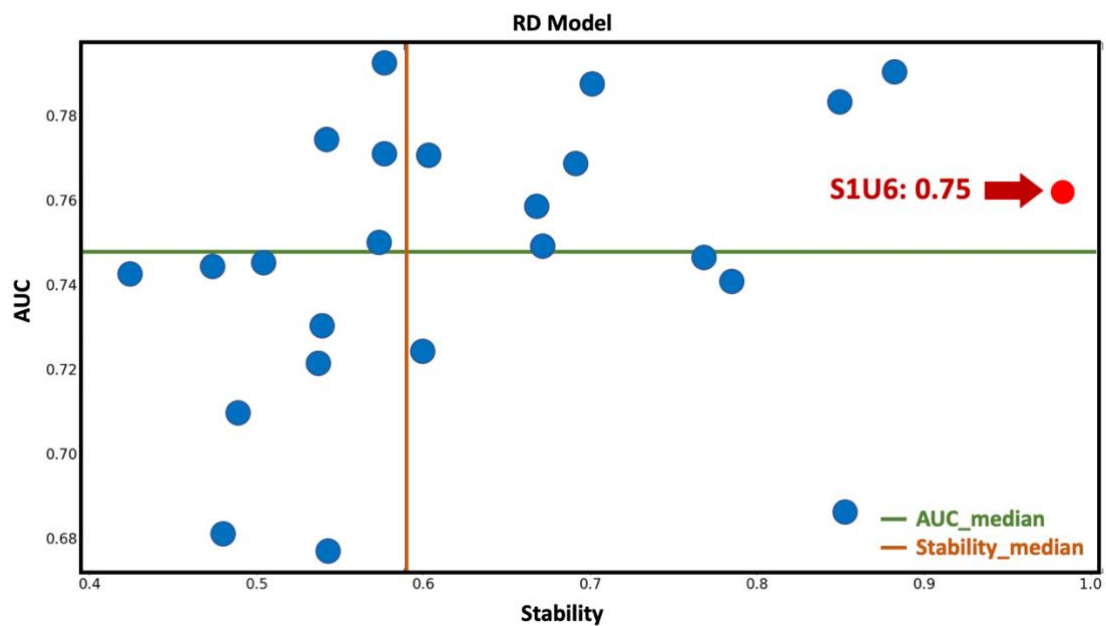
Figure 24. Results of decision graphs for the RD multi-omics model, with the

corresponding optimal FS algorithms and the product scores being S1U6 and 0.75,

respectively (as indicated by Red arrow).
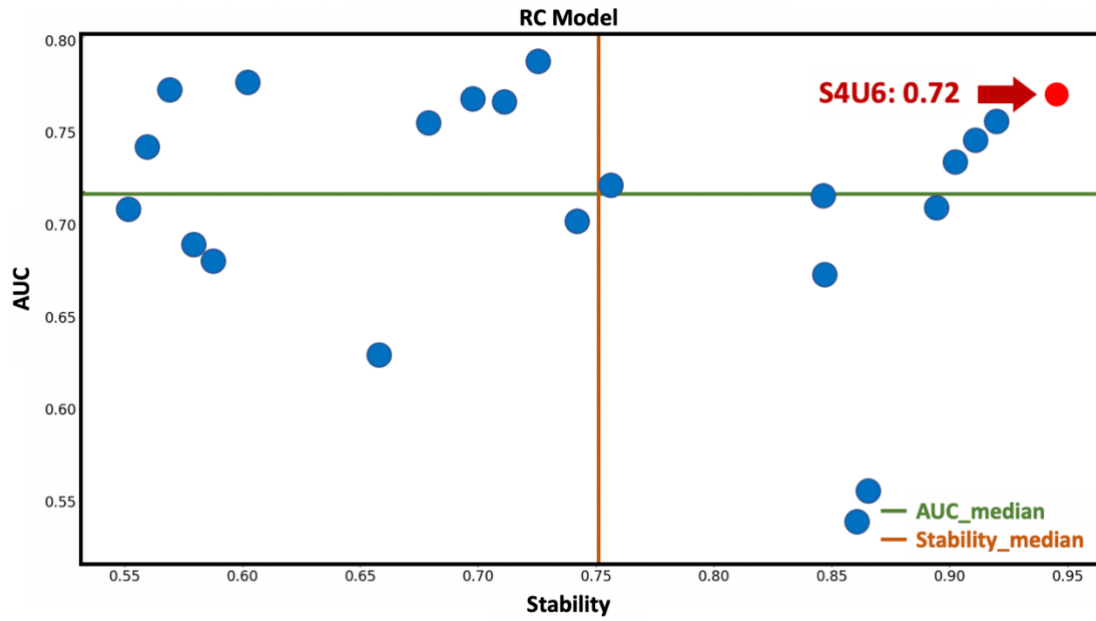
Figure 25. Results of decision graphs for the RC multi-omics model, with the corresponding optimal FS algorithms and the product scores being S4U6 and 0.72, respectively (as indicated by Red arrow).
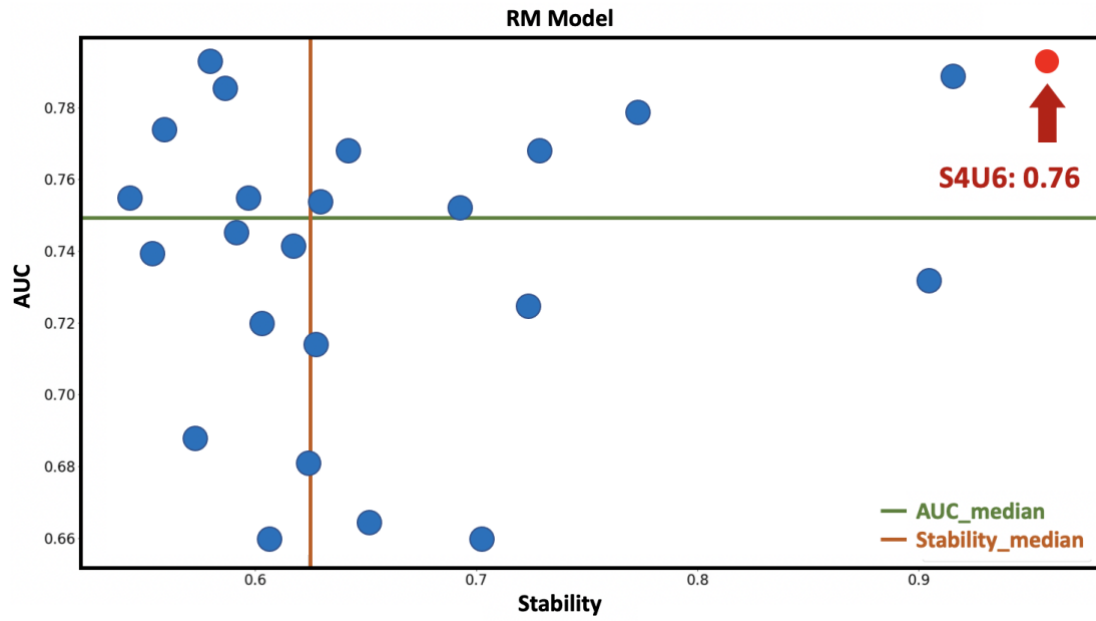
Figure 26. Results of decision graphs for the RM multi-omics model, with the corresponding optimal FS algorithms and the product scores being S4U6 and 0.76, respectively (as indicated by Red arrow).
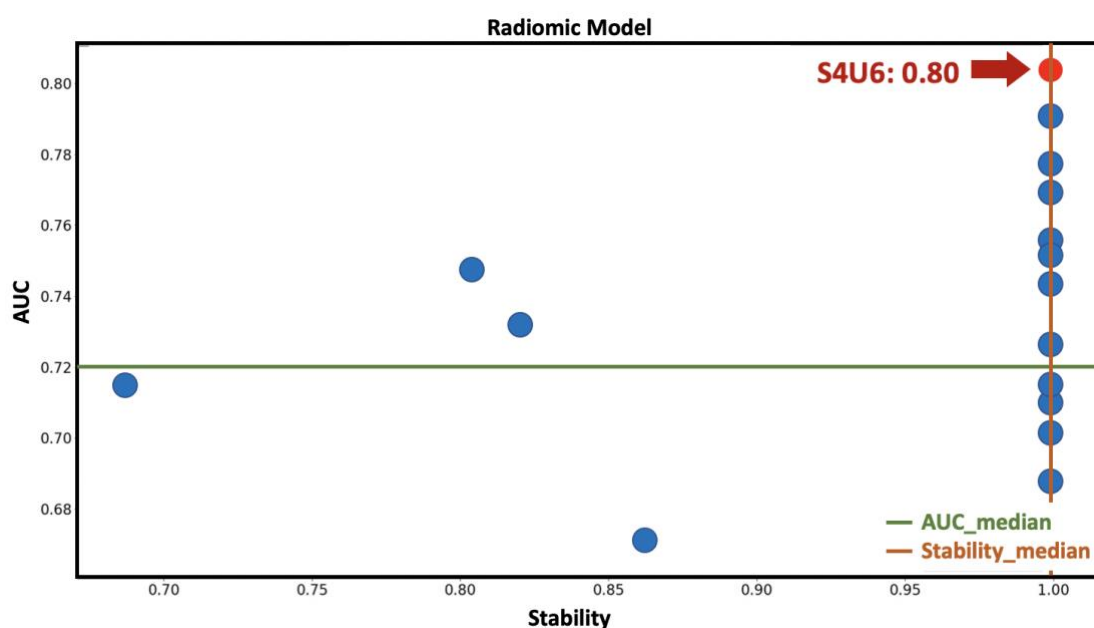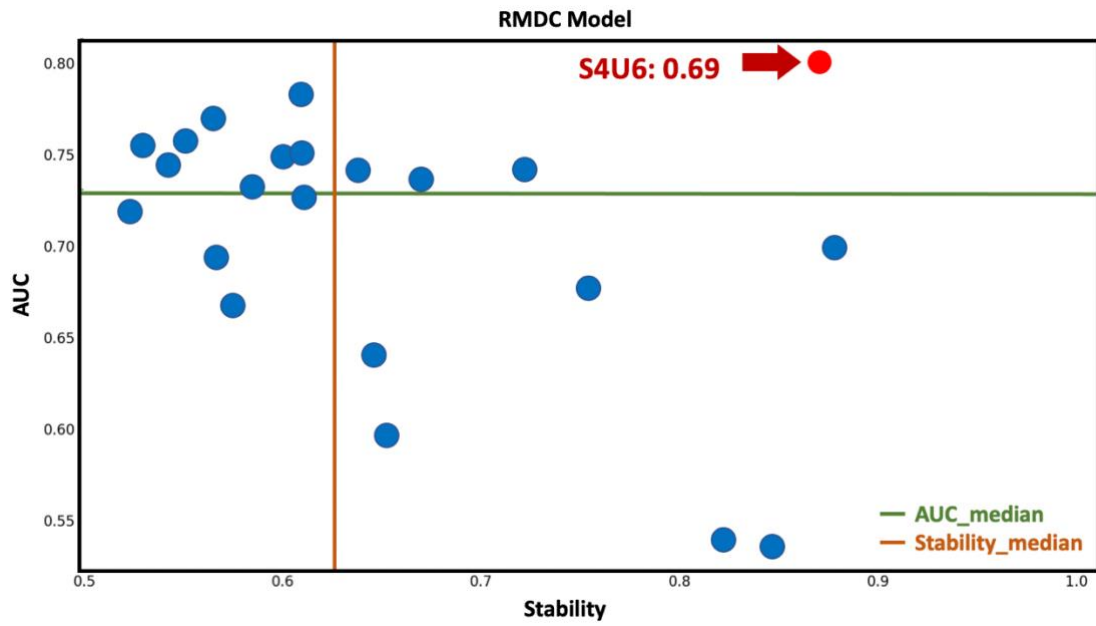
Figure 27. Results of decision graphs for the RMDC multi-omics model, with the corresponding optimal FS algorithms and the product scores being S4U6 and 0.69, respectively (as indicated by Red arrow).

Table 8. Results of optimal combinations of supervised and unsupervised feature selection (FS) algorithms for the four single-omics models and four multi-omics models.

