AUTOMATIC TREATMENT PLANNING FOR FUNCTIONAL LUNG AVOIDANCE RADIOTHERAPY

XIONG TIANYU

PhD

The Hong Kong Polytechnic University

2026

The Hong Kong Polytechnic University

Department of Health Technology and Informatics

Automatic Treatment Planning for Functional Lung Avoidance Radiotherapy

XIONG Tianyu

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

January 2026

Certificate of Originality

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

_____(Signed)

Xiong Tianyu (Name of Student)

Abstract

Background: Functional Lung Avoidance Radiotherapy (FLART) is a promising radiotherapy (RT) technique that may reduce pulmonary toxicity compared to conventional lung RT. Unlike conventional anatomy-guided RT, FLART features in utilizing functional images to account for lung function heterogeneity and preferentially sparing high-function lung (HFL) regions for better lung function protection. The incorporation of lung function information into treatment planning is crucial for FLART. However, the majority of current FLART studies rely on manual planning. The significant inter-patient variability and intra-lung heterogeneity of lung function distributions, coupled with limited clinical experience as an emerging RT technique, render manual FLART planning highly challenging. Therefore, manual FLART planning is not only time-consuming and labor-intensive but also prone to producing inconsistent or suboptimal FLART plans.

Purpose: This study aims to develop a novel multi-modality-guided dose prediction (MMDP)-based automatic planning (auto-planning) algorithm for FLART and evaluate its clinical performance and potential benefits.

Method: The study is structured in three parts. First, we developed a fully automatic planning algorithm for FLART. It integrates a dosimetric score-based beam angle selection method and a meta-optimization (MO)-based plan hyperparameter optimization method, both of which incorporate lung function information to guide dose redirection from HFL to low-functional lung (LFL). It is applicable to both contour-based FLART (cFLART) and voxel-based FLART (vFLART) optimization

options. A cohort of 18 lung cancer patients underwent planning CT and SPECT perfusion (Q) scans were collected to preliminarily evaluate the performance of the MO-based auto-planning algorithm. In the second part, we further developed an MMDP-based auto-planning algorithm for FLART, building on the MO-based method. We collected data from 196 lung cancer patients across three institutions, comprising 114/28 cases with lung ventilation (V) surrogate maps for training/validation and 21/33 cases with SPECT V/Q images for testing. The MO-based auto-planning algorithm was employed to generate ConvRT (MO-ConvRT) and FLART (MO-FLART) plans. High-quality plans were selected to train a novel MMDP model which features in extracting complementary features from multi-modality inputs for predicting optimal dose distributions. An innovative instance-weighting anatomy-to-function training strategy was tailored to enhance prediction accuracy. A function-guided dose mimicking algorithm was developed to convert predicted dose distributions into FLART (MMDP-FLART) plans, which were compared against MO-based and manual plans. In the third part, the proposed auto-planning algorithms were implemented in clinical treatment planning systems (TPS) and their clinical performance and benefits were evaluated. An additional 83 pairs of high-quality manual ConvRT/FLART plans were created for fine-tuning (50 pairs) and evaluation (33 pairs) of the auto-planning algorithms. Clinical performance was assessed using quantitative DVH/DFH metrics, normal tissue complication probabilities, and qualitative blind review by one senior medical physicist and two senior radiation oncologists. Additionally, the FLART planning results by a junior planner with and

without the assistance of the developed MMDP-FLART system were compared, aiming at evaluating the potential of the system in facilitating the broader clinical adoption of the emerging FLART technique.

Results: In the first part, automatic ConvRT plans generated by the MO-based approach exhibited similar quality compared to manual counterparts. Furthermore, compared to automatic ConvRT plans, HFL mean dose, $V_{20,}$ and V_5 were significantly reduced by 1.13 Gy ($p<.001$), 2.01% ($p<.001$), and 6.66% ($p<.001$) respectively for cFLART plans. The vFLART plans showed a decrease in functionally weighted mean lung dose (fMLD) by 0.64 Gy ($p<.01$), F_{20} by 0.90% ($p=0.099$), and F_5 by 5.07% ($p<.01$) respectively. Though inferior conformity was observed, all dose constraints were well satisfied. The ablation study results indicated that both function-guided beam angle selection and plan optimization significantly contributed to dose redirection. In the second part, the MMDP model achieved accurate dose predictions, with DVH score of 1.94 Gy, Dose score of 1.14 Gy, and prediction errors for fMLD of 0.10±0.57 Gy. In contrast, MMDP trained through naïve training and an anatomy-guided dose prediction model significantly overestimated ($p<.001$) fMLD by 0.40±0.58 Gy and 0.30±0.60 Gy respectively, validating the effectiveness of the proposed training strategy and multi-modality learning. Compared to manual FLART plans, MMDP-FLART plans exhibited significantly lower ($p<.001$) and comparable ($p=0.099$) fMLD on SPECT V and Q datasets respectively. Compared to manual ConvRT plans, MMDP-FLART plans effectively achieved lung dose painting and reduced fMLD by 0.80 Gy (11.9%, $p<.001$) and 0.46 Gy (6.0%, $p<.001$) on SPECT V

and Q datasets respectively. In the third part, both MO-based and MMDP-based auto-planning algorithms were integrated into clinical TPS. Compared to manual ConvRT plans, MMDP FLART plans reduced fMLD and HFL D_{mean} by 11.8% and 15.1%, respectively. For FLART-benefiting patients, MMDP FLART plans reduced the probability of grade ≥2 radiation pneumonitis by 6.25 percentage points (27%). The MMDP-based algorithm demonstrated superior lung function protection and clinical acceptability compared to the MO-based approach. Blind review by three senior clinicians revealed that 88% of MMDP FLART plans were clinically acceptable without modification, and 69% were rated as comparable or superior to manual FLART plans created by an eight-year-experience senior planner. In terms of planning efficiency, manual FLART planning required approximately 120–180 minutes per plan, whereas MO-based and MMDP-based auto-planning required approximately 78 minutes and 8 minutes per plan, respectively. Furthermore, assistance from the auto-planning system substantially enhanced the quality of FLART plans generated by a junior planner (with one year of planning experience), elevating the plan quality to a level comparable to that of a senior planner (with eight years of planning experience).

Conclusion: An innovative auto-planning algorithm for FLART based on multi-modality-guided dose prediction has been technically developed, clinically deployed, and comprehensively evaluated. It shows considerable promise in enhancing the efficiency, consistency, and quality of FLART planning, while also reducing radiation-induced lung toxicity compared to the current clinical standard. A demo for the clinically ready auto-planning system is available at the link.

Publications

Journal Articles (Published)

[1] **Tianyu Xiong#**, Guangping Zeng#, Zhi Chen, Yu-Hua Huang, Bing Li, Zongrui Ma, Dejun Zhou, Yang Sheng, Ge Ren, Qingrong Jackie Wu, Hong Ge, Jing Cai. Automatic lung dose painting for functional lung avoidance radiotherapy through multi-modality-guided dose prediction. *Physics in Medicine & Biology*. 2026. doi:10.1088/1361-6560/ae31c9 (**Early Career Researcher Focus Collection**) (IF:3.4; JCR Q1)

[2] **Tianyu Xiong**, Ge Ren, Zhi Chen, Yu-Hua Huang, Zongrui Ma, Zihan Li, Yang Sheng, Qingrong Jackie Wu, Jing Cai. A generalizable dose prediction model for automatic radiotherapy planning based on physics-informed priors and large-kernel convolutions. *Medical Physics*. 2026;53(1):e70272. doi: 10.1002/mp.70272 (IF:3.2; JCR Q1)

[3] **Tianyu Xiong**, Guangping Zeng, Zhi Chen, Yu-Hua Huang, Bing Li, Dejun Zhou, Xi Liu, Yang Sheng, Ge Ren, Qingrong Jackie Wu, Hong Ge, Jing Cai. Automatic planning for functional lung avoidance radiotherapy based on function-guided beam angle selection and plan optimization. *Physics in Medicine & Biology*. 2024;69(15):155007. doi:10.1088/1361-6560/ad5ef5 (IF:3.4 JCR Q1)

[4] Yu-Hua Huang, Zihan Li, **Tianyu Xiong**, Zhi Chen, Bing Li, Zhaoyang Lou, Yanjing Dong, Xinzhi Teng, Zongrui Ma, Hong Ge, Jing Cai. Constructing Surrogate Lung Ventilation Maps from 4DCT-derived Subregional Respiratory Dynamics. *International Journal of Radiation Oncology*Biology*Physics*. 2024.

doi: 10.1016/j.ijrobp.2024.11.074 (IF:6.4 JCR Q1)

[5] Zhi Chen, **Tianyu Xiong**, Yu-Hua Huang, Bing Li, Zongrui Ma, Hong Ge, Ge Ren, Jing Cai. Functional lung avoidance proton radiation therapy planning using anatomy-wise computed tomography-derived lung ventilation imaging. *Radiation Oncology*. 2026. doi:10.1186/s13014-026-02804-1. (IF:3.2 JCR Q1)

[6] Zhi Chen, Zihan Li, Yu-Hua Huang, Xinzhi Teng, Jiang Zhang, **Tianyu Xiong**, Yanjing Dong, Liming Song, Ge Ren, and Jing Cai. Anatomy-wise lung ventilation imaging for precise functional lung avoidance radiation therapy. *Physics in Medicine & Biology*. 2025;70: 045019. doi:10.1088/1361-6560/adb123 (IF:3.4 JCR Q1)

[7] Zhi Chen, Zihan Li, Yu-Hua Huang, **Tianyu Xiong**, Xinzhi Teng, Bing Li, John Kipritidis, Paul J. Keall, Joseph M. Reinhardt, Hong Ge, Jing Cai. A Review of Advances in Lung Function Imaging and Its Applications for Functional Lung Avoidance in Radiation Therapy. *Seminars in Nuclear Medicine*. 2025. doi: 10.1053/j.semnuclmed.2025.09.002 (IF:5.9 JCR Q1)

[8] Mayang Zhao, Tao Peng, Zhi Chen, **Tianyu Xiong**, Bing Li, Xiaoli Zheng, Yuhua Huang, Liming Song, Yao Pu, Zihan Li, Jing Cai and Ge Ren Functional-based multi-omics early prediction of radiation pneumonitis in NSCLC using AI-generated perfusion and ventilation from planning CT. *Physics in Medicine & Biology*. 2026;71(6):065013. doi: 10.1088/1361-6560/ae5209. (IF:3.4 JCR Q1)

[9] Meng Wang, Xi Liu, Haoze Li, Meixin Zhao, **Tianyu Xiong**, David Huang, Jing Cai, Weifang Zhang, Li-Sheng Geng and Ruijie Yang. 2026. Simultaneous

Synthesis of Perfusion and Ventilation Images from Ct Using a Dual-Decoder Residual Attention Network for Lung Disease Diagnosis. *Journal of Applied Clinical Medical Physics*. 27(3):e70498. doi: 10.1002/acm2.70498. (IF:2.2 JCR Q2)

Journal Articles (Unpublished)

[1] **Tianyu Xiong**, Zhi Chen, Chengcheng Fan, Chunyu He, Bing Li, Guangping Zeng, Vincent W. S. Leung, Yu-Hua Huang, Zongrui Ma, Ge Ren, Yang Sheng, Qingrong Jackie Wu, Hong Ge, Jing Cai. Clinical implementation and evaluation of an artificial intelligence-driven one-click automatic planning system for functional lung avoidance radiotherapy. (**Under Review**)

Conference Abstracts

[1] **Tianyu Xiong**, Guangping Zeng, Zhi Chen, Yu-Hua Huang, Bing Li, Dejun Zhou, Yang Sheng, Ge Ren, Qingrong Jackie Wu, Hong Ge, Jing Cai. Multi-modality learning-based dose prediction and automatic lung dose painting for functional lung avoidance radiotherapy. European SocieTy for Radiotherapy and Oncology (ESTRO), 2025. (**Mini-Oral**)

[2] **Tianyu Xiong**, Ge Ren, Zhi Chen, Yu-Hua Huang, Zihan Li, Riqiang Gao, Ali Kamen, Yang Sheng, Qingrong Jackie Wu, Jing Cai. GDP-Hmm 2nd Place Finisher. American Association of Physicists in Medicine (AAPM) Annual Meeting, 2025. (**Oral**)

[3] **Tianyu Xiong**, Guangping Zeng, Bing Li, Zhi Chen, Yu-Hua Huang, Dejun Zhou, Ge Ren, Hong Ge, Jing Cai. 基于迁移学习与多模态融合的功能肺回避放疗剂量预测(Dose Prediction for Functional Lung Avoidance Radiotherapy Based on Transfer Learning and Multi-Modality Fusion). Chinese Society of Medical Physics (CSMP) Annual Meeting, 2024. (**Oral, Excellent Student Presentation Award 3rd Prize**)

[4] **Tianyu Xiong**, Guangping Zeng, Zhi Chen, Yu-Hua Huang, Bing Li, Xi Liu, Yang Sheng, Ge Ren, Qingrong Jackie Wu, Hong Ge, Jing Cai. Automatic Planning for Functional Lung Avoidance Radiotherapy Based on Function-Guided Dose Redirection. AAPM Annual Meeting, 2024. (**Oral**)

Patents

[1] Jing Cai, **Tianyu Xiong**. Function-Guided Radiation Therapy Plan Generation Method, Apparatus, Electronic Device and Program Product (功能引导放疗计划生成方法、装置、电子设备及程序产品). Chinese Invention Patent: 202510316395.4

[2] Jing Cai, **Tianyu Xiong**, Ge Ren. FUNCTION-GUIDED RADIATION THERAPY PLAN GENERATION METHOD, APPARATUS, ELECTRONIC DEVICE AND PROGRAM PRODUCT. (US Patent Application No: 19/409,891)

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor, Prof. Jing Cai, for his invaluable guidance, support, and encouragement throughout my doctoral journey. His expertise, patience, and insightful feedback have been instrumental in shaping this thesis. Prof. Cai's passion for research and profound insights has been a continual source of inspiration, motivating me to strive for excellence in my work. He is, and will continue to be, my role model in both research and work.

I am also sincerely grateful to Prof. Qingrong Jackie Wu, Dr. Ge Ren, Dr. Tian Li, Dr. Yang Sheng, Prof. Liang-Ting Lin, Prof. Yevgeniy Vinogradskiy, and Prof. Xiangyang Tang for their constructive comments and suggestions, which have greatly improved the quality of this work.

I would also like to thank Prof. Hong Ge and Dr. Bing Li for their valuable support and assistance throughout this research project.

My heartfelt thanks go to Mr. Guangping Zeng, Dr. Zhi Chen, Mr. Jiaxing Han, Dr. Chunyu He, and Dr. Chengcheng Fan for their clinical expertise and generous support in clinical evaluation for our developed auto-planning system, which have greatly enhanced this study.

I am deeply appreciative of my group members, Dr. Zhi Chen, Dr. Yu-Hua Huang, Dr. Dejun Zhou, and Dr. Yanjing Dong, for their stimulating discussions and unwavering support. Special thanks also go to Dr. Chenyang Liu, Dr. Fan Zhang, Dr. Lu Wang, Dr. Xinzhi Teng, Dr. Jiang Zhang, Dr. Zongrui Ma, Mr. Xiang Wang, Mr. Yufeng Jiang, Mr. Jiarui Zhu, Ms. Xinyu Zhang, Ms. Qingpei Lai, Ms. Anran Liu, Mr. Jin Cao, and Mr. Jiaming Wu, among others, for fostering a collaborative and friendly environment that has made my research experience truly enjoyable.

I would like to express my appreciation to the staff of the Department of Health Technology and Informatics for their assistance and support.

I gratefully acknowledge the support provided by the Hong Kong Polytechnic University and the Department of Health Technology and Informatics, which made this research possible.

On a personal note, I am deeply thankful to my parents, family members, and friends for their unconditional love, patience, and encouragement. Their belief in me has been a constant source of motivation.

Finally, I would like to thank everyone who has contributed, directly or indirectly, to the completion of this thesis. Your support has been invaluable, and I am sincerely grateful.

Content

List of Figures

List of Tables

Abbreviations

3D	Three dimensional
4D	Four dimensional
AP	Automatic planning
CI	Conformity index
ConvRT	Conventional lung radiotherapy
CT	Computed tomography
CTPI	CT-derived perfusion imaging
CTVI	CT-derived ventilation imaging
DFH	Dose-function histogram
DIR	Deformable image registration
DL	Deep learning
DTH	Distance-to-target histogram
DTPA	Diethylene triamine penta-acetic
DVH	Dose-volume histogram
FLART	Functional lung avoidance radiotherapy
FMO	Fluence map optimization
FWL	Functionally weighted lung
GAN	Generative adversarial network
GBCA	Gadolinium-based contrast agents
HFL	High-function lung
HI	Homogeneity index
HU	Hounsfield unit
IMRT	Intensity modulated radiation therapy
IPOPT	Interior point optimizer

KBP	Knowledge-based planning
LFL	Low-function lung
MAA	Macro aggregated albumin
MAE	Mean absolute error
MCO	Multi-criteria optimization
MMDP	Multi-modality-guided dose prediction
MO	Meta-optimization
MRI	Magnetic resonance imaging
NM	Nuclear medicine
NSCLC	Non-small cell lung cancer
NT	Normal tissue
NTCP	Normal tissue complication probability
OAR	Organ at risk
OVH	Overlap volume histogram
PB	Pencil beam
PET	Positron emission tomography
PTV	Planning target volume
RILI	Radiation-induced lung injury
RL	Reinforcement learning
RT	Radiotherapy
SCLC	Small cell lung cancer
SPECT	Single-photon emission computed tomography
TPS	Treatment planning system
VMAT	Volumetric modulated arc therapy

1 Introduction

1.1 Lung cancer

Lung cancer is the most frequently diagnosed cancer and the leading cause of cancer-related mortality globally, responsible for approximately 2.5 million new cases (12.4% of all cancers) and 1.8 million deaths (18.7%) globally in 2022 [1]. Lung cancer can be classified into two major categories based on its origin. Primary lung cancer refers to malignancies that originate within the lung tissue, whereas lung metastases arise when tumor cells from other parts of the body migrate to the lungs. Among the various types of lung cancer, non-small cell lung cancer (NSCLC) is the most prevalent, accounting for approximately 80% – 85% of all cases [2]. Based on histological features, NSCLC is further subdivided into three main subtypes: adenocarcinoma, squamous cell carcinoma, and large-cell carcinoma. By contrast, small cell lung cancer (SCLC) accounts for 10%–15% of lung cancer cases and is characterized by its aggressive nature, with rapid growth and a high propensity for metastasis. Due to these characteristics, SCLC is often more responsive to chemotherapy and radiotherapy (RT); however, it is associated with a poorer overall prognosis compared to NSCLC [3].

1.2 Lung cancer radiotherapy

RT is a cornerstone of lung cancer management, with evidence-based indications for 77% of lung cancer patients. RT can be utilized for both curative and palliative purposes across all stages of lung cancer. For early-stage (stage I–II) NSCLC, RT is

considered one of the primary treatment modalities, particularly for patients who are either medically inoperable or unwilling to undergo surgery, as recommended by the National Comprehensive Cancer Network guidelines [4]. For stage III or more advanced disease, a combination of RT and chemotherapy is commonly employed. RT is also an effective treatment for patients with SCLC, both in the limited and extensive stages [5].

The administration of RT for the treatment of lung cancer presents several significant challenges [6]. On the one hand, there are numerous nearby organs at risk (OARs), including the esophagus, heart, spinal cord and surrounding healthy lung tissue, which are highly radiosensitive and play a critical role in patient survival. One the other hand, RT utilizes particles irradiated by treatment machines such as linear accelerator to damage tumor cells for cancer treatment. However, it inevitably involves unwanted radiation doses to healthy OARs such as healthy lung tissue due to the physical characteristics of particles, the physical constraints of treatment machines, the intra-fraction and inter-fraction variations of anatomical structures, the setup errors, and so on. The unwanted irradiation may damage the radiosensitive OARs and result in significant adverse effects such as radiation-induced lung injury (RILI), deteriorating patient prognosis and quality of life [6].

1.3 Lung cancer radiotherapy treatment planning

Therefore, the key to RT is to deliver high and homogeneous enough doses to tumors

while minimizing doses to healthy organs (referred to conventional dose objectives). To achieve this goal, physicians meticulously design treatment plans that can guide treatment machines to deliver clinically preferable dose distributions to patients. The planning process typically involves empirically selecting beam angles that can irradiate tumors while minimizing penetrating OARs, and iteratively setting and adjusting plan hyperparameters based on evaluation for intermediate treatment plans following established dosimetric objectives for tumors and OARs [7, 8].

Current treatment planning dosimetric objectives for lung derive from the assumption that lung function is homogeneous in spatial distribution and lung tissue is homogeneous in its response to radiation dose. However, for lung cancer patients, lung function distributions are spatially inhomogeneous due to the presence of tumors or thoracic comorbidities, which has been validated by multiple studies [9, 10]. Additional studies indicate that pulmonary toxicity is associated with doses to functional lung [11, 12].

1.4 Functional lung avoidance radiotherapy planning

Functional Lung Avoidance Radiation Therapy (FLART) is an emerging lung RT technique that utilizes functional images to account for lung function heterogeneity and preferentially spares high-function lung (HFL) regions for better lung function protection [13-15]. Substantial studies reveal that FLART can effectively reduce doses to HFL regions without severely damaging conventional dose metrics [16, 17].

Furthermore, clinical studies indicate that it is of potential to reduce pulmonary toxicity and improve treatment efficacy compared to conventional lung RT (ConvRT) [18, 19].

In the planning phase, FLART additionally aims to redirect doses from HFL regions to low-function lung (LFL) regions (referred to functional dose objectives) while maintaining other conventional dose objectives [20]. There are two effective venues to achieve dose redirection as illustrated by Figure 1. The first one is to select beam angles that preferentially spare HFL regions prior to plan optimization, which has been proven effective in reducing doses to HFL [21]. Another commonly used method is to modulate fluence maps by incorporating functional dose objectives in plan optimization to decrease doses to HFL [22]. The strategies of incorporating functional dose objectives can be divided into two categories: the contour-based strategy in which a HFL volume is contoured and dose-volume histogram (DVH) metrics of HFL are considered; and the voxel-based strategy in which a functional weight is assigned to each lung voxel and dose-function histogram (DFH) metrics of the functionally weighted lung (FWL) are used [11].

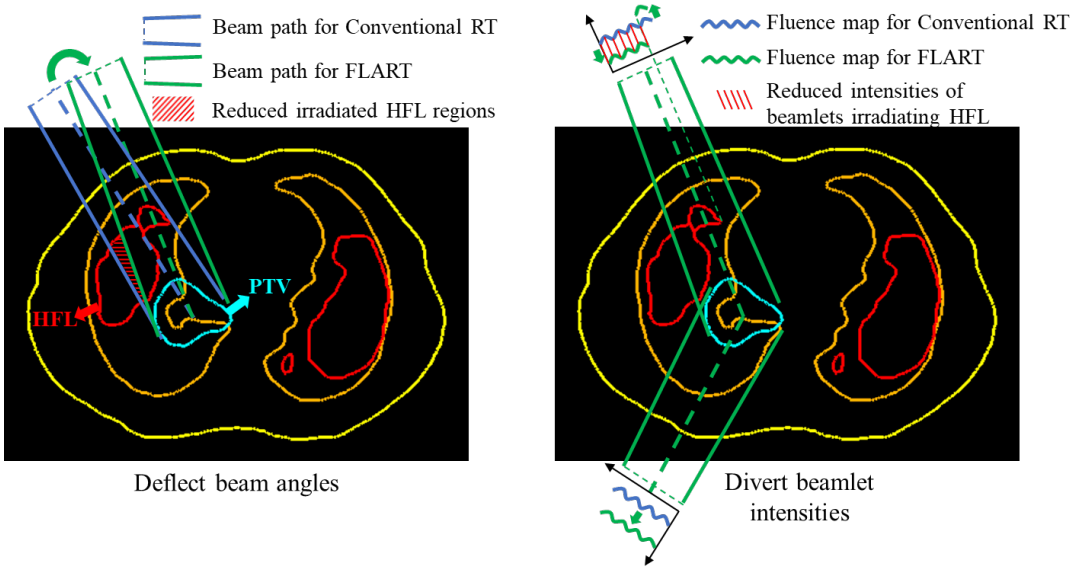


Figure 1. Illustration of two strategies for redirecting doses away from high-function lung regions to low-function lung regions.

Currently, most FLART studies have employed manual planning to achieve the FLART planning goal [14, 18, 19, 23-26]. As demonstrated by Figure 2, it requires manually setting beam angles that preferentially spare HFL regions and iteratively adjusting plan hyperparameters in plan optimization to achieve a subjectively deemed balance between functional and conventional dose objectives. This is a time-consuming and labor-intensive process and highly dependent on the planner's experience while it is relatively scarce as FLART is an emerging technique and has not been widely applied in clinical practice [27]. Additionally, unlike anatomical OARs that exhibit great similarity in population, the lung function distributions feature in significant inter-patient variability and intra-patient heterogeneity as demonstrated by Figure 3, making it more challenging for planners to select appropriate beam angles and estimate achievable functional dose objectives [27]. Consequently, manual planning may require considerable time and effort and yet

could still produce inconsistent or even suboptimal FLART plans.

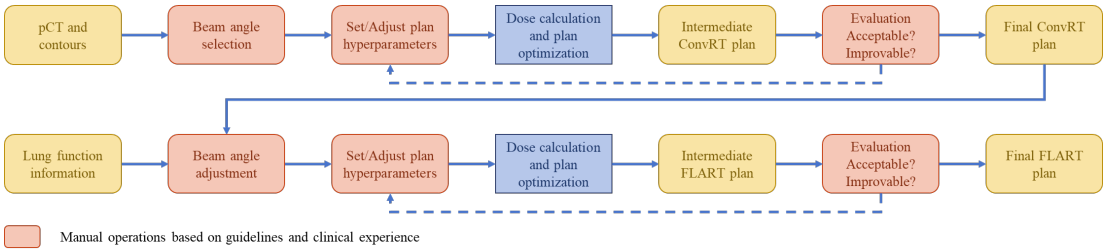


Figure 2. Workflow of manual FLART planning.

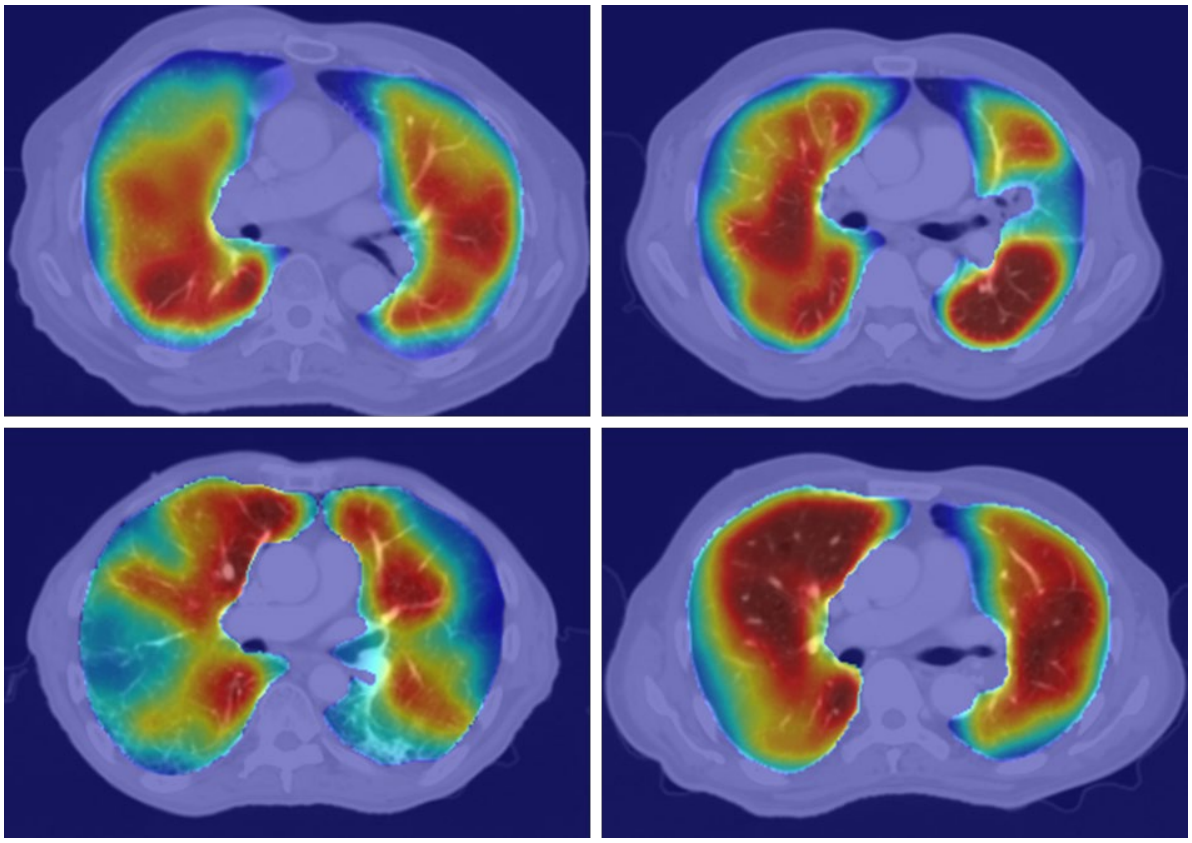


Figure 3. Planning CT images and SPECT perfusion images for four lung cancer patients.

1.5 Automatic planning for functional lung avoidance radiotherapy

The automation of RT planning has been an active area of research since the 1980s, aimed at assisting the design of RT plans [28-31]. Extensive studies have

demonstrated that automatic planning methods can significantly reduce the time and effort required for RT planning, mitigate inconsistencies across institutions and planners, and enhance the overall quality of treatment plans [28, 30-32]. Various automatic and semi-automatic RT planning algorithms have been developed, including knowledge-based [28], multi-criteria optimization (MCO)-based [33, 34], deep learning (DL)-based [35-38], and hyperparameter optimization-based approaches [39, 40]. Some of these methodologies have been commercialized and successfully implemented in clinical practice [28, 33, 34, 41, 42]. However, it is noteworthy that most existing auto-planning algorithms have predominantly focused on conventional anatomy-guided RT, with very few algorithms developed specifically for multi-modality-guided FLART.

The existing auto-planning algorithms for FLART can be categorized into knowledge-based planning (KBP) [27] and MCO-based [20]. *Faught et al*. adopted a KBP method to automate contour-based FLART planning by treating HFL as an additional OAR while it fails to achieve voxel-based FLART planning [27]. *Matuszak et al.* applied *a priori* MCO-based algorithm to sequentially optimize functional and conventional dose objectives according to a predefined priority list [20]. However, it relies on the fixed priority list that does not account for patient-specific anatomical and functional distributions, potentially leading to suboptimal FLART plans. Furthermore, the existing auto-planning methods primarily rely on direct prediction or sequential optimization of dose objectives, neglecting the spatial information of dose

distributions, which may compromise plan quality [36].

Recently, deep learning (DL) techniques have been adopted to predict optimal dose distributions for RT [35-37]. They typically involve collecting a dataset of high-quality RT plans to train a neural network model to predict voxel-wise optimal dose distributions, given planning CT images and contours of planning target volume (PTV) and OARs. The DL-based dose prediction approaches have proven to be effective in various tumor sites including head and neck [36], breast [43], liver [44], prostate [45], cervical [46], and lung [35]. However, the existing dose prediction models have been limited to conventional anatomy-guided RT, with no studies addressing the development of dose distribution prediction models for FLART.

The distinctive characteristics of FLART pose significant challenges in developing an effective dose prediction model. As a kind of data-driven approach, the performance of DL-based models is heavily contingent upon the quantity and quality of training plans [47]. However, the collection of a substantial number of high-quality FLART plans is particularly challenging, given that FLART is an emerging technique that has not yet been widely adopted in clinical practice. Moreover, in contrast to conventional anatomy-guided dose prediction models, dose prediction models for FLART shall incorporate an additional image modality—lung function images—to accurately predict optimal dose distributions. However, the significant inter-patient variability of lung function distributions amplifies the necessity for extensive training data so that

the model can learn to extract generalizable functional features from various lung function distributions.

1.6 Research objectives and hypotheses

Recognizing the research gaps and challenges, this study aims to develop and evaluate a novel auto-planning method for FLART based on multi-modality-guided dose prediction (MMDP). The MMDP model can predict personalized optimal dose distribution for FLART by extracting complementary features from patient-specific anatomical and functional information. When integrated with a function-guided dose mimicking algorithm, it has the potential to automate and streamline the FLART planning process. Specifically, the study aims to achieve the following three objectives:

1. To develop a meta-optimization (MO)-based auto-planning algorithm for FLART that can automatically produce high-quality FLART plans.
2. To develop an MMDP-based auto-planning algorithm for FLART, which builds upon and surpasses the MO-based method in planning quality and efficiency.
3. To facilitate the clinical application of the proposed auto-planning algorithms by conducting clinical implementation and evaluation.

The hypotheses of this study are as follows:

Primary Hypothesis:

The implementation of the proposed automated treatment planning system for FLART

will reduce the mean dose to high-function lung regions and the functionally weighted mean lung dose by at least 10% when compared to conventional manual planning techniques.