Feature Set	Optimal FS Combination
Single-omics Model	
Radiomics (R)	S4U6
Morphology (M)	S3U3
Dosiomics (D)	S2U1, S3U6
Contouromics (C)	S2U6
Multi-omics Model	
RM	S4U6
RD	S1U6
RC	S4U6
RMDC	S4U6

Figure 28 (A-D) and **Figure 29 (A-E)** illustrate the change of average AUC scores (and its STD shown in shadow) in both training and hold-out test sets against varying number of features for the 4 single-omics models and the 4 multi-omics models, respectively. Final models were determined when the average AUC scores on the hold-out test sets reached its maximum.
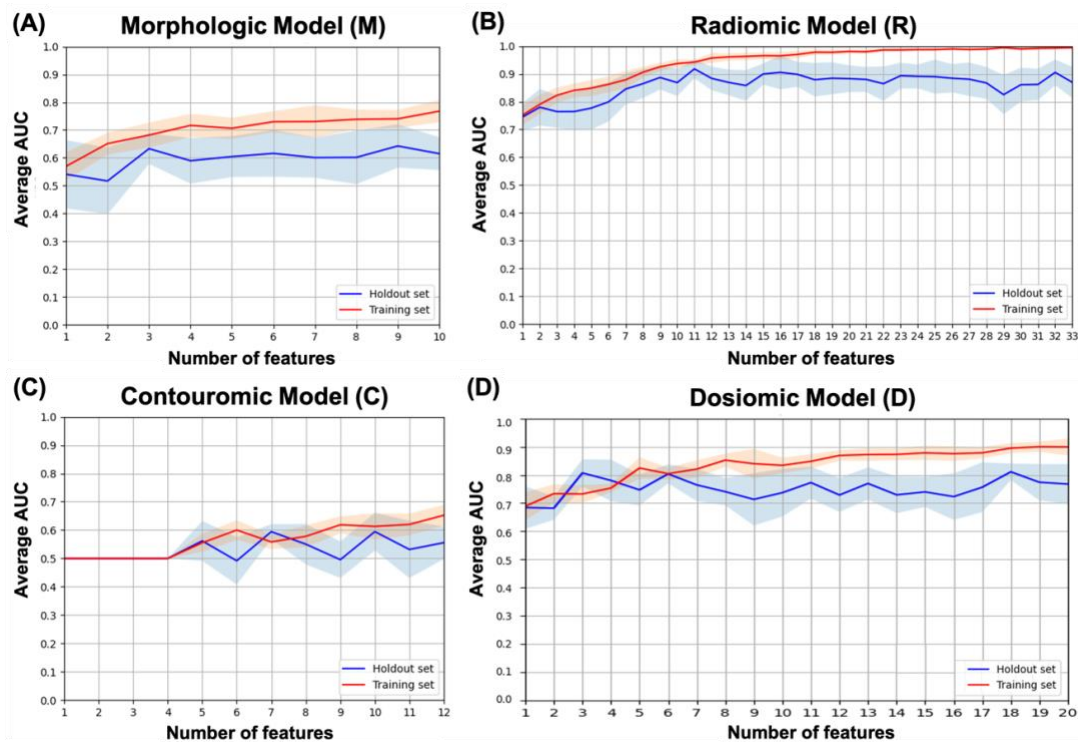
Figure 28 (A-D). Change of average AUC scores (and its STD shown in shadow) in both training (red curves) and hold-out test sets (blue curves) against varying number of features for the four single-omics models: Morphology (A), Radiomics (B), Contouromics (C), and Dosiomics (D), respectively.

(A)

RM Multi-omics Model (RM)

Average AUC vs Number of features

(B)

RD Multi-omics Model (RD)

Average AUC vs Number of features

(C) RC Multi-omics Model (RC)

(D) RMDC Multi-omics Model (RMDC)
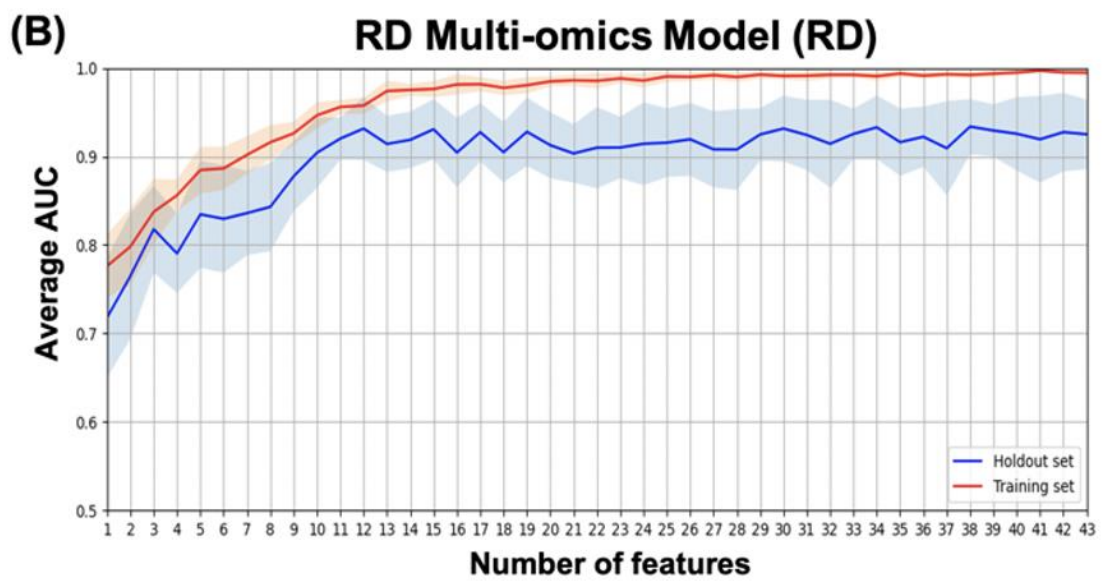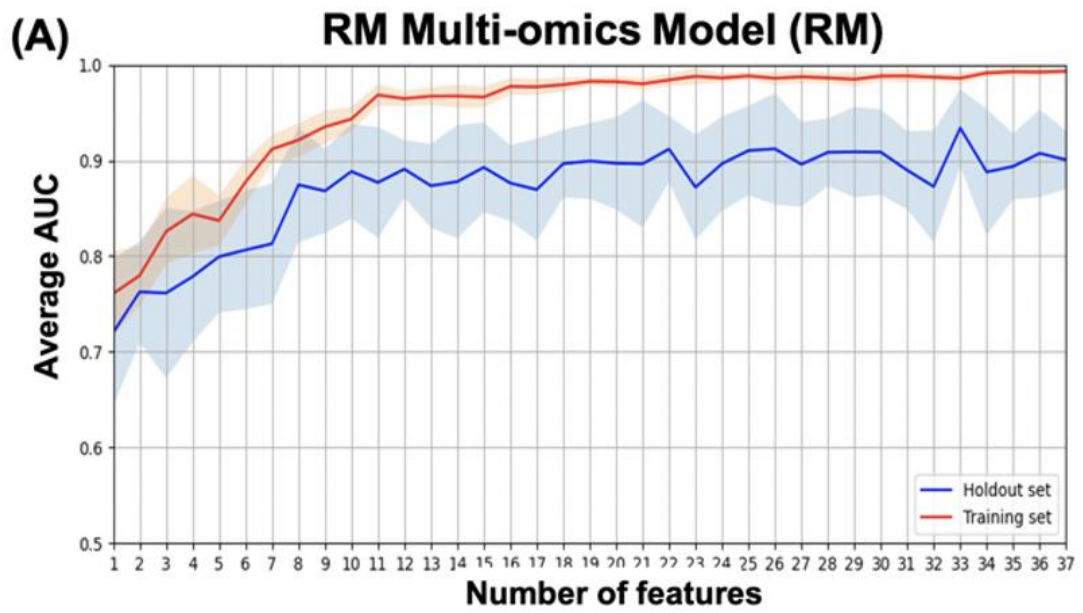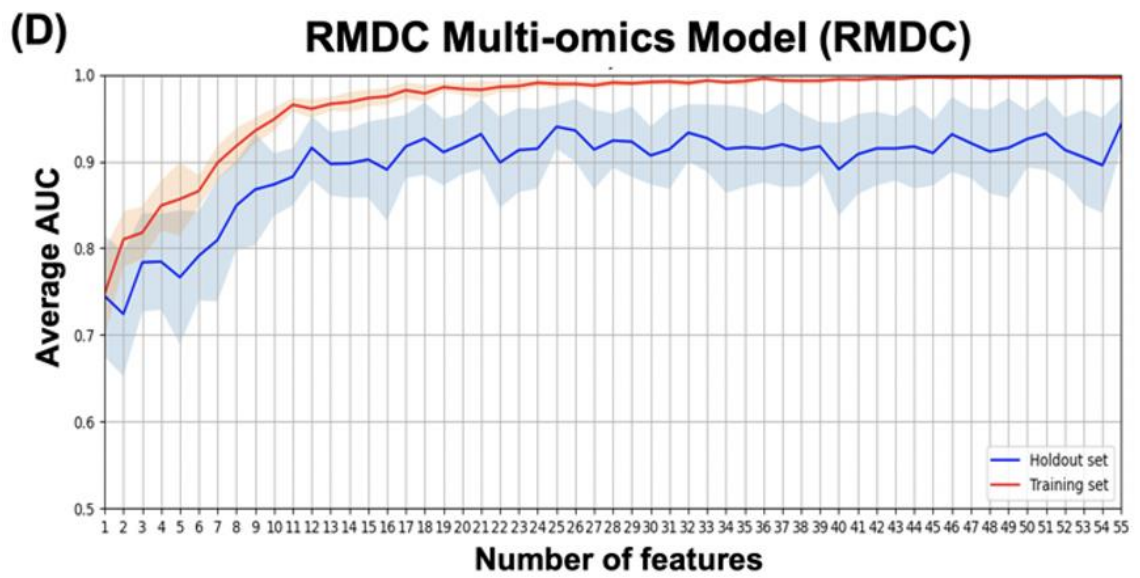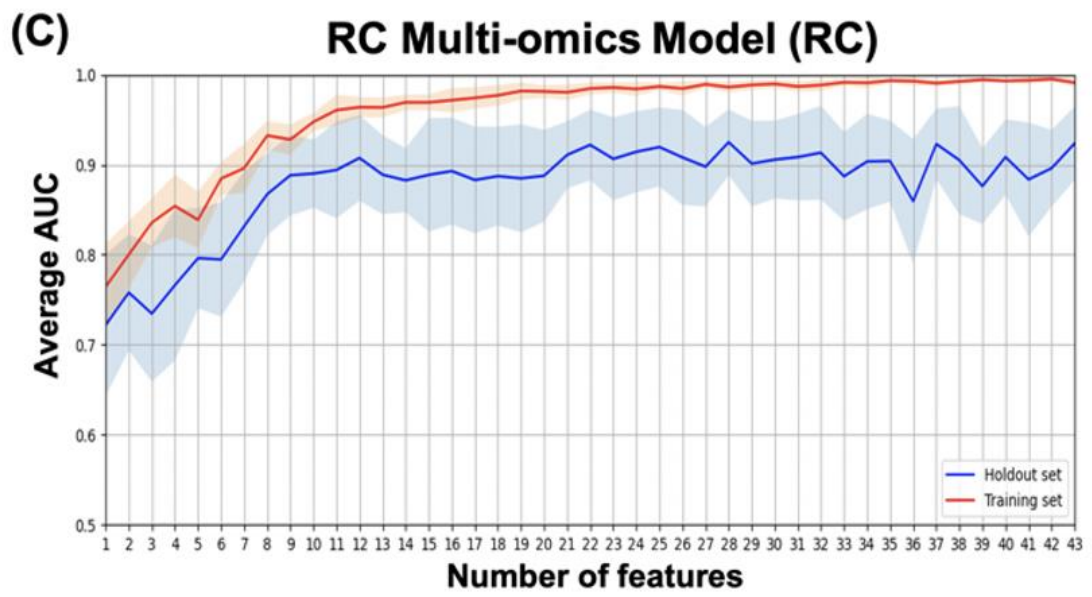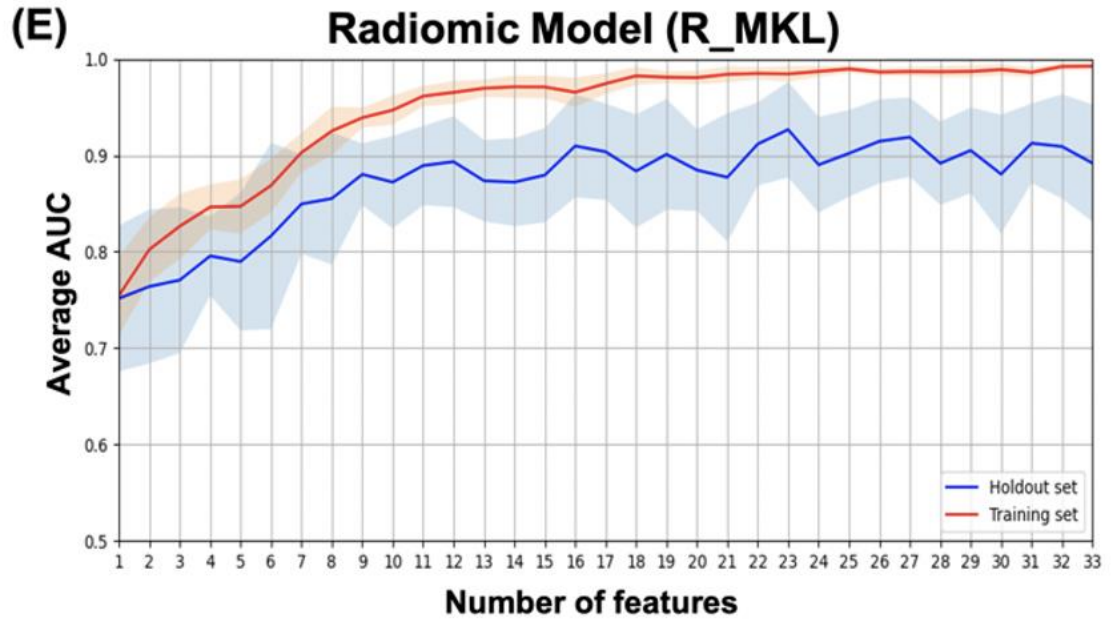
Figure 29 (A-E). Change of average AUC scores (and its STD shown in shadow) in both training (red curves) and hold-out test sets (blue curves) against varying number of features for the four multi-omics models: Radiomics+Morphology (A), Radiomics+Dosiomics (B), Radiomics+Contouromics (C), and Radiomics+Morphology+Dosiomics+Contouromics (D), and the Radiomics models trained by using MKL algorithm (E), respectively.