Secondary Hypotheses:

Clinical Utility: The automated FLART plans will achieve a clinical acceptability rate exceeding 80% upon clinical expert review, demonstrating their viability for routine clinical implementation.

Planning Efficiency: The auto-planning system will generate complete, deliverable FLART plans within a timeframe of under 10 minutes, demonstrating significantly improved planning efficiency.

Planning Quality: The dosimetric quality and expert review performance of the automated FLART plans are non-inferior to manual plans generated by medical physicists with approximately five years of clinical treatment planning experience.

The proposed auto-planning algorithms have potential to enhance the efficiency, consistency, and quality of FLART planning, and reduce radiation-induced lung toxicity compared to the current clinical standard, ultimately benefiting lung cancer patients worldwide.

1.7 Report layout

This report consists of seven sections. In the first section, we introduce the

background of this study, including lung cancer, lung cancer RT, planning for lung cancer RT, planning for FLART, auto-planning for FLART as well as research objectives. In the second section, we reviewed related literature about treatment planning for FLART and auto-planning for RT. In the third, fourth, and fifth section, we present the contents of three studies corresponding to the three research objectives respectively. The sixth and seventh sections are appendices and references respectively.

2 Literature review

2.1 Incorporation of lung function images into FLART planning

2.1.1 Lung function imaging techniques

The primary function of the lungs is to facilitate gas exchange between the circulatory system and the external environment. The ability to exchange gases can be measured by pulmonary ventilation/perfusion (V/Q) imaging [14]. A variety of imaging techniques are employed to quantify lung V/Q function distributions, including approaches based on Nuclear Medicine (NM), Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) [14].

NM-based techniques are currently regarded as the clinical gold standard for lung function imaging, encompassing single-photon emission computed tomography (SPECT) and positron emission tomography (PET). In SPECT V imaging, Technetium-99m diethylene triamine penta-acetic acid (99mTc-DTPA) [48] and Krypton-81m (81mKr) gas [49] are commonly utilized as contrast agents. For SPECT Q imaging, Technetium-99m labeled macro aggregated albumin (99mTc-MAA) particles are typically employed in clinical practice [50]. PET V imaging relies on radioactive gases, such as Gallium-68 labeled carbon monoxide (Galligas), while Gallium-68 labeled MAA serves as a widely used radioactive tracer for PET Q images [51]. Numerous prospective and retrospective studies have validated the feasibility and efficacy of these NM-based lung function imaging techniques in FLART [23, 51-54]. However, the requirement for additional NM scans, which are not part of the

current clinical routine for lung radiotherapy, introduces additional financial and time costs, as well as radiation exposure for patients.

MRI-based lung function imaging techniques can be categorized into contrast-based and non-contrast-based methods. Contrast-based MRI imaging techniques necessitate the inhalation or injection of contrast agents, such as helium-3 (^{3}He) and xenon-129 (^{129}Xe) for V [55], and Gadolinium-Based Contrast Agents (GBCA) for Q [56]. In contrast, non-contrast MRI techniques generate V/Q surrogate maps by utilizing Fourier decomposition to analyze dynamic changes in proton signal intensity with respect to respiratory and cardiac frequencies [57]. Compared to NM-based techniques, MRI-based lung function imaging techniques have the advantage of ionizing radiation free. However, MRI scans are not currently part of the standard clinical routine for lung RT. The requirement for specialized equipment and contrast agents in contrast-based MRI techniques, and the limited accuracy of non-contrast MRI techniques, pose significant challenges to the broader implementation of MRI-based lung function imaging in clinical practice.

Recently, CT-derived ventilation/perfusion imaging (CTVI/CTPI) techniques have been proposed to construct V/Q surrogate maps from standard CT or 4D CT images [58]. The CTVI/CTPI techniques can be classified into three categories: deformable image registration (DIR)-based, feature-based, and deep learning-based methods. The DIR-based algorithms generate V/Q surrogate maps by leveraging DIR methods to

extract regional lung volume/density changes across different phases [58-60]. The feature-based algorithms capture texture features such as radiomics features from CT images to construct V/Q surrogate maps [61, 62]. The deep learning-based algorithms typically employ neural networks to transform CT images into SPECT V/Q images [63-66]. A significant advantage of CTVI/CTPI techniques is that they do not require additional imaging scans, as planning CT scans are already part of the standard clinical protocol for lung RT. Several prospective studies have validated the efficacy and feasibility of CTVI in FLART [14, 18, 19, 25]. However, the accuracy of DIR-based CTVI/CTPI algorithms requires further enhancement, while the feature-based and deep learning-based approaches face challenges regarding their explainability.

2.1.2 Functional dose metrics for FLART planning

As demonstrated by Figure 4, the strategies of incorporating functional dose objectives into RT planning can be divided into two categories: the contour-based strategy in which a HFL volume is contoured and DVH metrics of HFL are considered; and the voxel-based strategy in which a functional weight is assigned to each lung voxel and DFH metrics of the functionally weighted lung (FWL) are used [11, 67].

In the contour-based strategy, the definitions of HFL can be categorized into two types. The first type is the threshold segmentation-based definition, where functional values derived from V/Q images are binned into percentiles [11]. The HFL is defined as lung

regions with function values exceeding a predetermined threshold. The second type is the defect identification-based definition, which involves dividing the lung into six regions [68]. Assuming a homogeneous distribution of lung function, each region would ideally contain approximately 16.7% (100% divided by 6) of the total function as measured from the V/Q map. Regions with function values deviating by a specified percentage from this idealized homogeneous value are identified as defect regions, while the remaining regions are classified as HFL [11]. The HFL contour can be directly integrated into current treatment planning systems (TPSs) as an additional OAR structure and aid in FLART planning. However, the contour-based strategy has limitations, as it neglects intra-contour function heterogeneity, and a standardized definition of HFL remains controversial [11].

In the voxel-based strategy [19, 69], linear weighting is commonly utilized to convert V/Q maps into FWL maps as described by Equation (2.1) [67]. W_i represents the weight value of voxel i in the resulted FWL map. F_i is the function value of voxel i from functional maps. N is the total number of voxels in the lung. *Faught et al.* proposed a non-linear weighting approach as formulated by equation (2.2). F_{max} represents the maximum function value. h_c and m are two adjustable hyperparameters [11].

$$W_i = \frac{N \bullet F_i}{\sum_i^N F_i} \tag{2.1}$$

$$W_i = \exp\left(-exp\left(\frac{-e \bullet m(F_i - h_c)}{F_{max}}\right)\right) \tag{2.2}$$

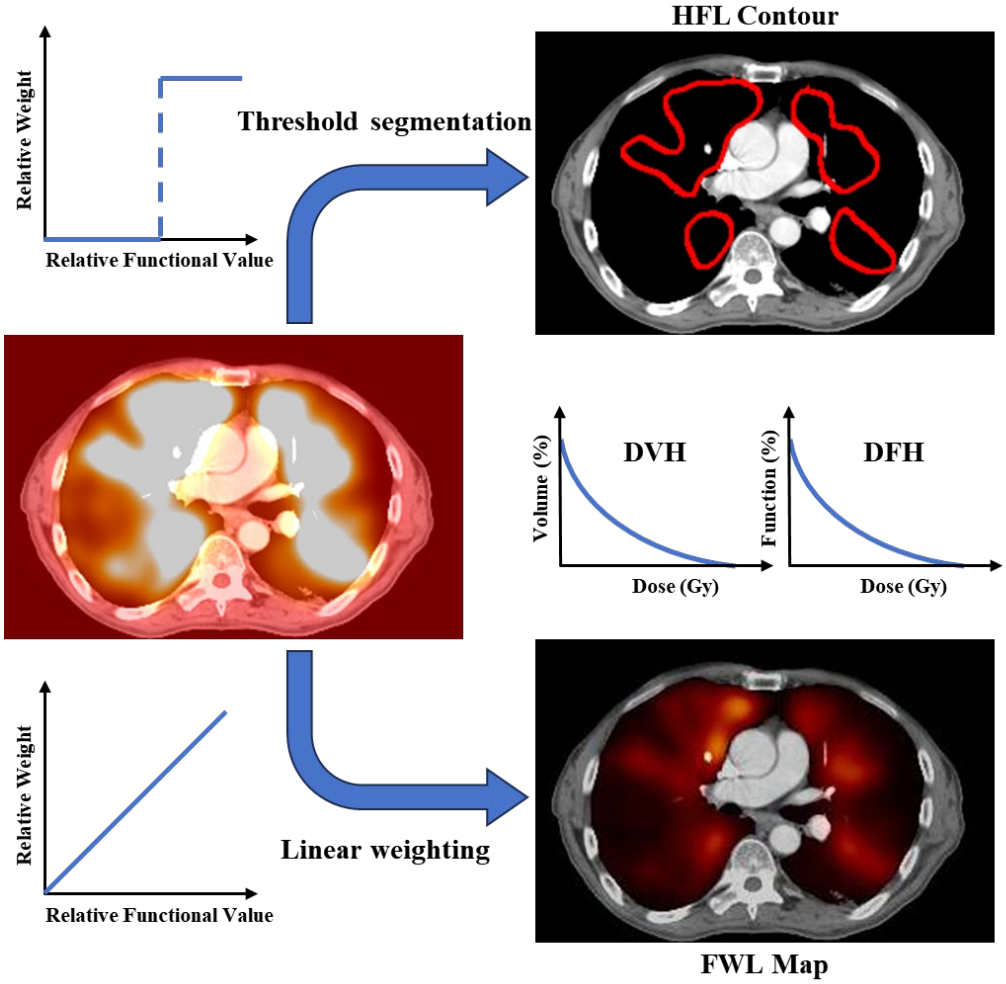


Figure 4. Illustration of contour-based and voxel-based FLART planning strategies.

To achieve the FLART planning goal of redirecting doses away from HFL to LFL, several dose metrics are incorporated into FLART planning due to their effectiveness in predicting radiation-induced pulmonary toxicity. In contour-based FLART planning, HFL V_{20} (portion of HFL volume receiving dose higher than 20 Gy) and HFL D_{mean} (mean dose to HFL) are two commonly used dose metrics. Similarly, F_{20} (portion of function receiving dose higher than 20 Gy) and fMLD (functionally weighted mean lung dose) are commonly used in voxel-based FLART planning [11, 67].

2.2 Automatic planning for radiotherapy

As shown in Figure 5 (a), in the manual planning process, treatment planners typically interactively adjust plan hyperparameters using an inefficient trial-and-error approach. The effects of these hyperparameters on the resulting treatment plan are often non-intuitive and can vary based on several factors, including model formulation, optimization algorithms, and anatomical distribution [32, 70, 71]. This method is not only time-consuming and labor-intensive but also heavily reliant on the clinical experience of the planners, which can lead to variability in plan quality. To reduce the time and effort required for RT planning, mitigate inconsistencies across institutions and planners, as well as to enhance the overall quality of treatment plans, various auto-planning algorithms have been proposed [30, 32, 70]. As shown in Figure 5, the existing auto-planning algorithms include KBP-based, MCO-based, hyperparameter optimization-based, and deep learning-based approaches.

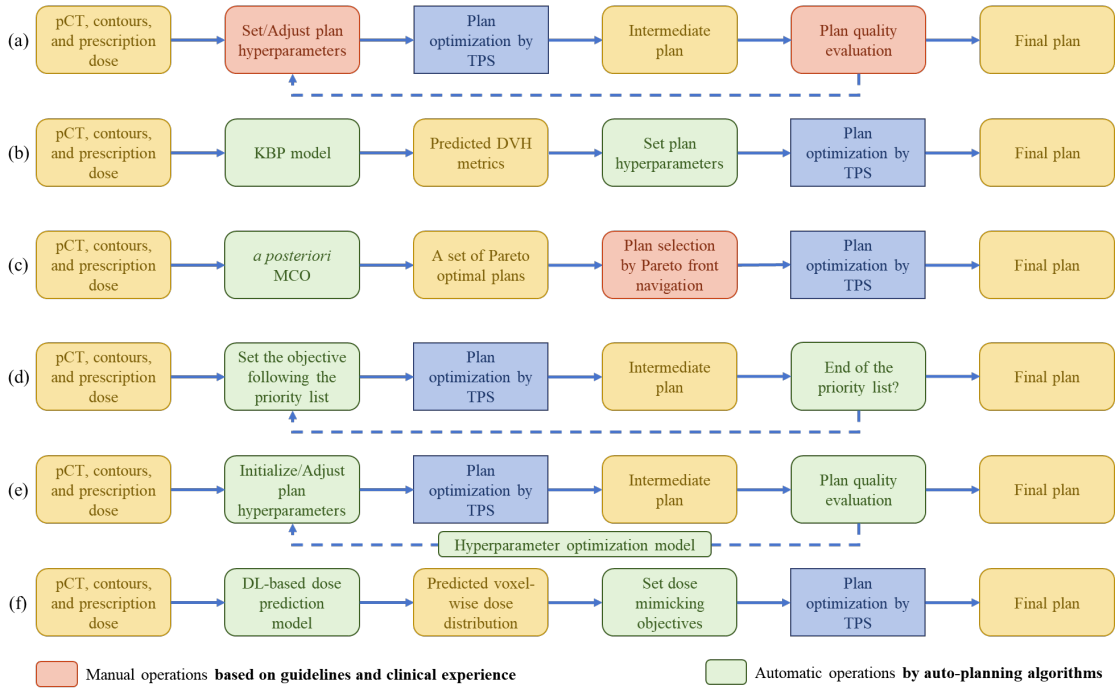


Figure 5. Comparison of flowcharts for (a) manual planning and auto-planning based

on (b) KBP, (c) *a posteriori* MCO, (d) *a priori* MCO, (e) hyperparameter optimization, and (f) deep learning.

2.2.1 Knowledge-based auto-planning algorithms for RT

Knowledge-based planning (KBP) models can predict planning results (DVH metrics) by extracting handcrafted features from anatomical information, thereby facilitating auto-planning [28]. The KBP methods can be classified into two categories: atlas-based methods and statistical modelling-based methods [28]. Both categories involve the compilation of datasets comprising high-quality RT plans to construct plan libraries or to train statistical models. A common approach is to gather manual plans developed by experienced physicists as part of the clinical RT workflow.

In atlas-based KBP methods [72], a plan library is constructed using the collected dataset of high-quality treatment plans. Given a new patient, one or more similar cases can be identified from the library by matching their anatomical features, which are extracted from their anatomical information. These identified plans can then be utilized to guide the planning process for the new case. The features used for plan matching are critical and typically handcrafted based on clinical expertise. Commonly employed features include overlap volume histogram (OVH) [73], tumor location [74], mutual information of the beam's eye view projections [72], and so on [28].

As for the statistical modeling-based methods [75], the collected dataset is utilized to

train machine learning models that establish statistical relationships between anatomical features and planning results such as DVH metrics. Once trained, the models can predict planning results by extracting anatomical features from anatomical information of a new patient. The predicted results can be used to guide planning for the new patient. Commonly used machine learning models in this context include support vector regression [76], multivariate regression [77], random forest regression [78], and so on. Various features such as OVH [79], tumor volume [80], distance-to-target histogram (DTH) [76] are employed in the modeling process.

Substantial studies have validated the effectiveness of the KBP methods [27, 81-83]. Moreover, the KBP method has been implemented in the commercially available Eclipse RapidPlan system (Varian Medical Systems, Palo Alto, CA, USA) and applied in clinical practice [81]. Nevertheless, the handcrafted features may fail to fully exploit anatomical information. Furthermore, conventional machine learning models exhibit relatively limited modeling capabilities compared to emerging DL models. The limitations may affect the matching/prediction accuracy, compromising plan quality [36].

2.2.2 MCO-based auto-planning algorithms for RT

RT planning constitutes a multi-objective optimization problem, involving optimization of multiple, potentially conflicting objectives. To address this complexity, some researchers have applied multi-criteria optimization (MCO) algorithms to

automate the RT planning process. The MCO-based auto-planning algorithms can be categorized into *a posteriori* and *a priori* MCO approaches [20, 84-86].

In *a posteriori* MCO approaches, a set of Pareto optimal plans are initially generated to construct/approximate a Pareto front [85, 86]. Pareto optimal plans are characterized by the fact that no improvement can be made to any objective without compromising one or more of the other objectives. Planners can navigate the approximate Pareto front to select a plan that achieves a subjectively determined balance among the various objectives [33]. The selected plan can guide final plan generation. The *a posteriori* MCO approach has been implemented in commercially available systems including RayStation (RaySearch Laboratories, Stockholm, Sweden) and Eclipse (Varian Medical Systems, Palo Alto, CA, USA). While navigation along the Pareto front assists planners in evaluating trade-offs among different objectives, the manual plan selection process can introduce inter-planner variability, and the final plan may be different from the selected option as the Pareto front is approximate [86].

As for *a priori* MCO approaches, a series of objectives are sequentially optimized following a predefined priority list to produce a Pareto optimal plan [20, 87, 88]. Compared with *a posteriori* MCO approaches, *a priori* MCO approaches can fully automate the planning process. The priority list is the key to *a priori* MCO approaches and typically designed based on clinical experience and intuition. However, the fixed priority list ignores the inter-patient variability in anatomical

distributions, which may result in plans that do not align with preferences of planners.

2.2.3 Hyperparameter optimization-based auto-planning algorithms for RT

Some studies have sought to leverage hyperparameter optimization algorithms to achieve the automatic adjustment of plan hyperparameters [71, 89, 90]. The paradigm usually involves two key components: a plan score function and a hyperparameter optimization algorithm.

The plan score function serves the purpose of quantifying the quality of intermediate plans. It generally incorporates various clinically important plan quality metrics, such as homogeneity index and conformity index for PTV, DVH metrics for OARs.

The hyperparameter optimization algorithm iteratively adjusts plan hyperparameters with the aim of improving plan quality quantified by the plan score function. However, solving this optimization problem presents challenges, as the plan score function is typically expensive to evaluate, non-convex, and lacks a closed-form expression in relation to the plan hyperparameters. To address these challenges, various gradient-free optimization algorithms have been employed, including evolutionary algorithms [91], Bayesian optimization [71, 89], and the Nelder–Mead simplex search method [90]. More recently, reinforcement learning (RL) has also been adopted for RT auto-planning, wherein an RL agent is trained to automatically adjust plan hyperparameters by evaluating and observing intermediate planning results [39, 40, 92]. Despite the

promising outcomes associated with hyperparameter optimization-based auto-planning algorithms, the reliance on a fixed plan score function can overlook inter-patient variability in anatomical distributions, potentially resulting in plans that are not clinically optimal.

2.2.4 Deep learning-based auto-planning algorithms for RT

Recently, deep learning techniques have also been adopted to facilitate RT auto-planning. The DL-based auto-planning algorithms can be divided into two categories: fluence map prediction [38, 93, 94] and dose distribution prediction [35-37, 46, 52, 94]. Similar to the KBP methods, DL-based methods demand a dataset of high-quality plans for model training. Retrospectively collecting clinical plans generated during routine RT practice is a common way for constructing the training dataset. The typical paradigm involves training a neural network to predict optimal fluence map or planning dose distribution given patient-specific anatomical information. The prediction results can be utilized to guide or automate plan generation. The neural networks commonly employ classic encoder-decoder architectures, such as 3D U-net [35, 36], ResNet [95], conditional generative adversarial network (cGAN) [96], and so on. The encoder can automatically extract features from input anatomical information, and the decoder can predict fluence map or dose distribution by using the extracted features.

The other auto-planning algorithms primarily rely on the direct prediction/sequential

optimization/automatic adjustment of dose objectives, which ignore the spatial information of dose distribution and may therefore compromise plan quality [36]. In contrast, DL-based dose prediction methods have the advantage of predicting voxel-wise dose distributions. Combined with the automatic feature extraction and powerful modelling capabilities of neural networks, DL-based dose prediction approaches have proven to be effective in various tumor sites including head and neck [36], breast [43], liver [44], prostate [45], cervical [46], and lung [35]. However, as a data-driven approach, the performance of the DL models highly depends on the quantity and quality of training plans. This presents a challenge, as it can be difficult to collect a sufficient number of high-quality plans for emerging radiotherapy techniques or for cancers with low incidence rates.

2.2.5 Auto-planning algorithms for FLART

Despite substantial efforts in developing various auto-planning algorithms for RT, most have predominantly focused on conventional anatomy-guided RT, with very few algorithms developed specifically for multi-modality-guided FLART. *Faught et al*. adopted a KBP method to automate contour-based FLART planning by treating HFL as an additional OAR while it fails to achieve voxel-based FLART planning [27]. *Matuszak et al.* applied *a priori* MCO-based algorithm to sequentially optimize functional and conventional dose objectives according to a predefined priority list [20]. However, it relies on the fixed priority list that does not account for patient-specific anatomical and functional distributions, potentially leading to suboptimal

FLART plans. Furthermore, the existing auto-planning methods rely on direct prediction or sequential optimization of dose objectives, ignoring spatial information of dose distributions, which may compromise plan quality [36].

Compared to conventional anatomy-guided RT, FLART additionally incorporates lung function information and additionally aims to achieve a balance between functional and conventional dose metrics. Unlike most anatomical OARs which demonstrate great similarity among different patients, lung function distributions exhibit significant inter-patient variability and intra-patient heterogeneity as demonstrated by Figure 3. This complexity, combined with the additional planning objectives, presents unique challenges for auto-planning algorithms. Additionally, as an emerging RT technique that has not yet been widely adopted in clinical practice, it is difficult to collect a substantial number of high-quality FLART plans, presenting an additional challenge for the development of auto-planning algorithms, particularly for data-driven approaches.

3 Towards full automation: Meta-optimization (MO)-based auto-planning for FLART

3.1 Introduction

Lung cancer, with an estimated 2.2 million new cases and 1.8 million deaths per year, remains the leading cause of cancer-related morbidity and mortality globally [97, 98]. Radiotherapy (RT) serves as a crucial treatment modality for lung cancer [99]. However, radiation-induced lung injury (RILI), such as radiation pneumonitis, is a common complication of lung RT and can affect the prognosis and life quality of patients [100]. Functional Lung Avoidance Radiotherapy (FLART) shows promise in reducing RILI by leveraging the heterogeneity of lung function distribution to selectively minimize dose to high functional lung (HFL) regions, thereby potentially enhancing lung function protection [11, 13, 19, 22, 25, 101-104], in contrast to conventional anatomical image-guided RT which uniformly treats lung without function guidance.

The key to RT is to deliver high and homogeneous enough doses to planning target volume (PTV) while minimizing doses to organs at risk (OARs). FLART additionally aims to redirect dose from HFL regions to low functional lung (LFL) regions while maintaining other conventional planning criteria [20]. There are two effective venues to achieve dose redirection. The first one is to select beam angles that preferentially spare HFL regions prior to plan optimization, which has been proven effective in reducing doses to HFL [21]. Another commonly used method is to modulate fluence

maps by incorporating functional dose objectives in plan optimization to decrease doses to HFL [22]. The strategies of incorporating functional dose objectives can be divided into two categories: the contour-based strategy in which a HFL volume is contoured and dose-volume histogram (DVH) metrics of HFL are considered [22]; and the voxel-based strategy in which a functional weight is assigned to each lung voxel and dose-function histogram (DFH) metrics of the functionally weighted lung (FWL) are used. The choice between the two strategies remains debated, due to the lack of an established HFL definition for the contour-based strategy and the limitations of current commercial treatment planning systems in implementing the voxel-based strategy.

Currently, most FLART studies have employed manual planning [13, 19, 25, 69, 101-103], which requires manually setting beam angles and iteratively adjusting plan hyperparameters in plan optimization to achieve the FLART planning goal. This labor-intensive process is dependent on the planner's experience while it is relatively scarce for the emerging FLART technique [27]. In addition, unlike anatomical OARs that exhibit great similarity in population, the lung function distributions are highly variable and patient-specific, making it more challenging for planners to select appropriate beam angles and estimate achievable dose objectives. Consequently, manual planning may require considerable time and effort and yet could still produce inconsistent or suboptimal FLART plans [28, 32]. Thus, it is necessary to develop an automatic planning (auto-planning) method for FLART to streamline the beam angle

selection and plan optimization process, enhancing FLART planning efficiency, consistency, and quality.

There are very few existing studies in the realm of auto-planning for FLART. To select high-quality beam angles for FLART, *McGuire et al.* [105] proposed a beam angle selection method based on orientation and dose-function histogram (DFH) contributions. *Bedford et al.* [106] evaluated the quality of each beam angle through conformal beam-based dose calculations and mean functionally weighted lung doses. As for automatic plan optimization, *Matuszak et al.* [20] implemented a priority-driven optimization algorithm to sequentially optimize various functional and conventional dose objectives strictly following a predefined priority list. *Faught et al.* [28] trained a knowledge-based planning (KBP) model to predict optimal plan hyperparameters for FLART.

Though effective dose reduction in HFL has been achieved by these existing methods, they suffer from a few limitations. First, the performance of the KBP model is contingent upon the number and quality of training data, a significant hurdle given FLART's emerging status and the scarcity of high-quality FLART plans. The other approaches solely or excessively focus on minimizing doses to HFL, which may underachieve conventional dose objectives [20]. Currently, the clinical benefits of HFL dose reduction still desire further clinical validation [13, 26]. Therefore, minimizing functional dose metrics with excessive compromises in conventional dose

metrics is undesirable [13, 20]. These studies focus on either beam angle selection or plan optimization for FLART, without offering a fully automatic planning framework. In addition, none of these studies achieve auto-planning for both the contour-based FLART (cFLART) and the voxel-based FLART (vFLART).

In this section, we aim to propose a fully Automatic Planning framework for FLART (AP-FLART) to address the aforementioned challenges. AP-FLART integrates and extends a dosimetric score-based beam angle selection method and a meta-optimization-based plan hyperparameter optimization approach, which can incorporate lung function information and balance conventional and functional dose objectives to facilitate dose redirection in the beam angle selection phase and the plan optimization phase respectively [90, 107]. Different from the existing methods, AP-FLART is training data-free and can comprehensively consider functional and conventional dose metrics to achieve dose redirection from HFL to LFL while minimizing compromises in conventional metrics. It can automate the whole planning process and generate plans for both cFLART and vFLART. In addition to FLART plans, AP-FLART can also produce conventional lung RT (referred to ConvRT) plans so that clinicians can explicitly evaluate the gain of better lung function protection and the potential cost of inferior conventional metrics following the current routine of FLART studies [25]. The proposed AP-FLART shall advance the research and clinical application of FLART by providing labor-free, consistent, and high-quality FLART plans.

3.2 Method

3.2.1 Study design

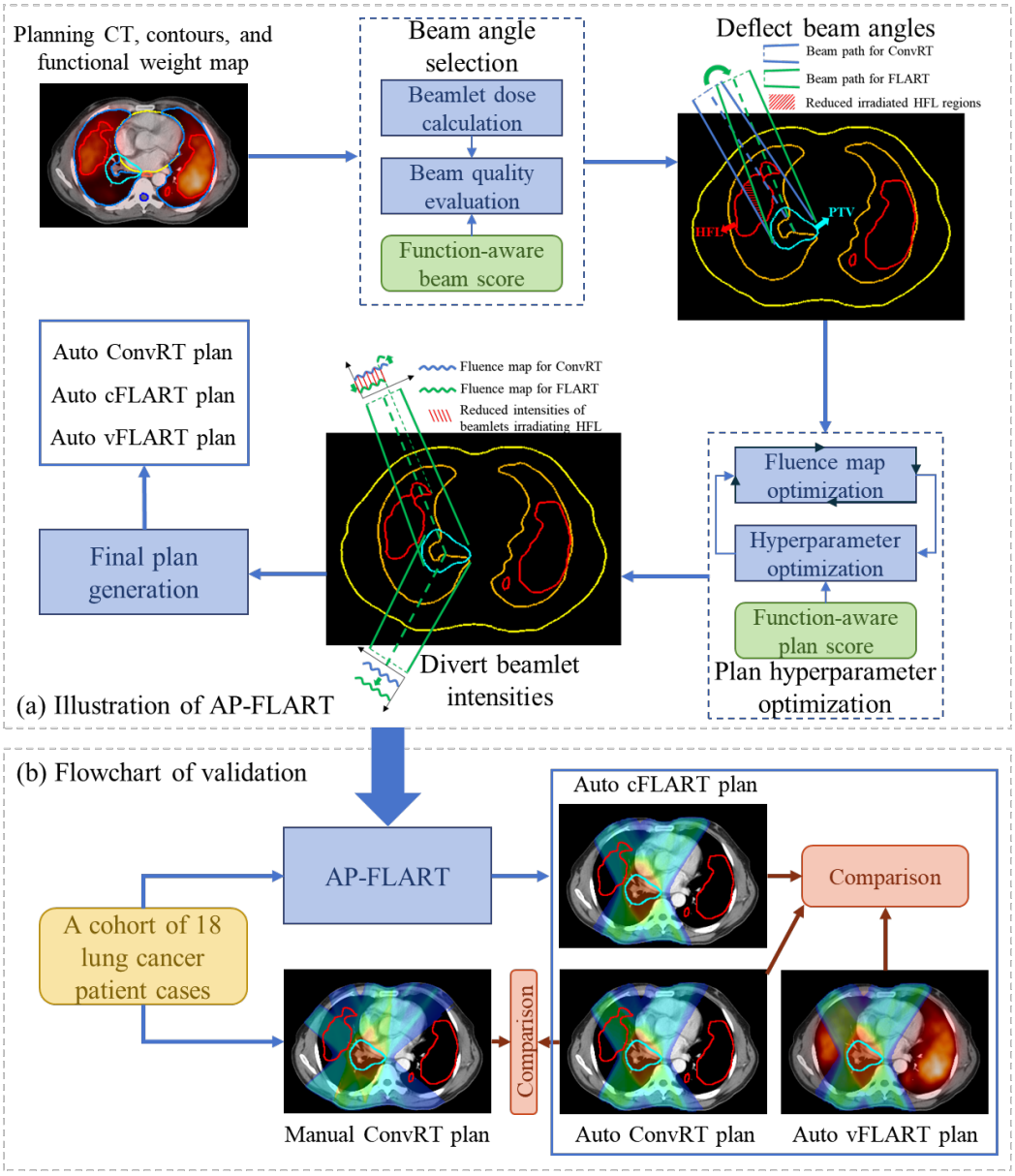


Figure 6. The overview of the study in this section including (a) the illustration of the proposed AP-FLART and (b) the flowchart of its validation.

Figure 6 depicts an overview of this section. AP-FLART can generate ConvRT, cFLART, and vFLART plans by sequentially determining beam angle configurations, optimizing dose objective weights, and generating final plans. The beam angle

selection method employs beam quality evaluation guided by beamlet-wise dose calculations and a beam score function. The plan hyperparameter optimization method optimizes the weights of different dose objectives by navigating the Pareto front guided by a plan score function. Notably, unlike ConvRT planning which solely utilizes anatomical information and considers anatomical dose objectives, FLART planning additionally utilizes lung function information and incorporates functional dose objectives into the beam angle selection and plan optimization processes. Consequently, beam angles that originally irradiate HFL in ConvRT plans tend to be deflected to irradiate LFL, and intensities of beamlets irradiating HFL in ConvRT plans tend to be diverted to those irradiating LFL, both contributing to dose redirection from HFL to LFL. A cohort of patient data was collected for the development and validation of AP-FLART. For all collected patient cases, AP-FLART was applied to generate ConvRT, cFLART, and vFLART plans and the results were compared to evaluate the performance of AP-FLART.

3.2.2 Patient data and preprocessing

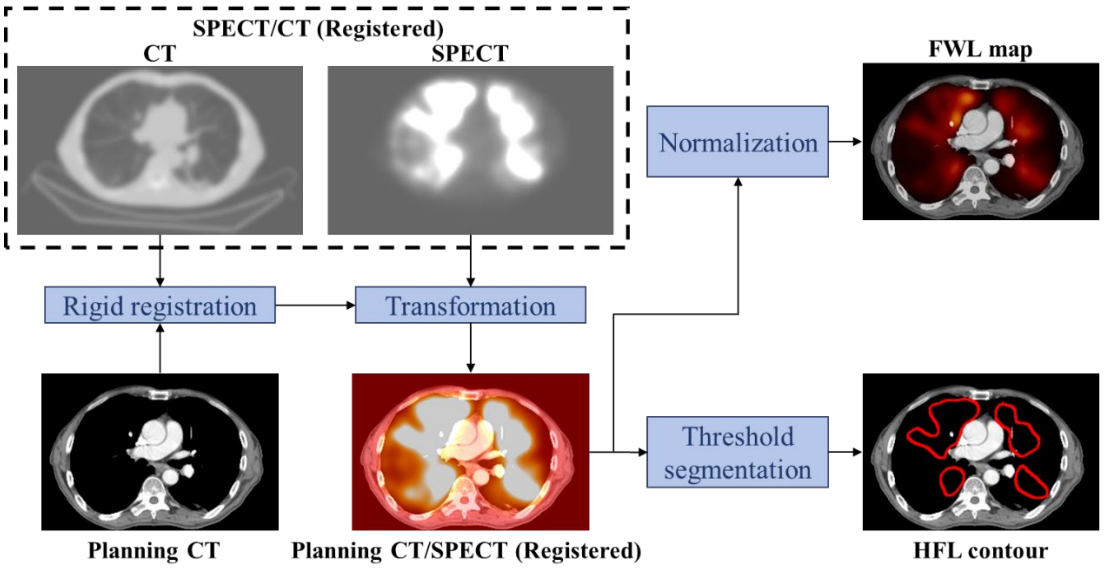


Figure 7. The flowchart of data preprocessing including the registration between SPECT and planning CT as well as the generation of FWL maps and HFL contours.