Table 9 summarizes the total number and distribution of the selected features in the final models. Interestingly, it can be observed that radiomic features are dominant in all the four multi-omics models, compared to M, C, and D features.

Table 9. A summary of total number and distribution of selected features in the final models. (*Not applicable)

	Number of Final Selected Features				
	Total	**R**	**M**	**C**	**D**
Radiomics (R)	11	11	*	*	*
Morphology (M)	9	*	9	*	*
Contouromics (C)	10	*	*	10	*
Dosiomics (D)	18	*	*	*	18
Radiomics (R_MKL)	23	23	*	*	*
RM	33	31	2	*	*
RC	28	27	*	1	*
RD	38	30	*	*	8
RDCM	55	36	3	9	7

5.3.3 Model Comparison

Figure 30(A-B) indicates box-whisker plots of the average AUC distributions for the final single-omics models, and **Figure 30 (C-D)** for the multi-omics models and the Radiomic models trained by using MKL algorithms, in training and hold-out test sets. A summary of the statistical estimates of model performance is provided in **Table 10A and 10B.**

From **Figure 30 (A-B)** and **Table 10A**, it can be seen that the Radiomic model (R) significantly outperformed all other studied single-omics models (p-value < 0.0001), achieving an average AUC of 0.942 (STD: 0.009, 95%CI: 0.938-0.946) in the training set and 0.918 (STD: 0.034, 95%CI: 0.903-0.933) in the hold-out set.

The Dosiomic model (D) was the second best single-omics model with an average AUC of 0.895 (STD: 0.018, 95%CI: 0.887-0.903) in the training set and 0.811 (STD: 0.029, 95%CI: 0.798-0.824) in the hold-out set. This was followed by the Morphologic model (M) which yielded an average AUC of 0.740 (STD: 0.032, 95%CI: 0.726-0.754) in the training set and 0.643 (STD: 0.078, 95%CI: 0.608-0.677) in the hold-out set, while the Contouromic model (C) was the most underperforming model, producing an average AUC of 0.664 (STD: 0.052, 95%CI: 0.641-0.687) in the training set and 0.550 (STD: 0.082, 95%CI: 0.514-0.586) in the hold-out test set.

From **Figure 30 (C-D)** and **Table 10B**, it can be observed that the RMDC model had the highest AUC of 0.997 (STD: 0.003, 95%CI: 0.995-0.998) in the training set and 0.943 (STD: 0.029, 95%CI: 0.931-0.956) in the hold-out set, compared to other types of multi-omics models. While it statistically outperformed the other three studied multi-

omics models (RM, RD, and RC) in the training set, it did not reach the statistical significant level in the hold-out test set.

Notably, when the R model was re-trained using MKL algorithm (referred to as R_MKL model), the average AUC boosted to 0.984 (STD: 0.008, 95%CI: 0.981-0.988) in the training set and 0.927 (STD: 0.050, 95%CI: 0.905-0.948) in the hold-out set. The development and performance of the R_MKL model can be seen in **Figure 30 (E)**, **Figure 30 (C-D)** and **Table 10B**. Surprisingly, further comparisons between the R_MKL model and all the 4 studied multi-omics models indicated that there were no significant differences in model discriminability between R_MKL and all other multi-omics models in the hold-out test set **(Figure 30D, Table 10B).**
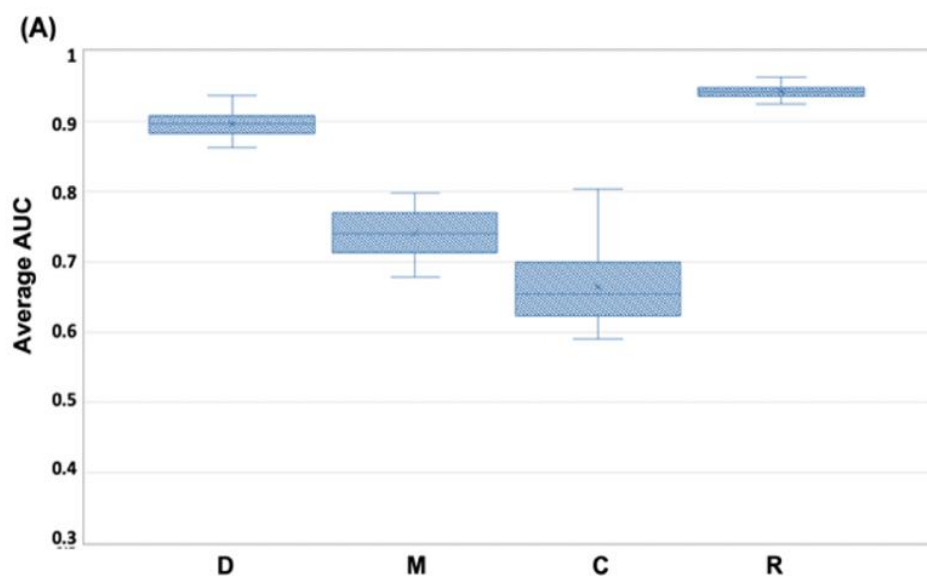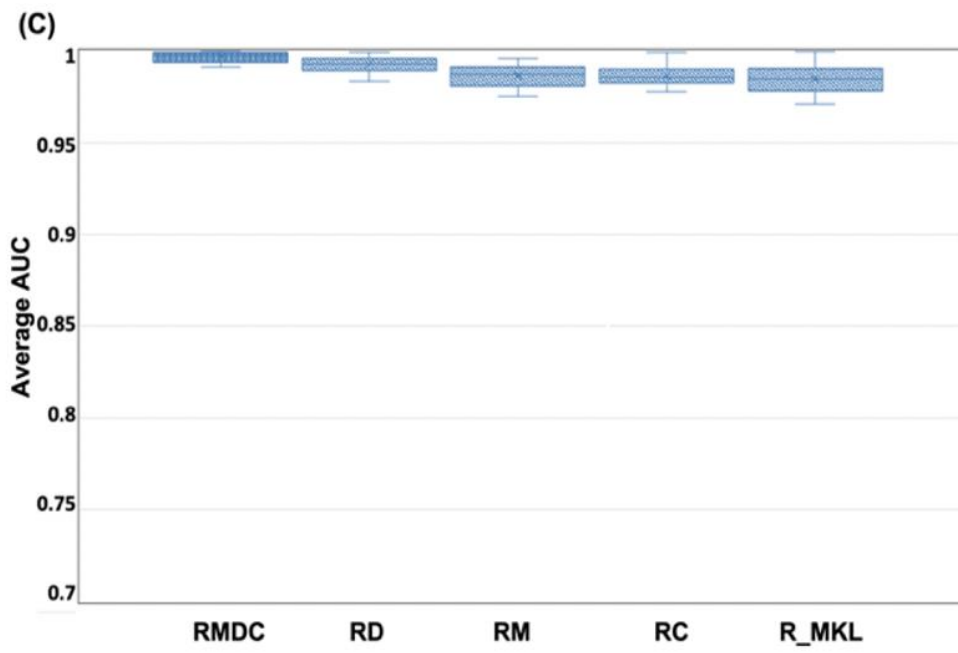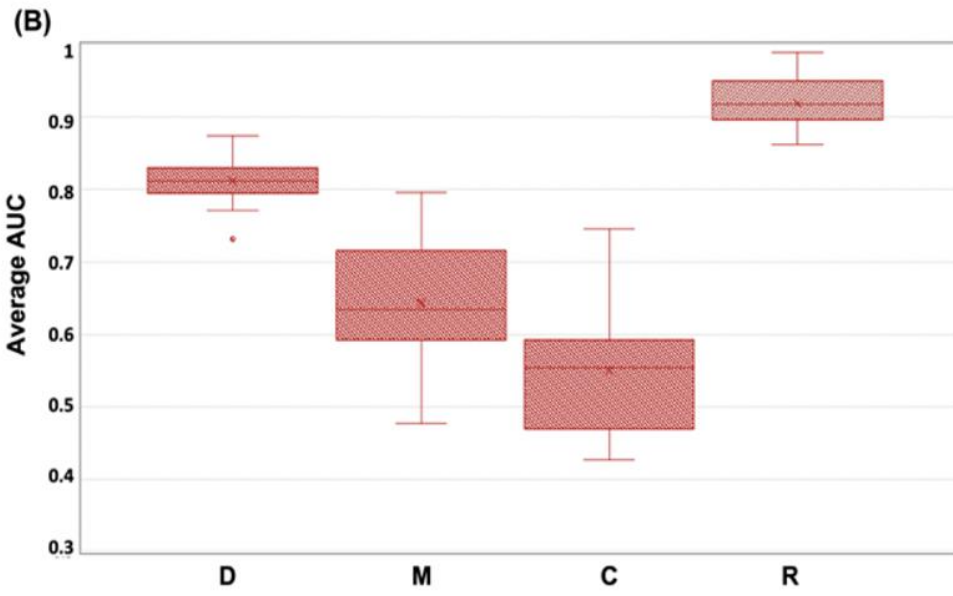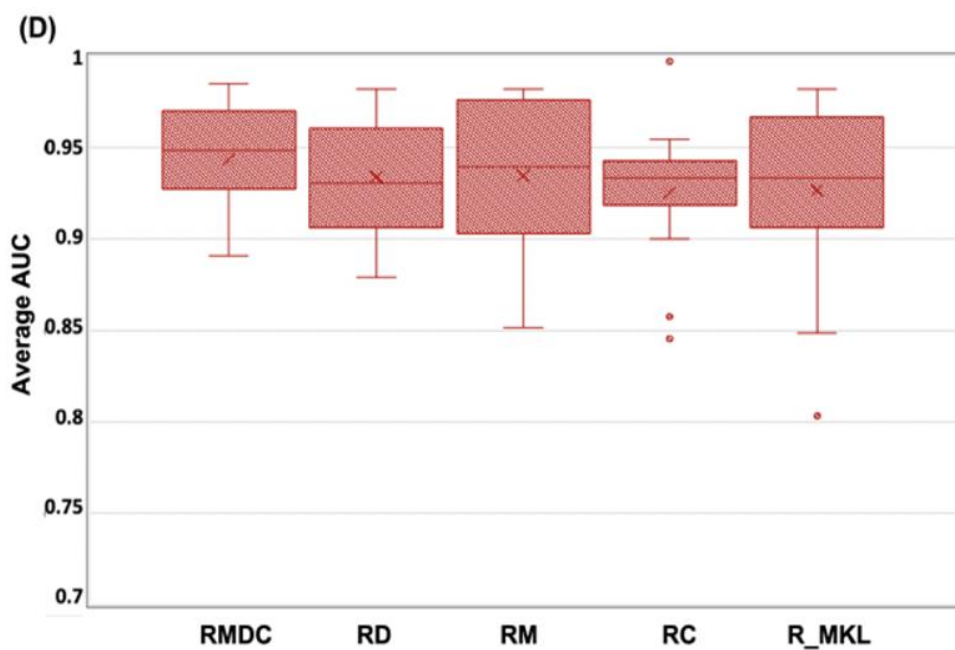
Figure 30 (A-D). Box-whisker plots of the average AUC distribution for the final single-omics models in training set (A) and hold-out test set (B), and for the multi-omics models and the Radiomic models trained by using MKL algorithms in training (C) and hold-out test (D).

Table 10 (A-B). A summary of statistical estimates on performance of single-omics models (A), multi-omics models and the Radiomic model trained by using MKL algorithm (B).