A total of 18 lung cancer patient cases underwent SPECT/CT scans and intensity modulated radiation therapy (IMRT) were collected from the Affiliated Cancer Hospital of Zhengzhou University with approval by the institutional review board. All cases contain perfusion SPECT/CT images, planning CT images, PTV and OARs contours, and conventional IMRT plans. All conventional IMRT plans were made using the Eclipse (Varian Medical Systems, Palo Alto, CA, USA) treatment planning system (TPS) with the analytical anisotropic algorithm dose engine and dynamic IMRT technique. The radiopharmaceutical Technetium-99m-labelled macroaggregated albumin (MAA) was injected into patients and then 3D perfusion SPECT/CT scans were acquired using a GE Discovery NM/CT 670 CZT scanner (GE HealthCare, Chicago, IL, USA). A total of 60 frames were acquired for each patient

with a frame rate of 10 seconds per frame. The SPECT images had dimensions of 128×128×128 with a voxel size of 4.42×4.42×4.42 mm^3, respectively. The planning CT images had an in-slice resolution of 1.172 mm and a slice thickness of 3 mm. The planning CT and SPECT/CT images were acquired under free-breathing conditions.

The principle of SPECT perfusion imaging is that the labelled MAA particles will be lodged within the pulmonary capillary beds, and their local concentration is directly related to the regional pulmonary blood flow [103]. The SPECT scanner can detect the distribution of these trapped MAA particles and generate SPECT perfusion images, which can quantify lung perfusion function distribution and serve as a surrogate for lung function [103, 108-110]. Therefore, SPECT perfusion images were utilized to guide FLART planning. As shown in Figure 7, the attenuation correction CT images from SPECT/CT were first registered with the planning CT images using rigid registration, and the resulting transformations were applied to register the SPECT images with the planning CT images [20]. All registration operations were performed within MIM platform (MIM Software Inc., Cleveland, OH, USA). Threshold segmentation with a cutoff value of 66^{th} percentile [65, 111] was then performed on the registered SPECT images to generate HFL contours. The intensity values of SPECT images were normalized to the mean intensity value within the lung contour to obtain FWL maps [20].

3.2.3 Function-guided dosimetric score-based beam angle selection

A novel function-guided beam angle selection method is developed by extending the dosimetric score-based method originally proposed for conventional lung IMRT by *Yuan et al.* [107]. The original method evaluates beam angle qualities by using beamlet-wise dose calculations and anatomical information. Our enhancement differs from the original one by comprehensively utilizing both anatomical and functional information, aiming to deflect beam angles from HFL to LFL.

The quantitative evaluation of each beam is based on beamlet-wise comparison. For each beam angle, this process involves computing dose distributions of each beamlet and then quantifying the beamlet's quality by calculating a function-aware beamlet-wise quality score. The beam score (*bs*) functions for ConvRT, cFLART and vFLART are defined as follows:

$$bs_{\alpha i} = \frac{\sum_{jOAR} w^{jOAR} \sum_{k \in jOAR} d_{\alpha i}^{k} + w^{NT} \sum_{k \in NT} d_{\alpha i}^{k}}{\sum_{k \in PTV} d_{\alpha i}^{k}}, for\ ConvRT \quad (3.1)$$

$$bs_{\alpha i} = \frac{\sum_{jOAR} w^{jOAR} \sum_{k \in jOAR} d_{\alpha i}^{k} + w^{NT} \sum_{k \in NT} d_{\alpha i}^{k} + w^{HFL} \sum_{k \in HFL} d_{\alpha i}^{k}}{\sum_{k \in PTV} d_{\alpha i}^{k}}, for\ cFLART \quad (3.2)$$

$$bs_{\alpha i} = \frac{\sum_{jOAR} w^{jOAR} \sum_{k \in jOAR} d_{\alpha i}^{k} + w^{NT} \sum_{k \in NT} d_{\alpha i}^{k} + w^{FWL} \sum_{k \in FWL} f^{k} d_{\alpha i}^{k}}{\sum_{k \in PTV} d_{\alpha i}^{k}}, for\ vFLART \quad (3.3)$$

Where $bs_{\alpha i}$ is the function-aware quality score for the beamlet *i* of the beam α. $d_{\alpha i}^{k}$ represents doses irradiated by the beamlet to voxel *k* of OARs, normal tissues (NT), HFL, FWL and PTV. The weights including w^{jOAR}, w^{NT}, w^{HFL}, and w^{FWL} reflect the relative importance of different regions to be protected. f^{k} refers to the functional

weight value of voxel k in FWL. A lower $bs_{\alpha i}$ indicates a preferred balance of minimizing doses to protected areas while delivering high enough doses to PTV, marking a beamlet as of high quality. Different from $bs_{\alpha i}$ for ConvRT, $bs_{\alpha i}$ for cFLART additionally includes a HFL term and $bs_{\alpha i}$ for vFLART replaces the lung term with a FWL term. The details are provided in Appendix 6.1. The quality of each individual beam is then quantified by calculating a beam-wise quality score (bs_{α}), the average of all beamlet-wise quality scores.

In addition to the quality of individual beams, the collective effects of multiple beams should also be considered. Clinical insights suggest that overly condensed beam angle distributions may lead to poor plan quality [107, 112, 113]. Therefore, a beam span constraint is introduced when selecting beam angles [107, 114]. The optimized beam angle configuration is selected by minimizing the following objective function:

$$OF_{\Pi} = \sum_{\alpha_i \in \Pi} bs_{\alpha_i} + \sum_{\alpha_i, \alpha_j \in \Pi; \alpha_i \neq \alpha_j} \frac{2k}{1 - \cos(|\alpha_i - \alpha_j|) + \delta} \qquad (3.4)$$

Where α_i and α_j represent two different beam angles in the beam angle configuration Π. The objective function consists of two terms: a sum of all beam-wise quality scores reflecting the individual quality of selected beam angles and a beam span term discouraging tightly clustered beam configurations. k stands for the weight of the beam span term and δ is a regularization parameter. The beam angle search space will be constrained to coplanar orientations, with candidate angles sampled at 5-degree intervals. The selected beam angle configuration, denoted as Π, will be initialized as an empty set. For each iteration of the optimization process, the beam angle search

space will be traversed to identify the specific angle that results in the smallest incremental increase in the objective function, OF_{Π}. This optimal beam angle will then be added to the selected configuration Π, and both this angle and its opposing angle will be removed from the remaining search space following clinical practice. This iterative process will continue until the total number of selected beam angles reaches the predefined target number, which is set to be 8 following the reference [107].

The angle selection method involves several parameters to be determined, including the weights of different regions to be protected as well as the weight and regularization parameter of the beam span term of Equation (3.4). For ConvRT, these parameters were fine-tuned based on the results from *Yuan et al*. [107]. The weights for the anatomical/functional lung term for FLART were then further fine-tuned to minimize functional dose metrics while maintaining conventional dose metrics. The details of parameter determination are presented in Appendix 6.1.

3.2.4 Function-guided meta-optimization-based plan optimization

The meta-optimization framework, proven to be effective for conventional head and neck as well as prostate RT planning [90], is further enhanced in this study for FLART plan hyperparameter optimization. The original meta-optimization method navigates the Pareto front to identify plans that achieve a balance between different conventional dose objectives guided by a plan score involving important conventional

metrics. In this study, we accommodate the approach to lung RT and extend it by incorporating additional lung function information to facilitate dose redirection for FLART. We also accelerate the framework by designing a novel initialization method to ensure the computation time is clinically acceptable.

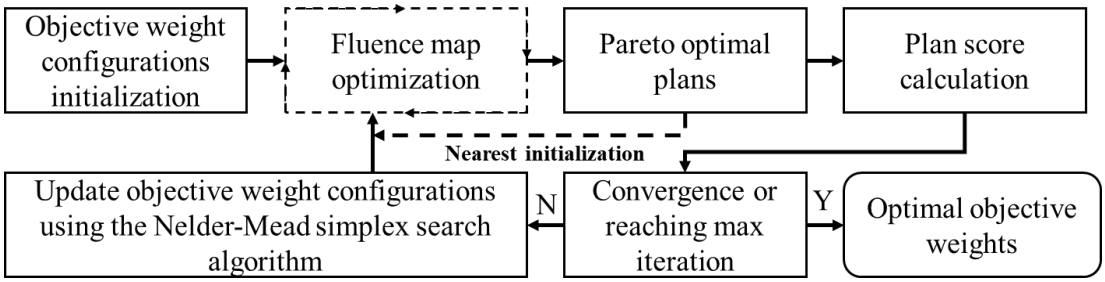


Figure 8. Flowchart of meta-optimization-based plan hyperparameter optimization.

The meta-optimization framework consists of two nested loops of optimization as demonstrated in Figure 8. In the inner loop, fluence map optimization (FMO) is performed to generate Pareto optimal plans by minimizing the weighted sum of conventional dose objectives and potential functional dose objectives as formulated by the following equations:

$$\min_{x} \quad \frac{w_{PTV}}{N_{PTV}} \sum_{i \in PTV} \left(d_i - D_p\right)^2 + \sum_{S \in OARs} \frac{w_S}{N_s} \sum_{i \in S} H(d_i - D_s)(d_i - D_s)^2 \,, for\ ConvRT \tag{3.5}$$

$$\frac{w_{PTV}}{N_{PTV}} \sum_{i \in PTV} \left(d_i - D_p\right)^2 + \sum_{S \in OARs} \frac{w_S}{N_s} \sum_{i \in S} H(d_i - D_s)(d_i - D_s)^2 + \frac{w_{HFL}}{N_{HFL}} \sum_{i \in HFL} H(d_i - D_{HFL})(d_i - D_{HFL})^2 \,, for\ cFLART \tag{3.6}$$

$$\frac{w_{PTV}}{N_{PTV}} \sum_{i \in PTV} \left(d_i - D_p\right)^2 + \sum_{S \in OARs} \frac{w_S}{N_s} \sum_{i \in S} H(d_i - D_s)(d_i - D_s)^2 + \frac{w_{FWL}}{N_{FWL}} \sum_{i \in FWL} H(d_i - D_{FWL}) f_i (d_i - D_{FWL})^2 \,, for\ vFLART \tag{3.7}$$

$$s.t. \quad x \geq 0$$

$$w \geq 0$$

$$\vec{d} = D\vec{x}$$

Where $\vec{x}$, D, and $\vec{d}$ refer to the fluence map, the dose influence matrix, and the dose distribution respectively. d_i and f_i represent the dose value and the functional weight value of voxel i respectively. w and N are the weights of different objectives and the number of voxels in different regions respectively. D_p, D_s, D_{HFL}, and D_{FWL} stand for reference dose of different objectives. H represents the Heaviside function. The details of dose objectives are summarized in Table 1.

In the outer loop, the gradient-free Nelder-Mead simplex search algorithm [90, 115, 116] is applied to iteratively optimize the weights of different objectives (w) to minimize a function-aware plan quality score (ps). This plan score (ps) integrating several plan quality metrics of clinical importance is used to quantify the clinical acceptability of different Pareto optimal plans generated from the inner loop. The outer loop is formulated as:

$$\min_{w} \qquad ps(w) = \sum_{QM} 2^{-P_{QM}} F_{QM} \qquad (3.8)$$

$$s.t. \qquad w \geq 0$$

$$F_{QM} = \begin{cases} 1, & QM \geq QM_+ \\ \dfrac{QM - QM_-}{QM_+ - QM_-}, & QM_- < QM < QM_+ \\ 0, & QM \leq QM_- \end{cases}$$

Where the P_{QM} and F_{QM} refer to the priorities and scoring values of different quality metrics (QM). The relationship between the score values and the QM values is formulated as a clipped linear function, and the linear range is defined as [QM_- QM_+]. The quality metrics of different structures and their priorities as well as linear ranges

are enumerated in Table 1. Among them, functional dose metrics including mean dose, V_5, and V_{20} for HFL, as well as functionally weighted mean lung dose (fMLD), F_5, F_{20} for FWL are utilized for FLART because they are widely used in FLART studies [11, 25, 69]. The other conventional dose metrics are extracted from the Radiation Therapy Oncology Group (RTOG) 0617 protocol and are commonly used in clinical plan evaluation [6, 26, 69, 102, 107]. Additionally, a novel nearest initialization method was introduced to accelerate the meta-optimization framework, achieving a 63% reduction in computation time, with details presented in Appendix 6.1. The implementation details of the meta-optimization framework including weight configuration initialization, convergence conditions, and so on are presented in the reference [90].

Table 1. The details of the dose objectives used in the inner loop, the quality metrics involved in the outer loop, and the dose constraints for OARs considered in evaluation.

Structure	Reference Dose (Gy)	Quality Metric (QM)	QM-	QM+	Priority	Dose Constraints
For ConvRT:						
PTV	60	HI	5	10	0	
		CI	0.6	0.7	1	
Spinal Cord	35	D_{max}	40.5Gy	45Gy	2	D_{max}<45Gy
Esophagus	0	D_{mean}	30.6Gy	34Gy	2	D_{mean}<34Gy
Heart	0	D_{mean}	40.5Gy	45Gy	2	V_{40}<30% D_{mean}<28Gy

		D_{mean}	0Gy	15Gy	1	D_{mean}<20Gy
Lung	0	V_5	25%	50%	1	V_5<60%
		V_{20}	0%	25%	1	V_{20}<30%
Body	30	D_{mean}	8.1Gy	9Gy	3	
Changes when converting from ConvRT to cFLART:						
						D_{mean}<20Gy
Lung	0	D_{mean}	18Gy	20Gy	2	V_5<60%
						V_{20}<30%
		D_{mean}	0Gy	15Gy	1	
HFL	0	V_5	25%	50%	1	
		V_{20}	0%	25%	1	
Changes when converting from ConvRT to vFLART:						
						D_{mean}<20Gy
Lung			None			V_5<60%
						V_{20}<30%
		fMLD	0Gy	15Gy	1	
FWL	0	F_5	25%	50%	1	
		F_{20}	0%	25%	1	

Note: CI (Conformity Index) = $\frac{(TV_{95Dp})^2}{TV \times V_{95D_p}}$; HI (Homogeneity Index) = $\frac{D_5 - D_{95}}{D_p} \times 100$; D_p: Prescription dose; D_{mean}: Mean Dose; D_{max}: Maximal Dose; V_x: Portion of structure volume irradiated by dose higher than x Gy; D_x: Minimum dose to the "hottest" x% of structure volume; F_x: Portion of lung function irradiated by dose higher than x Gy; fMLD: Functionally weighted mean lung dose.

Upon determining the optimal beam angles and dose objective weights, final plans are generated by sequentially performing FMO, leaf sequencing, and direct aperture optimization using the open-source TPS matRad [117]. The pencil beam dose calculation engine of matRad [118] and static IMRT technique [119] are utilized in the beam angle selection and plan optimization processes. We extend matRad to enable it to incorporate the voxel-wise FWL maps and optimize the dose-function objectives for vFLART.

3.2.5 Evaluation and ablation studies

To evaluate the performance of AP-FLART, comparison studies were conducted by using the collected cohort. For each patient case, a ConvRT plan, a cFLART plan, and a vFLART plan were automatically devised by using AP-FLART. Dose distributions of automatic plans and collected manual plans were normalized to guarantee that the prescription dose (60Gy) could cover 95% of PTV for consistent comparison. The analysis included comparing automated ConvRT plans with manual plans to validate the auto-planning framework's effectiveness. Since manual plans were made using Eclipse while automatic plans were generated based on matRad, to reduce discrepancies caused by using different TPSs, a voxel-based dose mimicking algorithm [36, 120] was applied to convert all manual Eclipse plans to matRad plans which would be used for comparison with automatic plans. The details of applying the dose mimicking algorithm for plan conversion are presented in Appendix 6.1. Additionally, comparisons between automated FLART and ConvRT plans were

executed to ascertain AP-FLART's ability of facilitating dose redirection.

There are two driving forces to achieve dose redirection in AP-FLART. The function-guided beam angle selection method can deflect beam angles irradiating HFL and the function-guided plan optimization approach can divert intensities of beamlets irradiating HFL. We conducted several ablation studies to quantify the contribution of the two driving forces. In the study "ConvBeam_cFLART" and "ConvBeam_vFLART", beam angles selected for ConvRT and plan hyperparameters optimized for FLART were used to generate cFLART and vFLART plans respectively. Another two studies adopted beam angles selected for cFLART and vFLART and plan hyperparameters optimized for ConvRT, and they were referred to as "cFLARTBeam_ConvRT" and "vFLARTBeam_ConvRT" respectively. All quality metrics listed in Table 1 were included in the evaluation criteria. In addition, D_1 of PTV, Esophagus, and Heart were also evaluated to monitor the change of hot spots.

3.3 Results

3.3.1 Comparison between manual and automatic ConvRT plans

Table 2. Comparison between manual and automatic ConvRT plans as well as comparison between automatic ConvRT and FLART plans. The mean and standard deviation of the dose metrics over all collected cases are listed. The p-values by Wilcoxon Signed-Rank test are also shown.

Dose metrics	Manual ConvRT VS Auto ConvRT			Auto FLART VS Auto ConvRT			
	Manual ConvRT	Auto ConvRT	p-value	Auto cFLART	p-value	Auto vFLART	p-value
HI	6.71±1.76	5.56±0.51	**0.030**	5.61±0.52	0.304	5.55±0.54	0.966
CI	0.63±0.09	0.76±0.06	**<0.001**	0.72±0.06	**0.006**	0.71±0.08	**0.004**
PTV D_1 (Gy)	64.61±1.17	63.93±0.35	**0.027**	64.01±0.41	0.276	63.96±0.43	1.000
HFL D_{mean} (Gy)	7.17±2.48	7.21±2.28	0.766	6.08±2.44	**<0.001**	6.18±2.39	0.001
HFL V_{20} (%)	12.88±5.66	13.01±5.55	1.000	11.00±5.75	**<0.001**	11.14±5.60	0.003
HFL V_5 (%)	26.63±9.03	32.64±10.50	0.038	25.98±9.32	**<0.001**	26.30±12.22	0.002
FWL fMLD (Gy)	8.62±2.51	8.32±2.22	0.246	7.76±2.44	0.010	7.68±2.37	**0.009**
FWL F_{20} (%)	15.75±5.61	15.18±5.23	0.284	14.71±5.57	0.551	14.28±5.55	0.099
FWL F_5 (%)	30.66±8.08	35.92±9.27	0.014	30.69±9.16	<0.001	30.85±10.87	**0.005**
Lung Dmean (Gy)	10.59±2.12	10.04±2.11	**0.021**	10.05±1.88	0.671	9.94±2.04	0.347
Lung V_{20} (%)	19.73±4.44	18.74±4.59	0.130	19.56±4.09	0.108	19.18±4.77	0.932
Lung V_5 (%)	35.05±5.76	39.79±8.61	**0.007**	36.55±7.74	0.007	36.26±7.75	0.016
SpinalCord D_{max} (Gy)	36.63±3.95	37.69±6.86	0.393	38.01±5.59	0.734	37.63±6.47	0.468
Esophagus D_{mean} (Gy)	13.32±6.39	13.29±8.38	0.609	12.05±7.80	0.005	12.91±8.07	0.099
Esophagus D_1 (Gy)	50.93±17.26	49.80±21.81	0.865	48.86±21.94	0.130	49.84±20.23	0.054
Heart D_{mean} (Gy)	7.37±5.95	6.37±6.21	0.734	6.69±6.38	0.799	7.03±6.33	0.832
Heart D_1 (Gy)	39.27±17.33	33.41±23.82	0.196	33.88±23.96	0.832	34.67±23.71	0.196
Heart V_{40} (%)	1.92±2.19	2.29±2.86	0.518	2.47±3.39	0.969	2.49±3.02	0.784
Body D_{mean} (Gy)	4.91±1.48	4.78±1.55	**0.048**	4.76±1.49	0.671	4.73±1.50	0.130

Note: CI (Conformity Index) $= \frac{(TV_{95Dp})^2}{TV \times V_{95D_p}}$; HI (Homogeneity Index) $= \frac{D_5 - D_{95}}{D_p} \times 100$; D_p: Prescription dose; D_{mean}: Mean Dose; D_{max}: Maximal Dose; V_x: Portion of structure volume irradiated by dose higher than x Gy; D_x: Minimum dose to the "hottest" x% of structure volume; F_x: Portion of lung function irradiated by dose higher than x Gy; fMLD:

Functionally weighted mean lung dose.

Table 2 reveals that automatic ConvRT plans generally match the quality of manual plans. Specifically, compared to manual ConvRT plans, automatic ConvRT plans significantly enhanced five out of thirteen conventional metrics including HI, CI, PTV D_1, mean doses to lung and body while significantly increased lung V_5. The other metrics showed no statistically significant change.

3.3.2 Comparison between automatic ConvRT and FLART plans

Table 2 also presents comparison between automatic ConvRT and FLART plans. Compared with ConvRT plans, FLART plans significantly reduced functional dose metrics with modest compromise in conventional dose metrics. The cFLART plans showed a significant decrease by 1.13 Gy for HFL mean dose ($p<.001$), 2.01% for HFL V_{20} ($p<.001$), and 6.66% for HFL V_5 ($p<.001$), with a decrease by 5.26% for CI ($p<.01$). As for vFLART plans, FWL fMLD, F_{20}, and F_5 were reduced by 0.64 Gy ($p<.01$), 0.90% ($p=0.099$), and 5.07% ($p<.01$) respectively while CI decreased by 6.58% ($p<.01$). Though inferior conformity was observed in FLART plans, all dose constraints listed in Table 1 were well satisfied. Figure 9 compares DVH and DFH curves of automatic FLART and ConvRT plans. Lower doses to HFL and FWL were observed in cFLART and vFLART plans. The cFLART and vFLART plans demonstrated similar mean DVH for HFL and mean DFH for FWL.

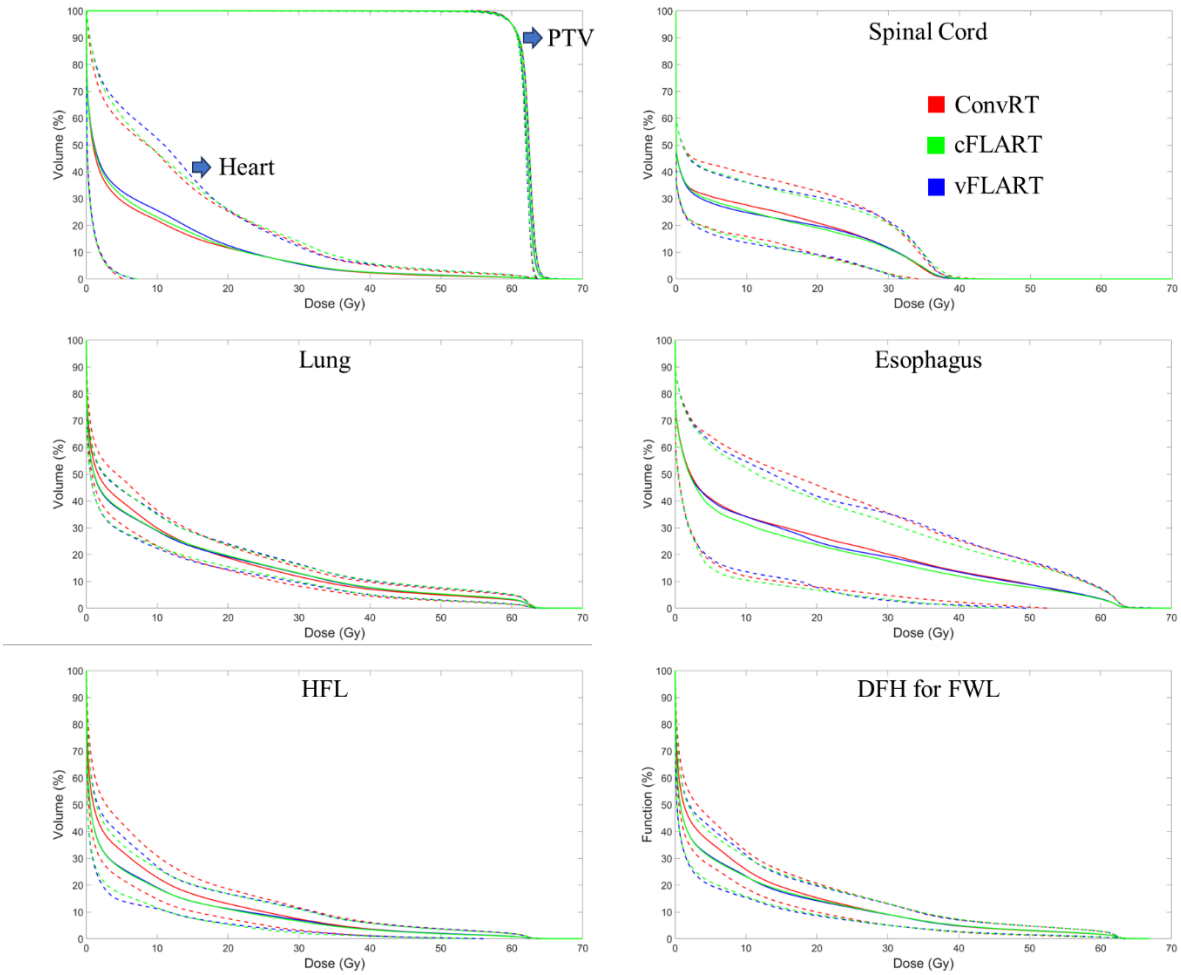


Figure 9. Comparison of DVH for anatomical structures and DFH for FWL among automatic ConvRT plans (red), cFLART plans (green), and vFLART plans (blue). The solid lines and dashed lines represent the means and the standard deviations over all collected cases.

Figure 10 showcases iso-dose curves of ConvRT and FLART plans in an example case. We observed effective dose redirection in cFLART and vFLART plans including dose reduction in high function regions indicated by red arrows and dose increase in low function regions indicated by blue arrows.

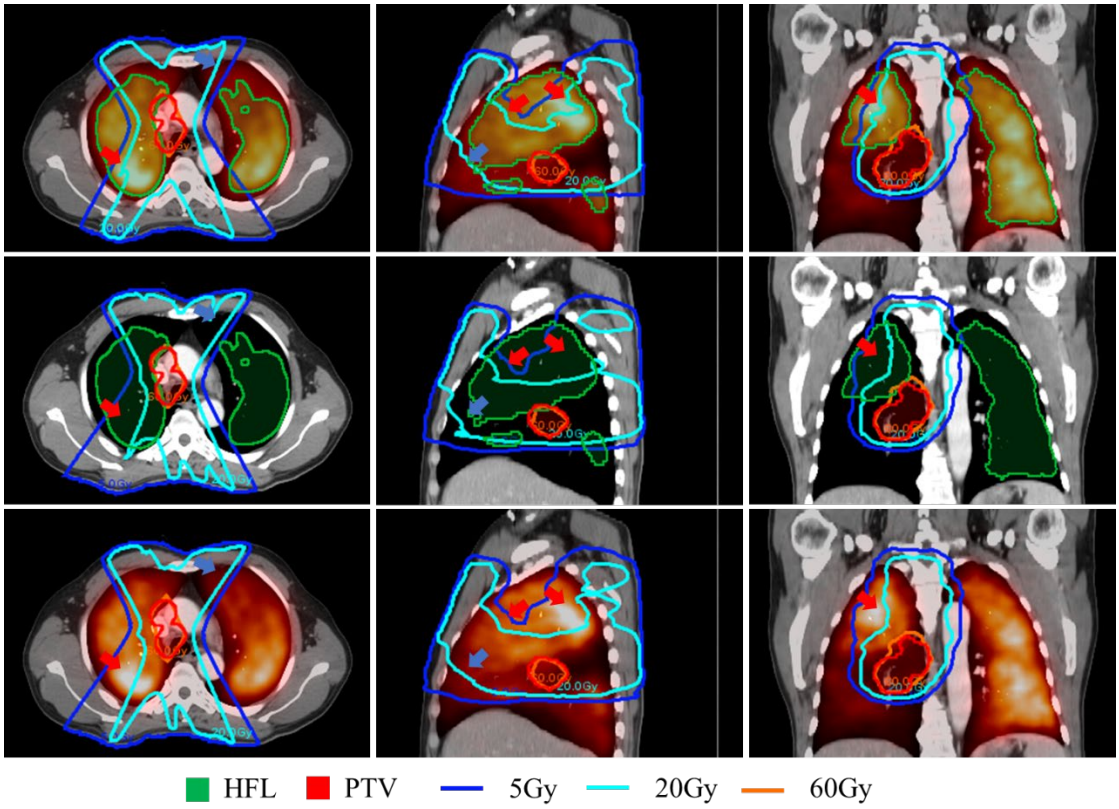


Figure 10. Comparison of iso-dose curves of 5 Gy (dark blue), 20 Gy (light blue), and 60 Gy (orange) between ConvRT (upper), cFLART (middle), and vFLART (lower) plans for an example case displayed in axial (left), sagittal (middle), and coronal planes (right).

3.3.3 Ablation study results

The ablation study results are depicted in Figure 11 and elaborated in Appendix 6.1. Compared with the automatic ConvRT plans, the plans of "cFLARTBeam_ConvRT" and "vFLARTBeam_ConvRT" reduced HFL mean dose by 5.13% (p=0.099) and FWL functionally weighted mean dose by 3.73% (p=0.099) respectively. The "ConvRTBeam_cFLART" and "ConvRTBeam_vFLART" plans achieved a reduction in HFL mean dose by 10.82% (p<.001) and FWL functionally weighted mean dose by

3.49% (p=0.108) respectively. With the combination of the two driving forces for shifting dose, cFLART further reduced HFL mean dose by 11.11% (p<.001) and 5.44% (p<.01) compared to the "cFLARTBeam_ConvRT" and "ConvRTBeam_cFLART" respectively. FWL functionally weighted mean dose further decreased by 4.12% (p<.01) and 4.36% (p<.001) in vFLART compared to the "vFLARTBeam_ConvRT" and "ConvRTBeam_vFLART".

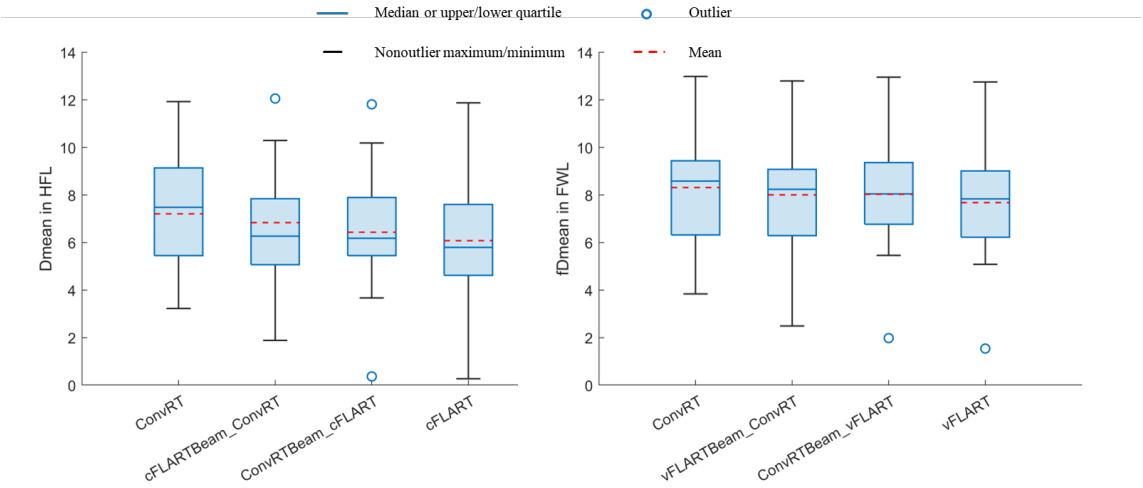


Figure 11. Results of the ablation studies including boxplots of HFL mean dose (left) and FWL functionally weighted mean dose (right).

3.3.4 Computation time

AP-FLART was evaluated on a workstation equipped with Intel(R) Xeon(R) Gold 6226R CPUs (2.9 GHz). The voxel size of dose grid and bixel width of beamlet grid were set to be 3×3×3 mm^3 and 5 mm respectively for beam angle selection. The plan hyperparameter optimization adopted dose grid and beamlet grid with a coarser resolution of 5×5×5 mm^3 and 7 mm respectively for improving computation efficiency. In the final plan generation process, the voxel size of dose grid and bixel

width of beamlet grid were set to be 2×2×2 mm^3 and 5 mm respectively. The parallel computation with the parallel number of 8 was adopted for beam angle selection and plan hyperparameter optimization. As a result, the computation time for beam angle selection, plan hyperparameter optimization, and final plan generation was 7.7±2.4 minutes, 36.7±11.5 minutes, and 44.6±15.4 minutes per plan respectively, culminating in a total computation time of 89.0±26.4 minutes per plan.