Table 10A	Training set				Hold-out test set			
	Avg. AUC	STD	95%CI	p-value	Avg. AUC	STD	95%CI	p-value
Single-omics Model								
Radiomics (R)	0.942	0.009	(0.938,0.946)	Ref	0.918	0.034	(0.903,0.933)	Ref
Morphology (M)	0.740	0.032	(0.726,0.754)	<0.0001*	0.643	0.078	(0.608,0.677)	<0.0001*
Contouromics (C)	0.664	0.052	(0.641,0.687)	<0.0001*	0.550	0.082	(0.514,0.586)	<0.0001*
Dosiomics (D)	0.895	0.018	(0.887,0.903)	<0.0001*	0.811	0.029	(0.798,0.824)	<0.0001*

Table 10B	Training set					Hold-out test set				
	Avg. AUC	STD	95%CI	p-value	p-value	Avg. AUC	STD	95%CI	p-value	p-value
Multi-omics Model										
RM	0.986	0.006	(0.983,0.989)	<0.0001*	0.473	0.934	0.041	(0.916,0.952)	0.362	0.622
RD	0.992	0.004	(0.990,0.994)	<0.01*	<0.001*	0.934	0.031	(0.920,0.947)	0.365	0.641
RC	0.986	0.006	(0.984,0.989)	<0.0001*	0.422	0.925	0.037	(0.909,0.941)	0.140	0.924
RMDC	0.997	0.003	(0.995,0.998)	Ref	<0.0001*	0.943	0.029	(0.931,0.956)	Ref	0.213
Radiomic Model (trained by MKL)										
R_MKL	0.984	0.008	(0.981,0.988)	<0.0001*		0.927	0.050	(0.905,0.948)	0.213	Ref

5.4 Discussion

ART aims to compensate for patient-specific anatomic variations between fractions in NPC patients, while routine ART implementation on patient basis would undoubtedly pose immense burden to clinic. Previously, we were the first to demonstrate the capability of tumoral Radiomics from pre-treatment MRI for prediction of ART eligibility in NPC patients [70]. In this study, we investigated a variety of single-omics and multi-omics models from multi-modal images, with an eye towards identifying their roles in predicting ART eligibility in NPC and providing insights into development of ART eligibility screening strategy in NPC in the long run. In this discussion, we attempted to highlight key findings of the study, scrutinize possible underlying reasons, and provide research community with potential directions in the future.

Results of the study showed that the R model significantly outperformed all other studied single-omics models (i.e., M, C and D models, all p-value < 0.0001), achieving an average AUC of 0.942 (STD: 0.009, 95%CI: 0.938-0.946) in the training set and 0.918 (STD: 0.034, 95%CI: 0.903-0.933) in the hold-out test set **(Figure 30A-B, Table 10A).** Among the studied multi-omics models, the RMDC had the highest average AUC in both cohorts **(Figure 30C-D, Table 10B)**, however, its difference to the other three models (RM, RD and RC) did not reach the level of statistical significance in the hold-out test sets **(Table 10B)**. Surprisingly, there was no statistical difference between the R_MKL and all the studied multi-omics models in the hold-out set **(Table 10B).** In other words, addition of other types of -omics features into a radiomic model did not demonstrate statistically significant improvement in model performance, suggesting the

dominant role of Radiomic features in prediction of multifactorial ART eligibility in NPC. Besides, Radiomic features accounted for majority of the final selected features, ranging from 64% to 94%, in all the studied multi-omics models **(Table 9).** We speculated that the dominant role of Radiomics found in this study could partially be explained by both the unique nature of Radiomics and the multi-factorial nature of the ART eligibility.

First, the outstanding predictability of Radiomics in this study may largely lie in its unique capability in unraveling intrinsic tissue property regarding response to treatment perturbations, which can be tissue-type dependent and patient-specific. There is mounting evidence in the literature showing the power of Radiomics in predicting treatment response in various cancer diseases [5, 75, 77, 78, 81, 112]. For instance, Hou et al. investigated CECT-based biomarkers for prediction of therapeutic response to chemo-radiotherapy in esophageal carcinoma and reported the discriminability of their model in AUC ranging from 0.686 to 0.727 [77]. Wang et al. developed a radiomic signature combining features from multi-modal MR imaging sequences for prediction of early treatment response to induction chemotherapy in NPC patients, achieving an AUC of 0.822 [5]. Piao et al. devised a MR-based radiomic model to distinguish sensitive and resistant tumors in NPC patients following induction chemotherapy, yielding an AUC of 0.905 [81]. In these studies, the tumor response was defined in accordance with the Response Evaluation Criteria in Solid Tumors (RECIST) via quantitative assessment of tumor shrinkage, which follows the same line of thought as in this present study. Apart from this, Ramella et al. performed radiomic analysis of pre-treatment CT images of replanned non-small-cell lung cancer patients and generated a radiomic signature for prediction of tumor shrinkage during chemo-

radiotherapy, yielding an AUC of 0.82 [112]. Yu et al. analyzed tumoral radiomic features from multi-parametric pre-treatment MRI of NPC patients and developed several prediction models for ART eligibility, achieving AUC ranging from 0.750 to 0.930 [70]. All the above evidence indicates the outstanding capability of Radiomics in divulging patient-specific intrinsic tissue biologic characteristics for discerning respondent and non-respondent cancer patients upon treatment perturbations, laying great foundation for predicting patient-specific anatomic variations for ART eligibility for NPC in this study.

By contrast, Dosimoics mainly characterizes aggressiveness of a specific treatment plan by capturing dose statistics from the entire three-dimensional dose distribution map within each of the studied organ structures, while it appears to convey little information on tissue responsiveness upon treatment perturbations. To a degree, this may shed some light on the well-recognized phenomenon where the same-staged patients experienced a diverse range of treatment outcome/response following identical treatment (same degree of treatment aggressiveness). Herein, we emphasize that results of the study do not deny the potential of Dosimoics in predicting treatment response. Indeed, it is worth noting that the D model was the second best-performing model in this study, giving rise to an average AUC of 0.895 (STD: 0.018, 95%CI: 0.887-0.903) in the training set and 0.811 (STD: 0.029, 95%CI: 0.798-0.824) in the hold-out test set **(Figure 30A-B, Table 10A).** This result appears in agreement with most of the previous studies investigating triggering factors for ART in NPC, where radiation dose deposited was regarded as a prime factor for morphologic volume shrinkage of targets and OARs during the RT course, which may in turn incur intolerable dosimetric deviations from initial treatment plan and hence trigger ART implementation. For instance, Cheng et al.

analyzed repeated planning CT and MR scans at 30 and 50-Gy intervals [50]. They reported that the shrinkage of both primary tumor and nodal lesions were higher when 50-Gy was delivered (13% and 29%, respectively) than that when 30-Gy was given (9% and 16%, respectively) and similar trend was also observed for bilateral parotid glands, which jointly led to significant increase in doses to numerous critical OARs, triggering implementation of ART. In this regard, several research groups have also suggested to incorporate dosimetric deviations in targets and/or OARs (such as parotid glands) as part of the ART regimen [58, 108, 114, 115]. Of note, although Dosiomics has recently been studied for prediction of toxicity [7, 9, 119, 121, 136] and prognosis [120, 121] in cancer patients, its potential in treatment response prediction, in particular on the basis of the RECIST criteria, has not been reported. Future studies in this aspect are recommended to confirm its capability in this regard.

On the other hand, Morphologic and Contouromic features merely depict initial morphometric characteristics and geometric relationship between organs, respectively. They share commonality in their distinct disparity against Radiomics in that they both carry little or no underlying biologic information of the studied organ structures. This may in part explain the fair-to-poor predictive performance of the M and C models in the study, yielding an AUC of 0.643 (STD: 0.078, 95%CI: 0.608-0.677) and 0.550 (STD: 0.082, 95%CI: 0.514-0.582) in the hold-out test set, respectively **(Figure 30A-B, Table 10A).**

In addition, the multifactorial nature of ART eligibility in the context of NPC disease may further elucidate why Radiomics plays a dominant role in this study, irrespective of additional types of -omics features. ART eligibility in NPC depends on

154

multiple organs located in a confined space of head-and-neck regions. GTVnp, GTVn and bilateral parotid glands are all bulky organ structures within the nose-pharynx ministry, responsiveness of these structures upon treatment perturbations jointly determines the degree of patient-specific alternations in anatomy, hence affecting the demand for ART. Given the unique superiority of Radiomics in unravelling intrinsic tissue biologic response, we inferred that the role of Radiomics could become increasingly important when more organ structures come into play in contributing to the studied outcome (i.e., the ART eligibility), compared with other types of -omics features. This may, to some extent, provide an insight into the findings that Radiomic features accounted for the largest proportion of the final selected features in all the studied multi-omics models **(Table 9);** and that the multi-organ-based R model performed far better than other single-omics models (all p-value < 0.0001) **(Table 10A)**; and that incorporating Morphologic and/or Dosiomic and/or Contouromic features into the radiomic model did not demonstrate statistically significant improvement in the hold-out test set **(Table 10B)** (all p-value > 0.05). Herein, we highlight that findings of this study may provide research community with valuable insights into development of pre-treatment stratification strategies for ART eligibility in NPC patients, potentially facilitating clinical implementation of ART in the future.

Although there exists a lack of studies on revealing multi-omics in prediction of multi-organ triggering outcome, results from a few studies in the literature may worth attention. Sheikh et al. investigated radiomics and dosimetric features from bilateral parotid and submandibular glands (i.e., four separated organ structures) for predicting xerostomia, and reported that addition of dosimetric and clinical factors into a joint-CT-MR radiomic model did not lead to statistically significant improvement in model

155

performance [137], which appears to be in line with the current findings. By contrast, Jiang et al. reported superior model performance when using both radiomic and dosimetric features from five lung sub-regions for predicting radiation pneumonitis than when using radiomic features alone [138], which may appear contradictive to the findings. However, it should be noted that the features in their studies were essentially derived from a single organ – the same lung tissue, rather than individual separated organ structures as in this current work. Further, unlike the present work, only CT-based radiomics was adopted in their study, which may lead to a relatively weaker predictive power than as if it were developed from multi-modal images that capture complementary tissue characteristics. Notwithstanding, this presents an interesting area to be explored and a close scrutinization of different types of features in prediction of a multi-organ contributing outcome is highly warranted in the future to further affirm the role of radiomics in context.

This study has several limitations. First, the models were developed and validated in a small-sized single cohort of NPC patients who received CCRT under Tomotherapy machine. While we believe such a homogeneous dataset is advantageous for model building, findings of the study require further validation in a large multi-cohort study. However, it is worth noting that the goal of this study was to assess the role of different omics-based prediction models for ART eligibility in NPC, instead of developing a generalizable model for clinical adoption. Thus, results of this study still deserve great attention in the community. Second, this study employed a large number of features for model building, which may lead to model overfitting in a small-sized cohort. In this regard, we deployed a strategic approach of determining optimal FS combinations that were used for feature dimensionality reduction prior to model development. The

remnant feature sets were of high outcome relevance and low feature redundancy, and only 10 to 33 and 37 to 55 features were input to the modelling algorithms for developing single-omics and multi-omics models, respectively.

To conclude, comparisons among all the studied models indicated that the Radiomic model was found to play a dominant role for ART eligibility in NPC patients; and Radiomic features accounted for the largest proportion of features in all the four multi-omics models, suggesting its governing power in ART eligibility prediction.

6.   Chapter 6: Summary and Reflection

6.1   Summary of Current Findings

Significant shrinkage of neck nodal lesions has been commonly reported in NPC patients undergoing RT treatment, which is a key contributor for ill-fitted thermoplastic mask (IfTM), a routinely used immobilization device for safeguarding radiation delivery accuracy and safety. In severe situation of IfTM, an ad hoc ART may be required to ensure accurate and safe radiation delivery and to maintain treatment efficacy. Presently, the entire procedure for evaluating an eligible ART candidate is time-consuming, resource-demanding, and highly inefficient. In Chapter 3, I developed a CT-based models using radiomic features extracted from neck nodal lesions of NPC patients in the hope of facilitate clinical implementation of IfTM-triggered ART. The results of the study showed that radiomics plays a key role in predicting IfTM risk in NPC patients in both discovery cohort (corrected AUC: 0.784, BCa 95%CI: 0.673, 0.859) and independent external cohort (corrected   ACU: 0.723, BCa 95%CI: 0.534, 0.859). The developed radiomic models also outperformed traditional clinical model. To a degree, the superiority of radiomics may be ascribed to its unique property of unravelling tissue bio- logic characteristics in response to treatment perturbations. Indeed, an enormous number of articles in the literature have demonstrated the capability of radiomics in predicting tumor responsiveness on the basis of Response Evaluation Criteria in Solid Tumors (RECIST). where the criteria are defined according to the degree of tumor volume shrinkage following treatment, which appears to follow the same line of thought as in the study. Detailed discussions of the selected features in the radiomic models and their relevance to the literature are provided in the discussion

section in Chapter 3. Overall, this CT Radiomic-based pre-treatment identification of NPC patients who are at a greater risk of having an IfTM is of high potential to alleviate the clinical workload, facilitate ART practice in the clinic, and achieve personalized RT.

Apart from the neck nodal shrinkage, enormous amounts of studies have demonstrated significant volume reduction of primary NPC tumor throughout the radiotherapy course. Indeed, when significant tumor shrinkage occurs, those critical organs might move into the original high dose region, leading to deleterious dosimetric impact on the surrounding organs and/or insufficient dose delivery to targets. It is worth-noting that pre-treatment multi-parametric MR images are standard-of-care of radiotherapy workflow in NPC patients and their versatile capability in highlighting tissue anatomic and physiologic information. In Chapter 4, a variety of MR-based radiomic models, including contrast-enhanced T1-weighted (CET1-w) and T2-weighted (T2-w) MR-based prediction model separately and in combination, were presented for evaluating their efficacy in assessing ART eligibility in a single cohort of NPC patients. The joint T1-T2 model achieved the highest AUC of 0.964 in the training cohort and 0.930 in the testing cohort, followed by CET1-w model with AUC of 0.962 in the training cohort and 0.852 in the testing cohort, and T2-w model with AUC of 0.895 in the training cohort and 0.750 in the testing cohort. Detailed discussions of the selected features in the radiomic models and their relevance to the literature are provided in the discussion section in Chapter 4. The study follow the same line of thought regarding capability of radiomics in predicting tumor shrinkage upon treatment perturbation on the basis of RECIST. The high predictability of the developed MR-

based radiomic model may offered valuable insights into screening at-risk NPC patients for ART implementation in clinic.