3.4 Discussion

This section introduced an auto-planning method for FLART (AP-FLART) integrating and enhancing a dosimetric score-based beam angle selection method and a meta-optimization-based plan hyperparameter optimization method. AP-FLART can generate ConvRT plans with similar or better quality in terms of all evaluated conventional metrics except for lung V_5 compared to manual plans, validating the effectiveness of the proposed auto-planning method. The higher lung V_5 of automatic plans is a compromise to achieve better PTV homogeneity and conformity. In addition, the increased lung V_5 (39.79±8.61%) is well below the dose constraints (60%). Furthermore, AP-FLART is capable of taking into account lung functional information in both beam angle selection and plan optimization phases using either the contour-based or voxel-based strategy and generating FLART plans. Compared with the automatic ConvRT plans, the automatic FLART plans significantly reduced HFL mean dose, V_5, and V_{20} by 1.13 Gy, 6.66%, and 2.01%, as well as FWL fMLD, F_5, and F_{20} by 0.64 Gy, 5.07%, and 0.90% respectively, which could potentially better protect patients' lung function and reduce the risk of pulmonary side effects [11]. The reduction in the functional dose metrics is similar to existing FLART planning studies. *Yaremko et al.* achieved a reduction in HFL mean dose, V_5, and V_{20} by 1.0 Gy, 3.5%, and 2.9% respectively [26]. *Vinogradskiy et al.* reported that HFL mean dose, V_5, and V_{20} were reduced by 1.4 Gy, 3.4%, and 3.5% respectively [25]. In the study by *Yamamoto et al.*, FWL fMLD and F_{20} decreased by 0.5 Gy and 3.8% respectively [19].

In addition to reducing functional dose metrics, FLART also aims to minimize

compromises in conventional dose metrics [13, 20]. For Pareto optimal plans, dose reduction in HFL will inevitably result in dose increase in other regions. Ideally, FLART redirects these doses from HFL to LFL instead of affecting other OARs, thus preserving conventional dose metrics. According to the results presented in Table 2, the automatic FLART plans significantly reduced HFL mean dose and FWL fMLD while maintaining anatomical lung mean dose similar, indicating an effective dose shift from HFL to LFL as demonstrated in Figure 10 as well. Though a reduction in conformity was observed, none of the other conventional dose metrics changed significantly. The conformity losses, often observed in FLART studies [20, 111, 121], due to lung function's anisotropic distribution around PTV, are considered as a common trade-off for dose redirection.

The ablation study results demonstrate that the function-guided beam angle selection and plan optimization jointly contributed to dose redistribution. Distinct from the functional dose metrics-oriented beam angle selection methods proposed by *McGuire et al.* and *Bedford et al.* which solely or excessively pursue optimal lung function protection [105, 106], our proposed beam angle selection method features in comprehensively considering both functional and conventional dose metrics, therefore advancing in minimizing compromises in conventional dose metrics. This is achieved by utilizing a function-aware beam score function to guide the beam selection process. The weights of different protected regions in the score function represent their relative importance. In the transition from ConvRT to FLART, the weights of conventional

OARs remain constant, and the relative weight of the anatomical/functional lung term is maintained. Consequently, beams originally irradiating HFL in ConvRT plans are preferably deflected to LFL instead of other OARs.

In the plan optimization phase, incorporating functional dose objectives can reduce doses to HFL. We adapted and enhanced the meta-optimization-based approach to optimize the weights of conventional and functional dose objectives to achieve a balance. Compared to the priority-driven approach [20] which results in statistically significant dose increase in heart, spinal cord, and esophagus, our method better maintains the conventional dose metrics. A small compromise in functional dose metrics may potentially considerably improve conventional dose metrics, and such a compromise will be preferably made in clinical practice. However, the priority-driven approach fails to do so as it strictly follows a priority list and the priority of minimizing doses to functional lung is higher than conventional OARs. In contrast, the meta-optimization-based approach is more flexible and small compromises in functional dose metrics can be made only if they are clinically beneficial indicated by a lower plan score.

Given the absence of established protocols for FLART currently, much effort is being made to develop a standardized protocol [13, 101]. AP-FLART can assist FLART research by providing labor-free, high-quality, and consistent FLART plans. For example, there are some controversies about which kind of patients may benefit from

FLART. The study by *Li et al.* [102] reported that patients with PTV surrounded by HFL may gain more dosimetric benefits from FLART plans while *Munawar et al.* observed inconsistent results [122]. The work by *Christian et al.* [22] concluded that patients with one large area of LFL benefit most from FLART plans while another study by *Lavrenkov et al.* [123] indicated that FLART plans are more beneficial in patients with non-uniform scattered-distributed LFL. These discrepancies may stem from the limited number of patient cases included in these studies and the plan variability caused by different planning strategies and different planners. AP-FLART alleviates these limitations by considerably lowering the barriers to including a larger dataset and ensuring plan consistency, thereby expediting the development of patient selection criteria and the standardization of FLART planning protocols.

Upon the establishment of FLART protocols, AP-FLART will boost the widespread clinical application of FLART. Considering the scarce clinical experience and the difficulties introduced by the complex and patient-specific lung function distributions in FLART planning [27], it is of clinical significance for AP-FLART to assist beam angle selection and plan optimization. In this study, we utilized a specific FLART planning strategy including SPECT perfusion-based lung function imaging, threshold segmentation-based HFL, and linear weighting-based FWL to demonstrate the feasibility and effectiveness of AP-FLART. Nevertheless, it is anticipated to be still effective when using other strategies whichever are adopted by the protocol without much modification. It can accommodate other lung function imaging modalities such

as CT-based perfusion/ventilation images [61, 65, 102], other HFL definitions such as homogeneity deviation-based HFL [27], and other FWL definitions such as nonlinear weighting-based FWL [11]. Furthermore, when the benefits of minimizing functional dose metrics become more evident, the weights of functional dose metrics in the beam and plan scores can be increased to divert more doses away from HFL.

A few limitations exist in the study presented in this section. First, the validation of AP-FLART's effectiveness was conducted on a limited dataset of 18 cases from a single center because data containing both RT and lung function imaging is scarce. Future studies will assess its performance on a larger multi-center dataset. Currently, AP-FLART is implemented in the open-source TPS matRad to demonstrate its technical feasibility and efficacy. Although we have adopted the dose mimicking algorithm to convert manual Eclipse plans into matRad plans for consistent comparison, there might still be some unaccounted differences since the reference manual plans were originally made using Eclipse TPSs. Future work will integrate the AP-FLART algorithm into clinical TPSs, such as Eclipse and RayStation to enable direct clinical evaluation. Implementing AP-FLART in clinical TPSs can also reduce planning time and facilitate broader clinical application through their efficient and well-validated dose calculation and plan optimization engines.

3.5 Conclusion

In this section, a novel Automatic Planning framework for FLART (AP-FLART) was developed and validated. It integrates and enhances a dosimetric score-based beam angle selection method and a meta-optimization-based plan hyperparameter optimization approach. AP-FLART demonstrates the ability to generate radiotherapy plans of similar quality to manual plans within clinically acceptable time. Furthermore, AP-FLART can incorporate lung function information, employing either contour-based or voxel-based strategies, to redirect doses from high function regions to low function regions without severely damaging conventional dose metrics. Effective dose redirection is jointly driven by function-guided beam angle selection and plan optimization. AP-FLART is of potential to improve FLART planning efficiency, consistency, and quality, and advance the research and clinical application of FLART.

4 Towards faster and better: Multi-modality guided dose prediction (MMDP)-based auto-planning for FLART

4.1 Introduction

In the third section, we proposed a meta-optimization (MO)-based auto-planning algorithm for FLART. An MO approach [90] was employed to automate the adjustment of anatomical and functional objective hyperparameters driven by a predefined plan score function. Despite its promising performance, it suffers from a few limitations. First, the fixed plan score function fails to account for patient-specific anatomical and functional distributions, potentially leading to suboptimal FLART plans. Second, it relies on the automatic adjustment of dose objectives, neglecting the spatial information of dose distributions, which may compromise plan quality [35, 36]. Additionally, the relatively long planning time (around 1.5 hours per plan even with parallel planning optimization) hinders its clinical application, particularly for online adaptive radiotherapy where fast planning is demanded.

Recently, deep learning (DL) techniques have been adopted to predict optimal dose distributions for RT [35, 36, 124-126]. They typically involve training neural network models to predict optimal dose distributions given planning CT images and contours of planning target volume (PTV) and OARs. The DL models can predict personalized, optimal voxel-wise dose distribution at real-time speeds by extracting features from patient-specific anatomical information. The DL-based dose prediction approaches have been proven to be effective in various tumor sites including head and neck [36],

breast [43], liver [44], prostate [45], cervical [46], and lung [35]. However, the existing dose prediction models have been limited to conventional anatomy-guided RT, with no studies addressing the development of dose distribution prediction models for FLART.

In contrast to conventional anatomy-guided dose prediction models, dose prediction models for FLART shall incorporate an additional image modality—lung function images—to accurately predict function-guided dose distributions. However, the significant inter-patient variability of lung function distributions [27] poses challenges for the model to extract generalizable functional features, particularly given limited available training plans as FLART is an emerging technique and has not been widely adopted.

Recognizing the research gap and challenges, this study [127] aims to develop an automatic lung dose painting algorithm that can automate voxel-based FLART planning. A novel multi-modality-guided dose prediction (MMDP) model is proposed to predict optimal function-painted dose distributions for FLART by extracting complementary features from anatomical and functional information. A super-voxel-based method [61] is employed to generate lung function surrogate maps from widely available planning CT images, enabling the MMDP model to learn to extract effective features from diverse lung function distributions. An instance-weighting anatomy-to-function training strategy is developed to enhance prediction accuracy. A

function-guided dose mimicking algorithm is proposed to translate the predicted dose distributions into FLART plans to achieve automatic lung dose painting. Notably, this study represents the first effort to develop a deep learning-based auto-planning method for FLART, aiming to improve FLART planning efficiency, consistency, and quality, ultimately advancing the research and clinical application of FLART.

4.2 Method

4.2.1 Study design

Figure 12 depicts the overall design of the study in this section. A total of 196 lung cancer patients underwent RT across three institutions were retrospectively or prospectively collected and divided into a training cohort and a testing cohort. For each case in the training cohort, a super-voxel-based CT-derived ventilation imaging (CTVI) algorithm [61] was utilized to generate lung ventilation surrogate maps. Each case in the testing cohort underwent SPECT scans—either perfusion (Q) or ventilation (V)—and the SPECT V/Q images served as lung function maps. The MO-based auto-planning algorithm [67] was applied to generate conventional RT and FLART plans (referred to as MO-ConvRT and MO-FLART plans) guided by the lung function maps. The MMDP model was constructed and trained using high-quality plans selected from the generated plans. An instance-weighting anatomy-to-function training strategy was utilized to facilitate MMDP training. In the evaluation phase, we first evaluated the dose prediction accuracy of the MMDP model by comparing the predicted dose distributions against planning dose distributions on the testing cohort. Additionally, a function-guided dose mimicking algorithm was implemented to convert the predicted dose distributions into FLART plans (referred to as MMDP-FLART plans) for all cases in the testing cohort. The MMDP-FLART plans were compared against MO-ConvRT plans, MO-FLART plans, manual conventional RT (MA-ConvRT) plans, and manual FLART (MA-FLART) plans respectively to assess the performance of the proposed MMDP-based automatic lung dose painting

algorithm.

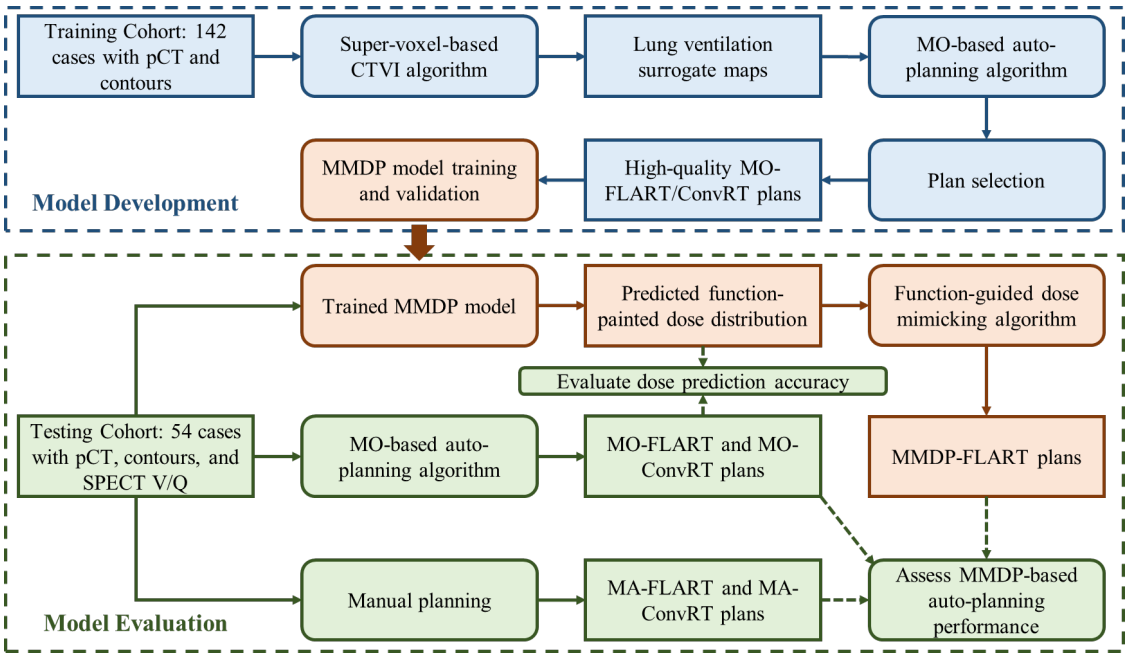


Figure 12. Overall design of the study in this section.

4.2.2 Patient data

Three retrospective datasets and one prospective dataset including a total of 196 lung cancer patient cases received RT were collected for model development and evaluation. The first dataset (Dataset_1) containing 122 patient cases was retrospectively collected from the Affiliated Cancer Hospital of Zhengzhou University. The second retrospective dataset (Dataset_2) containing 20 patient cases was 4D-Lung dataset, a publicly available dataset in the Cancer Imaging Archive from the National Cancer Institute [128, 129]. Each patient case in the first two datasets contains planning CT images and corresponding contours of tumors and OARs. The third dataset (Dataset_3) was prospectively collected from the Affiliated Cancer

Hospital of Zhengzhou University, and 33 patient cases were included. In addition to planning CT images and contours, each patient underwent perfusion SPECT/CT scans. The fourth dataset (Dataset_4) was from the VAMPIRE dataset [58], containing ventilation SPECT/CT images and planning CT images of 21 patients retrospectively obtained from Stanford University [130]. The tumors and OARs were contoured by oncologists.

4.2.3 Lung function mapping and RT planning

The super-voxel-based CTVI algorithm [61] was applied to generate lung ventilation surrogate maps from planning CT images for Dataset_1 and Dataset_2. For each patient, the Simple Linear Iterative Clustering algorithm was first adopted to segment the lung volume of the planning CT into hundreds of super voxels. The mean density value (DEN_{mean}) of each super voxel was then calculated and converted into corresponding mean function value (F_{mean}) according to the established relationship between DEN_{mean} and F_{mean}. Finally, interpolation was performed to generate voxel-wise lung function surrogate maps that matched the resolution and dimensions of the planning CT images. For patients in Dataset_3, rigid registration was performed to register the attenuation correction CT images from perfusion SPECT/CT scans with corresponding planning CT images based on MIM platform (MIM Software Inc., Cleveland, OH, USA). The resulting transformations were directly utilized to align SPECT Q images with planning CT images. Trilinear interpolation was then performed to match the dimensions of SPECT Q images with those of planning CT

images. The SPECT V images from Dateset_4 have been registered and interpolated to match the locations and dimensions of corresponding planning CT images in the VAMPIRE challenge. Following the VAMPIRE challenge [58], an iterative thresholding-based method [131] was adopted to mitigate the influence of spurious ventilation values from the SPECT V images. The generated lung function surrogate maps and matched SPECT V/Q images were normalized to the mean intensity value within the lung volume to generate final lung function maps, which were utilized to guide lung dose painting in FLART planning.

A fully automatic planning framework was applied to generate conventional RT and FLART plans for all collected cases, with high-quality plans selected for training and evaluating the MMDP model [67]. The auto-planning framework encompasses a dosimetric score-based beam angle selection algorithm and a MO-based plan hyperparameter optimization algorithm [90, 107] as detailed in Section 3. It can automatically generate intensity modulated radiation therapy (IMRT)-based conventional RT plans by sequentially executing beam angle selection, plan hyperparameter optimization, and final plan generation. The beam angle selection algorithm quantifies the quality of each beam angle through beamlet-wise dose calculations and a beam score function. The MO-based plan hyperparameter optimization algorithm employs the Nelder-Mead simplex search algorithm [116] to adjust weights of different dose objectives, minimizing a plan score function. Furthermore, given lung function maps, it can incorporate voxel-based functional

dose objectives into beam angle selection and plan optimization to achieve lung dose painting and produce FLART plans. The MO-based auto-planning framework was directly utilized without modifications, aside from adjustments to the number of beam angles, generating five paired MO-ConvRT and MO-FLART plans with beam angle numbers ranging from five to nine for each case. A total of 980 pairs of MO-ConvRT and MO-FLART plans corresponding to 196 patients were obtained. High-quality MO-FLART plans that effectively reduced functionally weighted mean lung dose (fMLD) compared to corresponding MO-ConvRT plans were selected for MMDP model training and evaluation. Consequently, 508 and 214 high-quality MO-FLART plans were chosen from the training and testing cohorts respectively. The training cohort was randomly divided at the patient level into a training dataset containing 114 patients and a validation dataset containing 28 patients.