Lastly, it has been discovered that not only primary tumors and neck nodal lesions may experience the above-mentioned volumetric changes following radiotherapy, surrounding healthy organs (such as parotid glands) may exhibit significant morphometric volume and/or geometric alterations, all of which may then individually and jointly alter patient anatomy and jeopardize the efficacy of the original treatment plan. Further, three-dimensional spatial dose distribution within the pertinent organ structures (reflected by Dosiomics), initial morphologic characteristics of pertinent organs (reflected by morphologic features), and initial geometric relationship between different internal organs (reflected by Contouromics), may also work in conjunction with radiomics attributes for determining the final patient anatomy that necessitates ART implementation. In Chapter 5, I developed a variety of single-omics models (Radiomics: R, or Dosiomics: D, or Contouromics: C or Morphologic: M) and various multi-omics models for different pertinent organ structures, with a hope to identify the role of different omics-based models in prediction of ART eligibility in NPC patients. Results of the study indicated that the R significantly outperformed all other studied single-omics models (i.e., M, C and D models, all p-value < 0.0001), achieving an average AUC of 0.942 (STD: 0.009, 95%CI: 0.938-0.946) in the training set and 0.918 (STD: 0.034, 95%CI: 0.903-0.933) in the testing set. Among the studied multi-omics models, the RMDC had the highest average AUC in both cohorts, however, its difference to the other three models (RM, RD and RC) did not reach the level of statistical significance in the testing sets. Surprisingly, addition of other types of -omics features into a radiomic model did not demonstrate statistically significant

160

improvement in model performance, suggesting the dominant role of Radiomic features in prediction of multifactorial ART eligibility in NPC. Besides, Radiomic features accounted for majority of the final selected features, ranging from 64% to 94%, in all the studied multi-omics models. This result suggested that the dominant role of Radiomics found in this study could partially be explained by the unique nature of Radiomics. On the contrary, Dosiomics may mainly characterize aggressiveness of a specific treatment plan by capturing dose statistics from the entire three-dimensional dose distribution map within each of the studied organ structures, while possibly conveying little information on tissue responsiveness upon treatment perturbations. Similarly, Morphologic and Contouromic features merely depict initial morphometric characteristics and geometric relationship between organs, respectively. In fact, ART eligibility in NPC depends on multiple organs located in a confined space of head-and-neck regions. GTVnp, GTVn and bilateral parotid glands are all bulky organ structures within the nose-pharynx ministry, responsiveness of these structures upon treatment perturbations jointly determines the degree of patient-specific alternations in anatomy, hence affecting the demand for ART. The best-performing radiomic model was developed by using radiomic features extracted from a number of pertinent organs, including primary NPC tumor, metastatic neck lymph nodes, and bi-lateral parotid glands, which appears to be in the same line of thoughts as the above interpretations. Results from the study may offer valuable insights into development of ART eligibility screening strategy in NPC in the long run.

It is worth noting that the 4 single-omics models developed by using Ridge logistic regression (linear) and 4 multi-omics models developed using MKL algorithm (non-linear) were studied for multi-organ omics-based prediction of ART eligibility in NPC

161

patients, as described in **Chapter 5.2.6**. Ridge classifier is a typical statistical approach for resolving bias-variance trade-off with the use of a linear function, while it is considered deficient in capturing the difference in representations of multi-omics data and non-linear relationship between predictors and prediction outcome. The MKL algorithm, on the other hand, was adopted for model development on multi-omics data. Unlike single-omics features, different types of multi-omics data may contain distinctly different data representations. In kernel learning, such data representations are implicitly selected through kernel. The use of kernel allows machine learning practitioners to define similarity between two types of -omics data and a proper regularization term for a learning task. The MKL algorithm was used to divulge complementary (non-linear) relationship between different types of -omics features and prediction outcomes. Specifically, two types of kernels (Gaussian and Polynomial) with a range of kernel parameters were applied. Each kernel was embedded into the feature space of a given multi-omics feature set for subsequent multi-omics fusion (an overall MKL framework is indicated in **Figure 19**).

To determine the final model for prediction of ART eligibility, discriminability of the four single-omics models (all linear) were first inter-compared in aspect of distribution of the AUC scores across the 20 iterations. It was found that the Radiomic model (R) outperformed other single-omics models, achieving average AUC of 0.942 (STD: 0.009, 95%CI: 0.938-0.946) and 0.918 (STD: 0.034, 95%CI: 0.903-0.933) in training and hold-out testing set (all p<0.0001, indicated in **Table 10A).** This best-performing R model was then re-trained using MKL algorithm, referred to as R_MKL model (non-linear), for the comparison against the four multi-omics models. When the R_MKL model was compared to other multi-omics models (all non-linear), it turned

out that these were no significant difference in aspects of AUC in the hold-out test set (all p>0.05, indicated in **Table 10B)**, suggesting that the dominant role of radiomic features in prediction of ART eligibility in NPC patients. Besides, it is worth pointing out that although the average AUC score of the R_MKL model (non-linear) in the hold-out testing set (AUC: 0.927, STD: 0.050, 95%CI: 0.905-0.0948) was higher compared to the R model (linear) (AUC: 0.918, STD: 0.034, 95%CI: 0.903-0.0933), the difference did not reach the level of statistical significance (p>0.05). The R_MKL model presented no statistical benefits in terms of classification power while containing twice amounts of predictors, compared to the R model. Therefore, the R model was finally chosen as a parsimonious model that achieved a satisfactory level of classification performance using as few explanatory variables as possible in order to avoid complex model interpretability when more and more predictors are involved and to reduce the risk of false positive findings when testing on new datasets. The intention of developing a parsimonious model in this dissertation project is also in line with a number of studies in the literature [139-141].

6.2   *Limitations & Recommendations for Future Studies*

The limitations of the three studies are described in each of the corresponding Chapters (**Chapter 3.4, Chapter 4.4, Chapter 5.4**). Particularly, three major limitations of methods and works done: **(1)** model and radiomic feature interpretability, **(2)** model reliability, and **(3)** limited sample size, are summarized in this section, followed by recommendations for future works.

(1) Radiomic model interpretability is the central premise for its wide-spreading clinical implementation. Although a well-established protocol - Image Biomarker Standardization Initiative (IBSI) has been widely used for calculation of radiomic features [98], the definitions of these features are based on complex mathematical equations and there often exists a complicated linkage between the radiomic feature and pathophysiological or biological mechanisms. Therefore, it is still challenging for this high-throughput radiomics to offer biologic underpinnings of the association between radiomic predictors and the prediction outcome-of-interest, which has largely impeded the clinical adoption of the developed radiomic models.

In Chapter 3, for example, although a number of CT-based features were identified for prediction of IfTM-triggered ART, which predicted the shrinkage of neck nodal lesions in NPC patients, there is still a lack of thorough understanding on the biologic relationship between the radiomic predictors and tumor shrinkage. In Chapter 5, despite that an attempt was made to enhance the model interpretability by incorporating dose-related and organ morphologic characteristics into the model development process, it turned out that the radiomic-alone model outperformed all other studied models after statistical analysis.

Multiple efforts in the past years have been made to explore biologic meaning of radiomic features [116, 141-144]. In particular, Tomaszewski et al. provided a comprehensive review on the biological meaning of radiomic features in several key perspectives [144], including genomic association (i.e., radio-genomics), histopathologic relationship, etc.

For radio-genomics, for instance, Rios Velazquez et al. revealed association of CT-based radiomics with somatic mutations in 763 lung cancer patients [143]. They developed a radiomic signature related to imaging heterogeneity that was associated with epidermal growth factor receptor (EFGR) and Kirsten rat sarcoma (KRAS) viral ono-gene homolog. Gevaert et al. developed gene-based signatures to predict PET/CT-based radiomic association in lung cancer patients, and vice versa, and found that the accuracy can be up to 86%, indicating strong association between genomic and radiomic features [142].

For histopathologic association with radiomics, Tunali et al. adopted immunohistochemistry (IHC) and genetic factors of non-small cell lung cancer patients for prognosis prediction of treatment response following immunotherapy [141]. They found that the CT-based radiomic feature, gray-level co-occurrence (GLCM) inverse difference, was positively related to tumor hypoxia. Sun et al. generated a CT-based radiomic signature for immune cell infiltration, and they found that the signature was associated with genetic expression of PD-L1 and anti-PD-1, pathologic findings, and overall survival [116].

Recommendations: With the increasing importance of biologic interpretability of radiomic features and models, researchers should make their greatest efforts in the future to include biologic tests, such as pathologic test, genetic test and tumor hypoxia information on PET images, at the stage of study design for the sake of performing correlation analysis between radiomic predictors and the tumor pathology and genotypes. More importantly, biological validation of the developed radiomic signature should be performed to validate the biological role of the prediction model, as this will

enhance the model's explanability from the biologic grounds, paving the way toward bench-to-bedside model translation in clinic.

(2) Model reliability against variabilities in intra-/inter- organ segmentations and imaging protocols, etc. is essential for model clinical utility. Variabilities in imaging protocols within or between institutions and organ segmentation have been considered as prime culprits that deleteriously influence feature reproducibility.

For imaging protocol variabilities, the potential impacts on feature reproducibility have been extensively studied in the literature [98], necessitating the need for image pre-processing prior to radiomic feature extraction. In my dissertation project, a series of image preprocessing steps were conducted in each of the corresponding chapters **(Chapter 3.2.3, Chapter 4.2.8, and Chapter 5.2.4).** For example, isotropic voxel size resampling to $1x1x1mm^3$ was made to mitigate impacts of difference in image acquisition protocol between medical centers; re-segmentation was conducted on CT images to confine the HU to the range of (-150,180) for eliminating non-soft-tissue components (such as air cavities and bony structure); intensity normalization of MR images was conducted to obtain comparable features between subjects, etc. However, it has been found that pre-processing steps, such as signal intensity normalization and voxel size resampling [145, 146], choices of scanner and noise [147, 148], may affect feature reproducibility, potentially resulting in unexpected impacts on predictive performance of the developed radiomic model.

In particular for isotropic voxel size resampling, linear interpolation was performed in this project to rescale the voxel size to $1 \times 1 \times 1 \text{ mm}^3$ for the sake of achieving inter-institutional comparison of features extracted by images obtained under varying

protocols and reconstruction parameters. In clinical practice, it is common to obtain anisotropic voxels in most imaging modalities, with slice thickness greater than the cross-sectional resolution. Three-dimensional isotropic resampling has been suggested for establishing conservative scales in 3D and eliminating direction-related bias in calculation of 3D radiomic features [149].

Such interpolations may affect reproducibility of radiomic feature [150-152], and the degree of impact may vary depending on the categories of radiomic features. For instance, calculations of shape-based and first-order features are less relevant to spatial distribution of imaging voxel intensities, therefore the impacts may be of less extent. However, texture radiomic features, which characterizes spatial variations of voxel intensities, may in principle subject more to the interpolation methods. For example, resampling to a larger voxel size, which increases inter-voxel spatial distance, may lead to information loss; while resampling to a smaller voxel dimension, which decreases inter-voxel spatial distance, lead to generation of artificial information. Extreme resampling methods may increase the risk of generating images with poor-quality or over-smoothing. Therefore, in this dissertation project, the dimension and methods of voxel size resampling were adopted in compliance with a well-established IBSI protocol for radiomic feature calculation [98].

Still, it is acknowledged that the adoption of interpolation in this dissertation, to some degree, affect the calculation of radiomic features and possibly the performance of radiomic model. Several studies have investigated potential impact of interpolation on robustness of radiomic features [150-152], mostly on CT images. An ideal potential biomarker or model should be robust to such an imaging pre-processing step, which,

167

however, was not conducted in this dissertation project and is regarded as one of the limitations of this work.

For organ segmentation variabilities, the issue of intra-observer variability in segmentation of tumor and organs-at-risk has been recognized in the literature, which may become severer in NPC due to its highly infiltrative nature [153-155]. Such variabilities have been thought to impact the textural content of the segmented regions, hence affect the reproducibility of the extracted radiomic features.

Multiple studies have been conducted to investigate the underlying impact on feature reproducibility [148, 156-159], among which there are two most commonly employed approaches: (1) test-rest reproducibility analysis [156] and (2) multi-expert tumor segmentation variability assessment [159]. For instance, Balagurunathan et al. performed a test-retest reproducibility analysis of 32 lung cancer patients who were subjected to 2 CT scans (15 mins apart) [156]. Segmentations were made on both scans separately, followed by radiomic feature calculations from the lung tumors. Concordance correlation coefficient (CCC) across the test-retest data revealed that only 30% of the features were robust (CCC $\geq$ 0.9) [156]. On the other hand, Qiu et al. investigated impact of tumor segmentation variations on CT-based radiomic features in hepatocellular carcinoma patients by recruiting five abdominal radiation oncologists and semi-automatic algorithms for tumor segmentation [159]. They found the reproducibility of radiomic features largely depended on segmentation approaches, with only 65% (out of 71) features had ICC larger than 0.75 for all segmentation methods [159].

Although these two commonly used strategies allow assessment of feature reproducibility using ICC and/or CCC, prior to predictive model development, it presents practical challenges in that the test-retest would requires every studied subject to undergo two separate scans in two different time points, followed by separated segmentations, which was practically not achievable in this dissertation project due to the retrospective study nature. Besides, although the recruitment of multiple oncologists for organ segmentation allows assessment of impact of inter-observer segmentation variability on feature reproducibility, this resource is highly demanding and difficult to obtain due to the existing heavy workload of clinicians. Therefore, feature reproducibility against organ segmentation variabilities was not investigated and remains to be one of the limitations in this project.

Recommendations: To resolve the above-mentioned challenges, there is an ongoing trend in application of semi-automatic or fully automatic segmentation methods, hoping to potentially improve the inter- and intra-observer segmentation consistency, hence alleviating the underlying impact of feature reproducibility. For example, Liu R et al. assessed stability of CECT-based radiomic features against segmentation variabilities by using different semi-automatic segmentation algorithms in oropharyngeal cancer patients, quantified by using intra-class correlation coefficients (ICC) and concordance correlation coefficients (CCC) metrics for inter-observer reliability assessment [158]. Alternatively, Zwanenburg et al. has developed an image perturbation algorithm that aims to simulate variabilities in image pre-processing steps (such as voxel size resampling), imaging noise, and tumor segmentation for the sake of assessing robustness of radiomic features [148]. This presents an exciting moment to waive the requirement of test-retest dataset and recruitment of multiple oncologists for

169

examining impact of such variabilities on radiomic features reproducibility, and is anticipated to plan an increasing important role in the future.

(3) The limited sample size used in this project is the third limitation of this project. Ideally, a prediction model should be developed in a large-sized training cohort, followed by validation in independent external cohorts for assessing model generalizability between datasets. Indeed, a small sample size may increase both the risk of type I and type II errors, and thus increase the risk of overfitting [160]. Therefore, there exist various recommendations regarding the minimum size of dataset for safeguarding statistical power and robustness of the developed predictive models. Bibault et al. suggested that the sample size should be 5-10 times higher than the number of features investigated, and a thorough pre-selection of features is needed to determine a reasonable sample size [161]. Chalkidou et al. recommended a minimum of 10-15 patients for model training and validation if one feature finally contributes to reasonably stable estimates [100].

Despite the efforts made in this dissertation project to minimize the potential impacts, for example, by using the dedicatedly designed feature selection approaches, the nested K-fold cross-validation (CV), and bootstrapping techniques **(Chapter 3.2.5, Chapter 4.2.10-4.2.11, and Chapter 5.2.5)**, the statistical power of the developed model was still limited by the small sample size. Indeed, the small sample size involved in the project was partly due to an attempt to obtain a cohort of patients with relatively homogenous characteristics. For instance, all of the studied patients presented in **Chapter 3 and 5** were treated by concurrent chemoradiotherapy (CCRT). Patients who received radiotherapy-alone, induction chemotherapy before CCRT, or adjuvant

chemotherapy post CCRT were excluded, for the sake of minimizing potential prediction bias due to the biologic effects caused by heterogenous treatment regimens. However, this has largely reduced number of eligibility patients in these studies.

Recommendations: A larger study cohort is warranted in the future to achieve a higher statistical inference of the prediction model. Alternatively, there is emerging imbalance adjustment strategies being investigated to mitigate impact of the sample size on model performance [162]. For instance, Zhang et al. developed an imbalance adjustment framework to alleviate the impact of imbalanced events in a small sample cohort for prediction of distant metastasis in NPC patients [162]. This presents an interesting technique to be incorporated into this project in the future. Besides, whenever possible, external model validation should be conducted for assessing model generalizability to patients treated in different institutions. In future study, investigations on datasets obtained from different oncology departments are needed to further assess model generalization power.

7. Chapter 7: Conclusion

ART aims to compensate for patient-specific anatomic variations between fractions, ensure accurate and safe radiation delivery, and to maintain treatment efficacy in NPC patients. Presently, the entire procedure for evaluating an eligible ART candidate is time-consuming, resource-demanding, and highly inefficient. Therefore, routine ART implementation on patient basis would undoubtedly pose immense burden to clinic. Tremendous efforts have been made to identify dosimetric factors that may relate to anatomic variations. Nevertheless, most of these factors require close monitoring throughout the radiotherapy course for each patient, pre-treatment screening strategy is severely lacking at present. Further, these factors are deficient in capturing inter-patient disparity in intrinsic biologic response of tissue upon receiving treatment perturbation. Until more recently, emerging Radiomics has opened up opportunities for divulging concealed biologic traits and genetic association of tumor and organ structures [96, 116, 117]. There is mounting evidence in the literature showing the power of Radiomics in predicting treatment response on the ground of volume shrinkage in various cancer diseases [5, 75, 77, 78, 81, 112], which has laid great foundation for Radiomics prediction of ART demand in cancer patients. It is also worth noting that ART eligibility is multifactorial in nature. Joint response of multiple organ structures upon treatment perturbations, treatment aggressiveness, and pre-treatment geometric and morphologic condition of patient anatomy, may all come into play for triggering ART. In my thesis, I have set out a series of studies for investigating potential of a variety of prediction models using -omics features from several pertinent organ structures, including primary NPC tumor, metastatic neck lymph node lesions, bi-

lateral parotid glands, etc., in the hope of facilitating ART clinical implementation in the long run.

This study evaluated capability of CT-based neck nodal radiomic features for prediction of IfTM-triggered ART, MR-based tumoral radiomic features and a variety of multi-organ single-omics (R, D, M, C) and multi-omics models (RD, RM, RC, RDCM) for prediction of ART eligibility prior to treatment commencement. There are still many unresolved challenges for future investigations. First, the model interpretability needs to be enhanced for bench-to-bedside translation, it could be achieved by incorporating biologic, genetic or functional imaging tests (e.g., PET imaging) at the stage of study design in future works. Second, prospective large multi-center studies are warranted in the future to improve generalizability and statistical inference of the prediction models. Third, feature reproducibility against VOI delineations need to be considered in order to obtain robust features for downstream model development and evaluation. If these challenges were overcome, it adds significant value to the assessment of -omics in prediction of ART eligibility in NPC patients, paving the way toward personalized oncology.

To summarize, a series of studies in this thesis demonstrated that CT-based neck nodal radiomics was capable of predicting IfTM-triggered ART events in NPC patients undergoing RT, showing higher predictability over traditional clinical predictors. MRI-based tumoral radiomics was shown promising in pre-treatment identification of ART eligibility in NPC patients. In particular, the joint T1-T2 model outperformed both T1-w and T2-w models. Multi-organ multi-omics analyses revealed that the Radiomic model played a dominant role for ART eligibility in NPC patients and Radiomic

173

features accounted for the largest proportion of predictors in all the four multi-omics models (RD, RC, RM and RDCM), suggesting its governing power in ART eligibility prediction in NPC patients. The overall findings may provide valuable insights for future study into developing an effective screening strategy for ART eligibility in NPC patients in the long run, ultimately alleviating the workload of clinical practitioners, streamlining ART procedural efficiency in clinics, and achieving personalized RT for NPC patients in the future.

8.   Reference

1.   Lambin, P., et al., *Radiomics: extracting more information from medical images using advanced feature analysis.* Eur J Cancer, 2012. **48**(4): p. 441-6.

2.   Panth, K.M., et al., *Is there a causal relationship between genetic changes and radiomics-based image features? An in vivo preclinical experiment with doxycycline inducible GADD34 tumor cells.* Radiother Oncol, 2015. **116**(3): p. 462-6.

3.   Zhuo, E.H., et al., *Radiomics on multi-modalities MR sequences can subtype patients with non-metastatic nasopharyngeal carcinoma (NPC) into distinct survival subgroups.* Eur Radiol, 2019. **29**(10): p. 5590-5599.

4.   Dong, D., et al., *Development and validation of an individualized nomogram to identify occult peritoneal metastasis in patients with advanced gastric cancer.* Ann Oncol, 2019. **30**(3): p. 431-438.

5.   Wang, G., et al., *Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma.* Eur J Radiol, 2018. **98**: p. 100-106.

6.   Placidi, L., et al., *A Multicentre Evaluation of Dosiomics Features Reproducibility, Stability and Sensitivity.* Cancers (Basel), 2021. **13**(15).

7.   Rossi, L., et al., *Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy.* Radiother Oncol, 2018. **129**(3): p. 548-553.

8. Buizza, G., et al., *Radiomics and Dosiomics for Predicting Local Control after Carbon-Ion Radiotherapy in Skull-Base Chordoma.* Cancers (Basel), 2021. **13**(2).

9. Adachi, T., et al., *Multi-institutional dose-segmented dosiomic analysis for predicting radiation pneumonitis after lung stereotactic body radiation therapy.* Med Phys, 2021. **48**(4): p. 1781-1791.

10. Chen, Y.P., et al., *Nasopharyngeal carcinoma.* Lancet, 2019. **394**(10192): p. 64-80.

11. Ferlay J, E.M., Lam F, *Global Cancer Observatory:cancer today.* Lyon, France: International Agency for Research on Cancer, 2018.

12. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.

13. Chang, E.T. and H.O. Adami, *The enigmatic epidemiology of nasopharyngeal carcinoma.* Cancer Epidemiol Biomarkers Prev, 2006. **15**(10): p. 1765-77.

14. Nicholls, J.M., et al., *The association of squamous cell carcinomas of the nasopharynx with Epstein-Barr virus shows geographical variation reminiscent of Burkitt's lymphoma.* J Pathol, 1997. **183**(2): p. 164-8.

15. Lee, A.W.M., et al., *The battle against nasopharyngeal cancer.* Radiother Oncol, 2012. **104(3)**: p. 272–278.

16. Chang, E.T., et al., *Active and Passive Smoking and Risk of Nasopharyngeal Carcinoma: A Population-Based Case-Control Study in Southern China.* Am J Epidemiol, 2017. **185**(12): p. 1272-1280.

17.   Guo, X., et al., *Evaluation of nonviral risk factors for nasopharyngeal carcinoma in a high-risk population of Southern China.* Int J Cancer, 2009. **124**(12): p. 2942-7.

18.   Liu, Z., et al., *Quantification of familial risk of nasopharyngeal carcinoma in a high-incidence area.* Cancer, 2017. **123**(14): p. 2716-2725.

19.   Liu, Z., et al., *Oral Hygiene and Risk of Nasopharyngeal Carcinoma-A Population-Based Case-Control Study in China.* Cancer Epidemiol Biomarkers Prev, 2016. **25**(8): p. 1201-7.

20.   Tsao, S.W., et al., *Etiological factors of nasopharyngeal carcinoma.* Oral Oncol, 2014. **50**(5): p. 330-8.

21.   Pathmanathan, R., et al., *Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx. Variants of Epstein-Barr virus-infected neoplasia.* Am J Pathol, 1995. **146**(6): p. 1355-67.

22.   Wang, H.Y., et al., *A new prognostic histopathologic classification of nasopharyngeal carcinoma.* Chin J Cancer, 2016. **35**: p. 41.

23.   Young, L.S. and C.W. Dawson, *Epstein-Barr virus and nasopharyngeal carcinoma.* Chin J Cancer, 2014. **33**(12): p. 581-90.

24.   Chan, A.T., P.M. Teo, and P.J. Johnson, *Nasopharyngeal carcinoma.* Ann Oncol, 2002. **13**(7): p. 1007-15.

25.   Wei, W.I. and J.S. Sham, *Nasopharyngeal carcinoma.* Lancet, 2005. **365**(9476): p. 2041-54.

26. Co, J., M.B. Mejia, and J.M. Dizon, *Evidence on effectiveness of intensity-modulated radiotherapy versus 2-dimensional radiotherapy in the treatment of nasopharyngeal carcinoma: Meta-analysis and a systematic review of the literature.* Head Neck, 2016. **38 Suppl 1**: p. E2130-42.

27. Kam, M.K., et al., *Prospective randomized study of intensity-modulated radiotherapy on salivary gland function in early-stage nasopharyngeal carcinoma patients.* J Clin Oncol, 2007. **25**(31): p. 4873-9.

28. Peng, G., 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.* Radiother Oncol, 2012. **104**(3): p. 286-93.

29. Pow, E.H., et al., *Xerostomia and quality of life after intensity-modulated radiotherapy vs. conventional radiotherapy for early-stage nasopharyngeal carcinoma: initial report on a randomized controlled clinical trial.* Int J Radiat Oncol Biol Phys, 2006. **66**(4): p. 981-91.

30. Zhang, B., et al., *Intensity-modulated radiation therapy versus 2D-RT or 3D-CRT for the treatment of nasopharyngeal carcinoma: A systematic review and meta-analysis.* Oral Oncol, 2015. **51**(11): p. 1041-1046.

31. Ng, W.T., et al., *Clinical outcomes and patterns of failure after intensity-modulated radiotherapy for nasopharyngeal carcinoma.* Int J Radiat Oncol Biol Phys, 2011. **79**(2): p. 420-8.

32. Kam, M.K., et al., *Intensity-modulated radiotherapy in nasopharyngeal carcinoma: dosimetric advantage over conventional plans and feasibility of dose escalation.* Int J Radiat Oncol Biol Phys, 2003. **56**(1): p. 145-57.

33. Chan, A.T., et al., *Overall survival after concurrent cisplatin-radiotherapy compared with radiotherapy alone in locoregionally advanced nasopharyngeal carcinoma.* J Natl Cancer Inst, 2005. **97**(7): p. 536-9.

34. Chen, Y., et al., *Progress report of a randomized trial comparing long-term survival and late toxicity of concurrent chemoradiotherapy with adjuvant chemotherapy versus radiotherapy alone in patients with stage III to IVB nasopharyngeal carcinoma from endemic regions of China.* Cancer, 2013. **119**(12): p. 2230-8.

35. Kwong, D.L., et al., *Concurrent and adjuvant chemotherapy for nasopharyngeal carcinoma: a factorial study.* J Clin Oncol, 2004. **22**(13): p. 2643-53.

36. Lee, A.W., et al., *Randomized trial of radiotherapy plus concurrent-adjuvant chemotherapy vs radiotherapy alone for regionally advanced nasopharyngeal carcinoma.* J Natl Cancer Inst, 2010. **102**(15): p. 1188-98.

37. Lin, J.C., et al., *Phase III study of concurrent chemoradiotherapy versus radiotherapy alone for advanced nasopharyngeal carcinoma: positive effect on overall and progression-free survival.* J Clin Oncol, 2003. **21**(4): p. 631-7.

38. Wee, J., et al., *Randomized trial of radiotherapy versus concurrent chemoradiotherapy followed by adjuvant chemotherapy in patients with American Joint Committee on*

Cancer/International Union against cancer stage III and IV nasopharyngeal cancer of the endemic variety. J Clin Oncol, 2005. **23**(27): p. 6730-8.

39. Wu, X., et al., *Long-term follow-up of a phase III study comparing radiotherapy with or without weekly oxaliplatin for locoregionally advanced nasopharyngeal carcinoma.* Ann Oncol, 2013. **24**(8): p. 2131-6.

40. Wagner, J.M. and L.M. Karnitz, *Cisplatin-induced DNA damage activates replication checkpoint signaling components that differentially affect tumor cell survival.* Mol Pharmacol, 2009. **76**(1): p. 208-14.

41. Bocci, G. and R.S. Kerbel, *Pharmacokinetics of metronomic chemotherapy: a neglected but crucial aspect.* Nat Rev Clin Oncol, 2016. **13**(11): p. 659-673.

42. Chen, Y.L., M.C. Chang, and W.F. Cheng, *Metronomic chemotherapy and immunotherapy in cancer treatment.* Cancer Lett, 2017. **400**: p. 282-292.

43. Liu, Y.C., et al., *Prognostic impact of adjuvant chemotherapy in high-risk nasopharyngeal carcinoma patients.* Oral Oncol, 2017. **64**: p. 15-21.

44. Twu, C.W., et al., *Metronomic adjuvant chemotherapy improves treatment outcome in nasopharyngeal carcinoma patients with postradiation persistently detectable plasma Epstein-Barr virus deoxyribonucleic acid.* Int J Radiat Oncol Biol Phys, 2014. **89**(1): p. 21-9.

45. Cao, S.M., et al., *Neoadjuvant chemotherapy followed by concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally*

advanced nasopharyngeal carcinoma: A phase III multicentre randomised controlled trial. Eur J Cancer, 2017. **75**: p. 14-23.

46.   Frikha, M., et al., *A randomized trial of induction docetaxel-cisplatin-5FU followed by concomitant cisplatin-RT versus concomitant cisplatin-RT in nasopharyngeal carcinoma (GORTEC 2006-02).* Ann Oncol, 2018. **29**(3): p. 731-736.

47.   Li, W.F., et al., *Concurrent chemoradiotherapy with/without induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma: Long-term results of phase 3 randomized controlled trial.* Int J Cancer, 2019. **145**(1): p. 295-305.

48.   Sun, Y., et al., *Induction chemotherapy plus concurrent chemoradiotherapy versus concurrent chemoradiotherapy alone in locoregionally advanced nasopharyngeal carcinoma: a phase 3, multicentre, randomised controlled trial.* Lancet Oncol, 2016. **17**(11): p. 1509-1520.

49.   Barker, J.L., Jr., et al., *Quantification of volumetric and geometric changes occurring during fractionated radiotherapy for head-and-neck cancer using an integrated CT/linear accelerator system.* Int J Radiat Oncol Biol Phys, 2004. **59**(4): p. 960-70.

50.   Cheng, H.C., et al., *A prospective study on volumetric and dosimetric changes during intensity-modulated radiotherapy for nasopharyngeal carcinoma patients.* Radiother Oncol, 2012. **104**(3): p. 317-23.

51.   Hu, Y.C., et al., *Which nasopharyngeal cancer patients need adaptive radiotherapy?* BMC Cancer, 2018. **18**(1): p. 1234.

52.   Jin, X., et al., *A modified VMAT adaptive radiotherapy for nasopharyngeal cancer patients based on CT-CT image fusion.* Radiat Oncol, 2013. **8**: p. 277.

53.   Lee, H., et al., *Tumor Volume Reduction Rate during Adaptive Radiation Therapy as a Prognosticator for Nasopharyngeal Cancer.* Cancer Res Treat, 2016. **48**(2): p. 537-45.

54.   Tan, W., et al., *Target volume and position variations during intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma.* Onco Targets Ther, 2013. **6**: p. 1719-28.

55.   Zhao, L., et al., *The role of replanning in fractionated intensity modulated radiotherapy for nasopharyngeal carcinoma.* Radiother Oncol, 2011. **98**(1): p. 23-7.

56.   Surucu, M., et al., *Decision Trees Predicting Tumor Shrinkage for Head and Neck Cancer: Implications for Adaptive Radiotherapy.* Technol Cancer Res Treat, 2016. **15**(1): p. 139-45.

57.   Hansen, E.K., et al., *Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer.* Int J Radiat Oncol Biol Phys, 2006. **64**(2): p. 355-62.

58.   Bhide, S.A., et al., *Weekly volume and dosimetric changes during chemoradiotherapy with intensity-modulated radiation therapy for head and neck cancer: a prospective observational study.* Int J Radiat Oncol Biol Phys, 2010. **76**(5): p. 1360-8.

59.   Munshi, A., et al., *Weight loss during radiotherapy for head and neck malignancies: what factors impact it?* Nutr Cancer, 2003. **47**(2): p. 136-40.

60. Noble, D.J., et al., *Anatomical change during radiotherapy for head and neck cancer, and its effect on delivered dose to the spinal cord.* Radiother Oncol, 2019. **130**: p. 32-38.

61. Mnejja, W., et al., *Dosimetric impact on changes in target volumes during intensity-modulated radiotherapy for nasopharyngeal carcinoma.* Rep Pract Oncol Radiother, 2020. **25**(1): p. 41-45.

62. Bahl, A., et al., *Analysis of volumetric and dosimetric changes in mid treatment CT scan in carcinoma nasopharynx: implications for adaptive radiotherapy.* J Exp Ther Oncol, 2019. **13**(1): p. 33-39.

63. Chitapanarux, I., et al., *A dosimetric comparison of two-phase adaptive intensity-modulated radiotherapy for locally advanced nasopharyngeal cancer.* J Radiat Res, 2015. **56**(3): p. 529-38.

64. Deng, S., et al., *Three-Phase Adaptive Radiation Therapy for Patients With Nasopharyngeal Carcinoma Undergoing Intensity-Modulated Radiation Therapy: Dosimetric Analysis.* Technol Cancer Res Treat, 2017. **16**(6): p. 910-916.

65. Lu, J., et al., *Assessment of anatomical and dosimetric changes by a deformable registration method during the course of intensity-modulated radiotherapy for nasopharyngeal carcinoma.* J Radiat Res, 2014. **55**(1): p. 97-104.

66. Luo, Y., Y. Qin, and J. Lang, *Effect of adaptive replanning in patients with locally advanced nasopharyngeal carcinoma treated by intensity-modulated radiotherapy: a propensity score matched analysis.* Clin Transl Oncol, 2017. **19**(4): p. 470-476.

67. Wang, R.H., et al., *Volume and dosimetric variations during two-phase adaptive intensity-modulated radiotherapy for locally advanced nasopharyngeal carcinoma.* Biomed Mater Eng, 2014. **24**(1): p. 1217-25.

68. Wang, W., et al., *Clinical study of the necessity of replanning before the 25th fraction during the course of intensity-modulated radiotherapy for patients with nasopharyngeal carcinoma.* Int J Radiat Oncol Biol Phys, 2010. **77**(2): p. 617-21.

69. Yang, H., et al., *Replanning during intensity modulated radiation therapy improved quality of life in patients with nasopharyngeal carcinoma.* Int J Radiat Oncol Biol Phys, 2013. **85**(1): p. e47-54.

70. Yu, T.T., et al., *Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients.* Front Oncol, 2019. **9**: p. 1050.

71. Ouyang, F.S., et al., *Exploration and validation of radiomics signature as an independent prognostic biomarker in stage III-IVb nasopharyngeal carcinoma.* Oncotarget, 2017. **8**(43): p. 74869-74879.

72. Zhang, B., et al., *Radiomic machine-learning classifiers for prognostic biomarkers of advanced nasopharyngeal carcinoma.* Cancer Lett, 2017. **403**: p. 21-27.

73. Zhang, B., et al., *Advanced nasopharyngeal carcinoma: pre-treatment prediction of progression based on multi-parametric MRI radiomics.* Oncotarget, 2017. **8**(42): p. 72457-72465.

74. Zhang, B., et al., *Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma.* Clin Cancer Res, 2017. **23**(15): p. 4259-4269.

75. Colen, R.R., et al., *Radiomics analysis for predicting pembrolizumab response in patients with advanced rare cancers.* J Immunother Cancer, 2021. **9**(4).

76. Coroller, T.P., et al., *Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC.* J Thorac Oncol, 2017. **12**(3): p. 467-476.

77. Hou, Z., et al., *Radiomic analysis in contrast-enhanced CT: predict treatment response to chemoradiotherapy in esophageal carcinoma.* Oncotarget, 2017. **8**(61): p. 104444-104454.

78. Klaassen, R., et al., *Feasibility of CT radiomics to predict treatment response of individual liver metastases in esophagogastric cancer patients.* PLoS One, 2018. **13**(11): p. e0207362.

79. Santiago, R., et al., *CT-based radiomics model with machine learning for predicting primary treatment failure in diffuse large B-cell Lymphoma.* Transl Oncol, 2021. **14**(10): p. 101188.

80. Trebeschi, S., et al., *Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers.* Ann Oncol, 2019. **30**(6): p. 998-1004.

81. Yongfeng, P., et al., *The Usefulness of Pretreatment MR-Based Radiomics on Early Response of Neoadjuvant Chemotherapy in Patients With Locally Advanced Nasopharyngeal Carcinoma.* Oncol Res, 2021. **28**(6): p. 605-613.

82.	Wang, X., et al., *Patterns of lymph node metastasis from nasopharyngeal carcinoma based on the 2013 updated consensus guidelines for neck node levels.* Radiother Oncol, 2015. **115**(1): p. 41-5.

83.	Lang, J., et al., *Comprehensive treatment of squamous cell cancer of head and neck: Chinese expert consensus 2013.* Future Oncol, 2014. **10**(9): p. 1635-48.

84.	Li, M., et al., *Effects of omitting elective neck irradiation to nodal Level IB in nasopharyngeal carcinoma patients with negative Level IB lymph nodes treated by intensity-modulated radiotherapy: a Phase 2 study.* Br J Radiol, 2016. **89**(1065): p. 20150621.

85.	Lin, C.G., et al., *Comparison of set up accuracy among three common immobilisation systems for intensity modulated radiotherapy of nasopharyngeal carcinoma patients.* J Med Radiat Sci, 2017. **64**(2): p. 106-113.

86.	Brouwer, C.L., et al., *Identifying patients who may benefit from adaptive radiotherapy: Does the literature on anatomic and dosimetric changes in head and neck organs at risk during radiotherapy provide information to help?* Radiother Oncol, 2015. **115**(3): p. 285-94.

87.	Castadot, P., et al., *Assessment by a deformable registration method of the volumetric and positional changes of target volumes and organs at risk in pharyngo-laryngeal tumors treated with concomitant chemo-radiation.* Radiother Oncol, 2010. **95**(2): p. 209-17.

88. Castelli, J., et al., *Adaptive radiotherapy for head and neck cancer.* Acta Oncol, 2018. **57**(10): p. 1284-1292.

89. Hamming-Vrieze, O., et al., *Evaluation of tumor shape variability in head-and-neck cancer patients over the course of radiation therapy using implanted gold markers.* Int J Radiat Oncol Biol Phys, 2012. **84**(2): p. e201-7.

90. Morgan, H.E. and D.J. Sher, *Adaptive radiotherapy for head and neck cancer.* Cancers Head Neck, 2020. **5**: p. 1.

91. Hansen, C.R., et al., *Comparison of three immobilisation systems for radiation therapy in head and neck cancer.* Acta Oncol, 2014. **53**(3): p. 423-7.

92. Neubauer, E., et al., *Assessment of shoulder position variation and its impact on IMRT and VMAT doses for head and neck cancer.* Radiat Oncol, 2012. **7**: p. 19.

93. Su, J., et al., *Different setup errors assessed by weekly cone-beam computed tomography on different registration in nasopharyngeal carcinoma treated with intensity-modulated radiation therapy.* Onco Targets Ther, 2015. **8**: p. 2545-53.

94. Fung, W.W., V.W. Wu, and P.M. Teo, *Developing an adaptive radiation therapy strategy for nasopharyngeal carcinoma.* J Radiat Res, 2014. **55**(2): p. 293-304.

95. Aerts, H.J., et al., *Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.* Nat Commun, 2014. **5**: p. 4006.

96. Gillies, R.J., P.E. Kinahan, and H. Hricak, *Radiomics: Images Are More than Pictures, They Are Data.* Radiology, 2016. **278**(2): p. 563-77.

97. Pinker, K., et al., *Precision Medicine and Radiogenomics in Breast Cancer: New Approaches toward Diagnosis and Treatment.* Radiology, 2018. **287**(3): p. 732-747.

98. Zwanenburg, A., et al., *The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping.* Radiology, 2020. **295**(2): p. 328-338.

99. Bourbonne, V., et al., *Radiomics analysis of 3D dose distributions to predict toxicity of radiotherapy for lung cancer.* Radiother Oncol, 2021. **155**: p. 144-150.

100. Chalkidou, A., M.J. O'Doherty, and P.K. Marsden, *False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review.* PLoS One, 2015. **10**(5): p. e0124165.

101. Zhang, L.L., et al., *Pretreatment MRI radiomics analysis allows for reliable prediction of local recurrence in non-metastatic T4 nasopharyngeal carcinoma.* EBioMedicine, 2019. **42**: p. 270-280.

102. Tsamardinos, I., E. Greasidou, and G. Borboudakis, *Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation.* Mach Learn, 2018. **107**(12): p. 1895-1922.

103. Ramspek, C.L., et al., *External validation of prognostic models: what, why, how, when and where?* Clinical kidney journal, 2020. **14(1)** p. 49–58.

104. Ng, W.T., et al., *Head and neck cancer in Hong Kong.* Jpn J Clin Oncol, 2018. **48**(1): p. 13-21.

105. Chen, A.M., et al., *Clinical outcomes among patients with head and neck cancer treated by intensity-modulated radiotherapy with and without adaptive replanning.* Head Neck, 2014. **36**(11): p. 1541-6.

106. Schwartz, D.L., et al., *Adaptive radiotherapy for head and neck cancer--dosimetric results from a prospective clinical trial.* Radiother Oncol, 2013. **106**(1): p. 80-4.

107. Schwartz, D.L., et al., *Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial.* Int J Radiat Oncol Biol Phys, 2012. **83**(3): p. 986-93.

108. Brown, E., et al., *Predicting the need for adaptive radiotherapy in head and neck cancer.* Radiother Oncol, 2015. **116**(1): p. 57-63.

109. Wu, Q., et al., *Adaptive replanning strategies accounting for shrinkage in head and neck IMRT.* Int J Radiat Oncol Biol Phys, 2009. **75**(3): p. 924-32.

110. Lambin, P., et al., *Radiomics: the bridge between medical imaging and personalized medicine.* Nat Rev Clin Oncol, 2017. **14**(12): p. 749-762.

111. Xu, C.J., et al., *Statistical validation of normal tissue complication probability models.* Int J Radiat Oncol Biol Phys, 2012. **84**(1): p. e123-9.

112. Ramella, S., et al., *A radiomic approach for adaptive radiotherapy in non-small cell lung cancer patients.* PLoS One, 2018. **13**(11): p. e0207455.

113. Du, R., et al., *Radiomics Model to Predict Early Progression of Nonmetastatic Nasopharyngeal Carcinoma after Intensity Modulation Radiation Therapy: A Multicenter Study.* Radiol Artif Intell, 2019. **1**(4): p. e180075.

114. Gai, X., et al., *Clinical study of the time of repeated computed tomography and replanning for patients with nasopharyngeal carcinoma.* Oncotarget, 2017. **8**(16): p. 27529-27540.

115. Yao, W.R., et al., *Replanning Criteria and Timing Definition for Parotid Protection-Based Adaptive Radiation Therapy in Nasopharyngeal Carcinoma.* Biomed Res Int, 2015. **2015**: p. 476383.

116. Sun, R., et al., *A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study.* Lancet Oncol, 2018. **19**(9): p. 1180-1191.

117. Xing, L., E.A. Krupinski, and J. Cai, *Artificial intelligence will soon change the landscape of medical physics research and practice.* Med Phys, 2018. **45**(5): p. 1791-1793.

118. Gregoire, V., et al., *CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC,RTOG consensus guidelines.* Radiother Oncol, 2003. **69**(3): p. 227-36.

119. Lee, S.H., et al., *Multi-view radiomics and dosiomics analysis with machine learning for predicting acute-phase weight loss in lung cancer patients treated with radiotherapy.* Phys Med Biol, 2020. **65**(19): p. 195015.

120. Wu, A., et al., *Dosiomics improves prediction of locoregional recurrence for intensity modulated radiotherapy treated head and neck cancer cases.* Oral Oncol, 2020. **104**: p. 104625.

121.   Gabrys, H.S., et al., *Design and Selection of Machine Learning Methods Using Radiomics and Dosiomics for Normal Tissue Complication Probability Modeling of Xerostomia.* Front Oncol, 2018. **8**: p. 35.

122.   Buettner, F., 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.* Radiother Oncol, 2012. **103**(1): p. 82-7.

123.   Wu, B., et al., *Patient geometry-driven information retrieval for IMRT treatment plan quality control.* Med Phys, 2009. **36**(12): p. 5497-505.

124.   Li J, C.K., et al. , *Feature Selection: A Data Perspective. .* ACM Comput Surv 2017. **50**(6): p. 1–45.

125.   Li, Z. and J. Tang, *Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control.* IEEE Trans Image Process, 2015. **24**(12): p. 5343-55.

126.   Robnik-Šikonja M, K.I., *Theoretical and Empirical Analysis of ReliefF and RReliefF.* Mach Learn, 2003. **53**(1/2): p. 23-69.

127.   S., W., *The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating.* Evolution, 1965. **19**(3): p. 395-420.

128.   Valizade Hasanloei, M.A., et al., *A combined Fisher and Laplacian score for feature selection in QSAR based drug design using compounds with known and unknown activities.* J Comput Aided Mol Des, 2018. **32**(2): p. 375-384.

129.   Tadist K., et al., *Feature Selection Methods and Genomic Big Data: A Systematic Review.* J Big Data 2019. **6**(1).

130. Arthur E. Hoerl, R.W.K., *Ridge Regression — 1980: Advances, Algorithms, and Applications.* American Journal of Mathematical and Management Sciences, 1981. **1**(1): p. 5-83.

131. Rakotomamonjy A, B.F., Canu S, Grandvalet Y. , *SimpleMKL.* J Mach Learn Res 2008. **9**: p. 2491-521.

132. Hinrichs, C., et al., *Q-MKL: Matrix-induced Regularization in Multi-Kernel Learning with Applications to Neuroimaging.* Adv Neural Inf Process Syst, 2012. **2012**: p. 1430-1438.

133. Hu, E., et al., *Semisupervised kernel matrix learning by kernel propagation.* IEEE Trans Neural Netw, 2010. **21**(11): p. 1831-41.

134. Lanckriet, G.R., et al., *A statistical framework for genomic data fusion.* Bioinformatics, 2004. **20**(16): p. 2626-35.

135. Deng, Z., et al., *Generalized hidden-mapping ridge regression, knowledge-leveraged inductive transfer learning for neural networks, fuzzy systems and kernel methods.* IEEE Trans Cybern, 2014. **44**(12): p. 2585-99.

136. Liang, B., et al., *Prediction of Radiation Pneumonitis With Dose Distribution: A Convolutional Neural Network (CNN) Based Model.* Front Oncol, 2019. **9**: p. 1500.

137. Sheikh, K., et al., *Predicting acute radiation induced xerostomia in head and neck Cancer using MR and CT Radiomics of parotid and submandibular glands.* Radiat Oncol, 2019. **14**(1): p. 131.

192

138. Jiang, W., et al., *Dosimetric Factors and Radiomics Features Within Different Regions of Interest in Planning CT Images for Improving the Prediction of Radiation Pneumonitis.* Int J Radiat Oncol Biol Phys, 2021. **110**(4): p. 1161-1170.

139. Perez-Morales, J., et al., *Peritumoral and intratumoral radiomic features predict survival outcomes among patients diagnosed in lung cancer screening.* Sci Rep, 2020. **10**(1): p. 10528.

140. Tunali, I., et al., *Novel clinical and radiomic predictors of rapid disease progression phenotypes among lung cancer patients treated with immunotherapy: An early report.* Lung Cancer, 2019. **129**: p. 75-79.

141. Tunali, I., et al., *Hypoxia-Related Radiomics and Immunotherapy Response: A Multicohort Study of Non-Small Cell Lung Cancer.* JNCI Cancer Spectr, 2021. **5**(4).

142. Gevaert, O., et al., *Non-small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data--methods and preliminary results.* Radiology, 2012. **264**(2): p. 387-96.

143. Rios Velazquez, E., et al., *Somatic Mutations Drive Distinct Imaging Phenotypes in Lung Cancer.* Cancer Res, 2017. **77**(14): p. 3922-3930.

144. Tomaszewski, M.R. and R.J. Gillies, *The Biological Meaning of Radiomic Features.* Radiology, 2021. **299**(2): p. E256.

145. Mottola, M., et al., *Reproducibility of CT-based radiomic features against image resampling and perturbations for tumour and healthy kidney in renal cancer patients.* Sci Rep, 2021. **11**(1): p. 11542.

146. Whybra, P., et al., *Assessing radiomic feature robustness to interpolation in (18)F-FDG PET imaging.* Sci Rep, 2019. **9**(1): p. 9649.

147. He, L., et al., *Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule.* Sci Rep, 2016. **6**: p. 34921.

148. Zwanenburg, A., et al., *Assessing robustness of radiomic features by image perturbation.* Sci Rep, 2019. **9**(1): p. 614.

149. Depeursinge, A., et al., *Three-dimensional solid texture analysis in biomedical imaging: review and opportunities.* Med Image Anal, 2014. **18**(1): p. 176-96.

150. Mackin, D., et al., *Harmonizing the pixel size in retrospective computed tomography radiomics studies.* PLoS One, 2017. **12**(9): p. e0178524.

151. Shafiq-Ul-Hassan, M., et al., *Voxel size and gray level normalization of CT radiomic features in lung cancer.* Sci Rep, 2018. **8**(1): p. 10545.

152. Shafiq-Ul-Hassan, M., et al., *Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels.* Med Phys, 2017. **44**(3): p. 1050-1062.

153. Breen, S.L., et al., *Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers.* Int J Radiat Oncol Biol Phys, 2007. **68**(3): p. 763-70.

154. Jiang, J., et al., *Variability of Gross Tumor Volume in Nasopharyngeal Carcinoma Using 11C-Choline and 18F-FDG PET/CT.* PLoS One, 2015. **10**(7): p. e0131801.

155. Yamazaki, H., et al., *Quantitative assessment of inter-observer variability in target volume delineation on stereotactic radiotherapy treatment for pituitary adenoma and meningioma near optic tract.* Radiat Oncol, 2011. **6**: p. 10.

156. Balagurunathan, Y., et al., *Test-retest reproducibility analysis of lung CT image features.* J Digit Imaging, 2014. **27**(6): p. 805-23.

157. Haniff, N.S.M., et al., *Stability and Reproducibility of Radiomic Features Based Various Segmentation Technique on MR Images of Hepatocellular Carcinoma (HCC).* Diagnostics (Basel), 2021. **11**(9).

158. Liu, R., et al., *Stability analysis of CT radiomic features with respect to segmentation variation in oropharyngeal cancer.* Clin Transl Radiat Oncol, 2020. **21**: p. 11-18.

159. Qiu, Q., et al., *Reproducibility and non-redundancy of radiomic features extracted from arterial phase CT scans in hepatocellular carcinoma patients: impact of tumor segmentation variability.* Quant Imaging Med Surg, 2019. **9**(3): p. 453-464.

160. Vabalas, A., et al., *Machine learning algorithm validation with a limited sample size.* PLoS One, 2019. **14**(11): p. e0224365.

161. Bibault, J.E., P. Giraud, and A. Burgun, *Big Data and machine learning in radiation oncology: State of the art and future prospects.* Cancer Lett, 2016. **382**(1): p. 110-117.

162. Zhang, Y.P., et al., *Integration of an imbalance framework with novel high-generalizable classifiers for radiomics-based distant metastases prediction of advanced nasopharyngeal carcinoma.* Knowledge-Based Systems, 2022. **235**.