4.2.4 MMDP model architecture and training

A novel multi-modality neural network was developed and trained to predict voxel-wise optimal function-guided dose distributions given anatomical and functional inputs. As shown in Figure 13(a), the MMDP model was built upon the classic 3D U-net model [132], featuring two encoders and attention-based fusion modules that differentiate it from the standard 3D U-Net. The dual encoders are designed to extract useful and complementary features from anatomical and functional inputs respectively. The extracted anatomical and functional features are fused and enhanced via the attention-based fusion modules consisting of channel attention blocks [133] and

spatial attention blocks [134]. As illustrated in Figure 13(a), the channel attention blocks can highlight salient feature channels, and the spatial attention blocks can highlight salient feature regions. The decoder then utilizes the fused features to generate the optimal dose distributions. Detailed specifications of the MMDP model architecture are provided in Appendix 6.2.

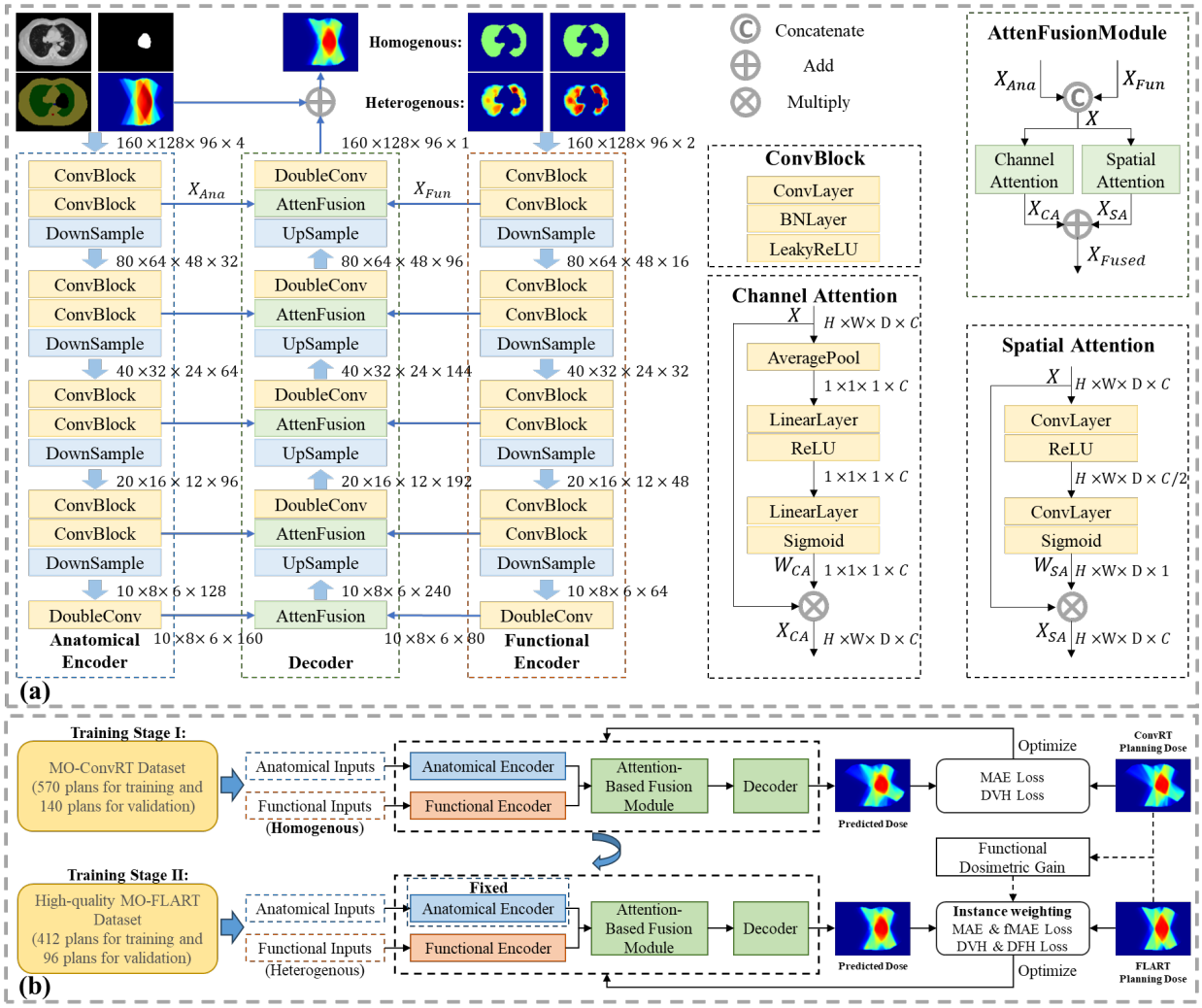


Figure 13. (a) The architecture of the proposed MMDP model. (b) Illustration of the instance-weighting anatomy-to-function training strategy.

As illustrated in Figure 13(b), an anatomy-to-function two-stage training strategy

based on transfer learning was employed to improve the accuracy of MMDP prediction. Given the substantial similarities between conventional lung RT planning and FLART planning, the MMDP model was initially pretrained using MO-ConvRT plan data and subsequently fine-tuned with MO-FLART plan data. This approach emulates the process by which human planners accumulate experience, progressing from simpler conventional lung RT planning to the more complex FLART planning. By leveraging knowledge from related conventional lung RT plans, this strategy can enhance the performance of the MMDP model, particularly when the availability of FLART plans is limited. To further enhance the MMDP's performance, an instance-weighting strategy was devised to prioritize instances with more FLART-specific knowledge. Specifically, for each high-quality MO-FLART plan, a functional dosimetric gain—defined as the difference in fMLD compared to corresponding MO-ConvRT plan—was calculated. It can quantify the disparity of the MO-FLART plan relative to its MO-ConvRT counterpart, indicating the degree of FLART-specific knowledge involved. In the second training stage, the loss of each training instance was weighted according to its functional dosimetric gain to encourage the MMDP model to extract salient and useful functional features.

In the first training stage, the loss function consists of mean absolute error (MAE) loss and dose-volume histogram (DVH) loss [135]. The MAE loss provides voxel-wise guidance to induce the predicted dose distributions close to the ground truth. The DVH loss is applied to improve the prediction accuracy of clinically important DVH

metrics [135]. In the second stage, in addition to MAE and DVH losses, a functionally weighted mean absolute error (fMAE) loss and a dose function histogram (DFH) loss are included to promote the prediction accuracy of function-guided dose distributions. The definitions of MAE, fMAE, DVH, and DFH loss are presented at Appendix 6.2. The instance weighting strategy is formulated as Equation (4.1) and (4.2). $fMLD_j^{MO-FLART}$ and $fMLD_j^{MO-ConvRT}$ indicate fMLD values of the *j*th MO-FLART plan and its corresponding MO-ConvRT plan respectively. $n_{MO-FLART}$ is the total number of high-quality MO-FLART plans in the training cohort. Each training instance's loss function is weighted by the normalized functional dosimetric gain value.

$$W_j = \frac{n_{MO-FLART}(fMLD_j^{MO-ConvRT} - fMLD_j^{MO-FLART})}{\sum_{j \in n_{MO-FLART}}(fMLD_j^{MO-ConvRT} - fMLD_j^{MO-FLART})} \quad (4.1)$$

$$L_{IW}^j = W_j L_{FLART}^j \quad (4.2)$$

The MMDP model was trained using the Adam optimizer with momentum parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$. In the first training stage, the MMDP model was trained to predict conventional RT planning dose maps. The initial learning rate was set to 0.0001 and halved after 100 epochs with the maximum epoch number of 200. The batch size was set to be 1. In the second training stage, the pre-trained MMDP model was trained to predict function-guided dose maps with the anatomical encoder frozen. The learning rate was set to 0.00005, with a maximum epoch number of 50 and a batch size of 1. An early stopping strategy was employed in both stages, halting the training process when the model's performance on the validation dataset no longer showed improvement. The model was developed and trained using PyTorch, utilizing

an NVIDIA RTX 4090D GPU.

4.2.5 Function-guided dose mimicking

A function-guided dose mimicking algorithm, which can incorporate voxel-wise lung function information to assist lung dose painting, was developed to convert function-guided dose maps predicted by the MMDP model into FLART plans [36, 120]. The predicted dose maps were first preprocessed to obtain clinically preferable dose distributions by increasing dose homogeneity of PTV and reducing dose to OARs as formulated by Equation (4.3) [136]. D_i^{pred}, D_i^{pro}, and D_i^{pd} denote predicted dose, processed dose, and prescription dose respectively. The factors f_{PTV} and f_{OARs} control the extent of the preprocessing, set to 0.5 and 0.05 respectively.

$$D_i^{pro} = \begin{cases} (1 - f_{PTV})D_i^{pred} + f_{PTV}D_i^{pd}, i \in PTV \\ (1 - f_{OARs})D_i^{pred}, i \in OARs \\ D_i^{pred}, \ i \in Body - OARs - PTV \end{cases} \quad (4.3)$$

The processed dose maps were then utilized to guide plan generation through dose mimicking, formulated as:

$$\min_x \quad \frac{w_{PTV}}{n_{PTV}} \sum_{i \in PTV} \left(D_i^{cal} - D_i^{pro}\right)^2 + \sum_{s \in S} \frac{w_s}{n_s} \sum_{i \in s} H\left(D_i^{cal} - D_i^{pro}\right)\left(D_i^{cal} - D_i^{pro}\right)^2 + \frac{w_F}{n_{lung}} \sum_{i \in lung} H\left(D_i^{cal} - D_i^{pro}\right) f_i \left(D_i^{cal} - D_i^{pro}\right)^2 \quad (4.4)$$

$$s.t. \quad x \geq 0$$

$$\overrightarrow{D^{cal}} = D\vec{x}$$

Where $\vec{x}$, and $\overrightarrow{D^{cal}}$ represent the fluence map and the calculated dose distribution respectively. f_i represents the functional weight to lung voxel *i*. *w* and *n* refer to the weights and the total number of voxels for various anatomical or functional structures. *H* is the Heaviside function. The weights assigned to PTV, functional lung, spinal cord,

esophagus, heart, and normal tissue were 800, 800, 600, 400, 400, and 200, respectively, reflecting their clinical priority. The voxel-wise function-guided dose mimicking algorithm was implemented based on the open-source treatment planning system (TPS) matRad [117]. Specifically, the pencil beam dose engine was employed to calculate the dose influence matrix. The Interior Point OPTimizer (IPOPT)-based fluence map optimization algorithm was utilized to generate optimal fluence maps by solving the optimization problem (4.4). Static segmentation-based leaf sequencing and direct aperture optimization were then sequentially performed to generate final plans. The dose calculation and plan optimization algorithms used in the dose mimicking process were the same as those used in MO-based planning for consistent comparison.

4.2.6 Dose prediction accuracy evaluation

To evaluate the prediction accuracy of the MMDP model, the predicted dose distributions were compared against planning dose distributions in the testing dataset consisting of 214 high-quality MO-FLART plans. The predicted and planning dose distributions were normalized to ensure mean dose to PTV equal to the prescription dose of 60 Gy for consistent comparison. The evaluation metrics include prediction errors for DVH and DFH metrics of clinical importance, as well as the DVH score and Dose score utilized in the OpenKBP challenge [124]. The DVH score is defined as the MAE of predicted DVH metrics including D_{99}, D_{95} and D_1 for PTV, D_{mean} and $D_{0.1cc}$ for spinal cord, heart, lung, and esophagus. The Dose score is the voxel-wise

MAE of predicted dose maps within the body.

Three ablation studies were conducted to validate the effectiveness of the proposed instance-weighting anatomy-to-function training strategy and multi-modality learning. In the first study, the MMDP model was directly trained from scratch using the sparing MO-FLART plans without the instance weighting loss function (referred to as MMDP_NT). In the second study, the MMDP model was trained through anatomy-to-function training without the instance weighting loss function (referred to as MMDP_A2F). For the final study, the functional encoder and its inputs were removed from the MMDP model, resulting in a traditional anatomy-guided dose prediction (ADP) model. Notably, the channel numbers of the anatomical encoder were increased accordingly to exclude the unwanted influence of parameter numbers to model performance. The ablation models were compared against the full MMDP model trained through instance-weighting anatomy-to-function training. The ablation models were evaluated on sparing plans from the external testing dataset as the MMDP model was trained to predict dose distributions close to sparing FLART plans.

4.2.7 Automatic planning performance evaluation

To assess the quality of FLART plans guided by the MMDP model, the predicted dose maps were converted into FLART plans through function-guided dose mimicking, producing 270 MMDP-FLART plans for 54 patients in the testing cohort. The

MMDP-FLART plans were compared against MO-ConvRT and MO-FLART plans to assess the performance of the proposed automatic lung dose painting framework.

The MMDP-FLART plans were also compared against MA-ConvRT and MA-FLART plans to evaluate the auto-planning performance. All manual plans were created using Eclipse TPS (version 16.1, Varian Medical Systems, Palo Alto, CA). For each testing case, an experienced physicist first created a ConvRT plan following institutional guidelines. With the MA-ConvRT plan as a baseline, a FLART plan was then created by additionally considering lung function information. Another experienced physicist reviewed and approved all manual plans. The details of manual planning are presented at Appendix 6.2. A total of 54 pairs of MA-ConvRT and MA-FLART plans were created and compared against MMDP-FLART plans. All plans were normalized to ensure PTV D_{95} equal to the prescription dose of 60 Gy for consistent comparison. Two critical DFH metrics for radiation pneumonitis prediction including fMLD and F_{20} were evaluated [11].

4.3 Results

4.3.1 Dose prediction accuracy evaluation

Table 3. The prediction errors and prediction absolute errors in relevant DFH and DVH metrics for the MMDP model on the high-quality MO-FLART plan testing dataset.

	Error				Absolute error
	25%	50%	75%	Mean ± SD	Mean ± SD
fMLD (Gy)	-0.25	0.06	0.40	0.10 ± 0.57	0.41 ± 0.40
F_{20} (%)	-0.56	0.23	1.58	0.55 ± 2.02	1.46 ± 1.50
PTV D_{99} (Gy)	-1.05	-0.42	0.21	-0.45 ± 1.01	0.87 ± 0.69
PTV D_{95} (Gy)	-0.91	-0.56	-0.21	-0.56 ± 0.47	0.61 ± 0.41
PTV D_5 (Gy)	0.07	0.28	0.56	0.32 ± 0.33	0.37 ± 0.28
PTV D_1 (Gy)	0.00	0.28	0.63	0.34 ± 0.56	0.48 ± 0.44
Lung D_{mean} (Gy)	-0.27	-0.03	0.24	0.03 ± 0.49	0.36 ± 0.34
Lung V_{20} (%)	-0.59	0.26	1.38	0.42 ± 2.00	1.45 ± 1.45
Lung V_5 (%)	0.06	0.84	2.27	1.42 ± 2.23	1.73 ± 2.00
SpinalCord D_{max} (Gy)	-4.76	-0.88	1.72	-1.03 ± 5.67	4.37 ± 3.76
Esophagus D_{mean} (Gy)	-0.25	0.26	1.53	1.02 ± 2.46	1.63 ± 2.11
Esophagus D_1 (Gy)	-0.68	-0.07	3.26	2.36 ± 6.36	3.95 ± 5.52
Heart D_{mean} (Gy)	-0.03	0.30	1.00	0.43 ± 1.27	0.91 ± 0.98
Heart V_{40} (%)	0.00	0.00	0.26	0.01 ± 0.94	0.41 ± 0.84

Abbreviations: fMLD = functionally weighted mean lung dose; F_x = portion of lung function irradiated by dose higher than x Gy; D_{mean} = mean dose; D_{max} = maximal dose; V_x = portion of structure volume irradiated by dose higher than x Gy; D_x = minimum dose to the "hottest" x% of structure volume.

Table 3 presents the prediction errors of the MMDP model across various clinically important DVH and DFH metrics, with the high-quality MO-FLART plans in the testing cohort as references. Overall, the median prediction errors for all assessed DVH and DFH metrics were within ±1Gy/±1%. The mean DVH score and Dose score were 1.94 Gy and 1.14 Gy respectively. These results indicate that the MMDP model could predict dose distributions close to high-quality FLART plans.

4.3.2 Ablation study results

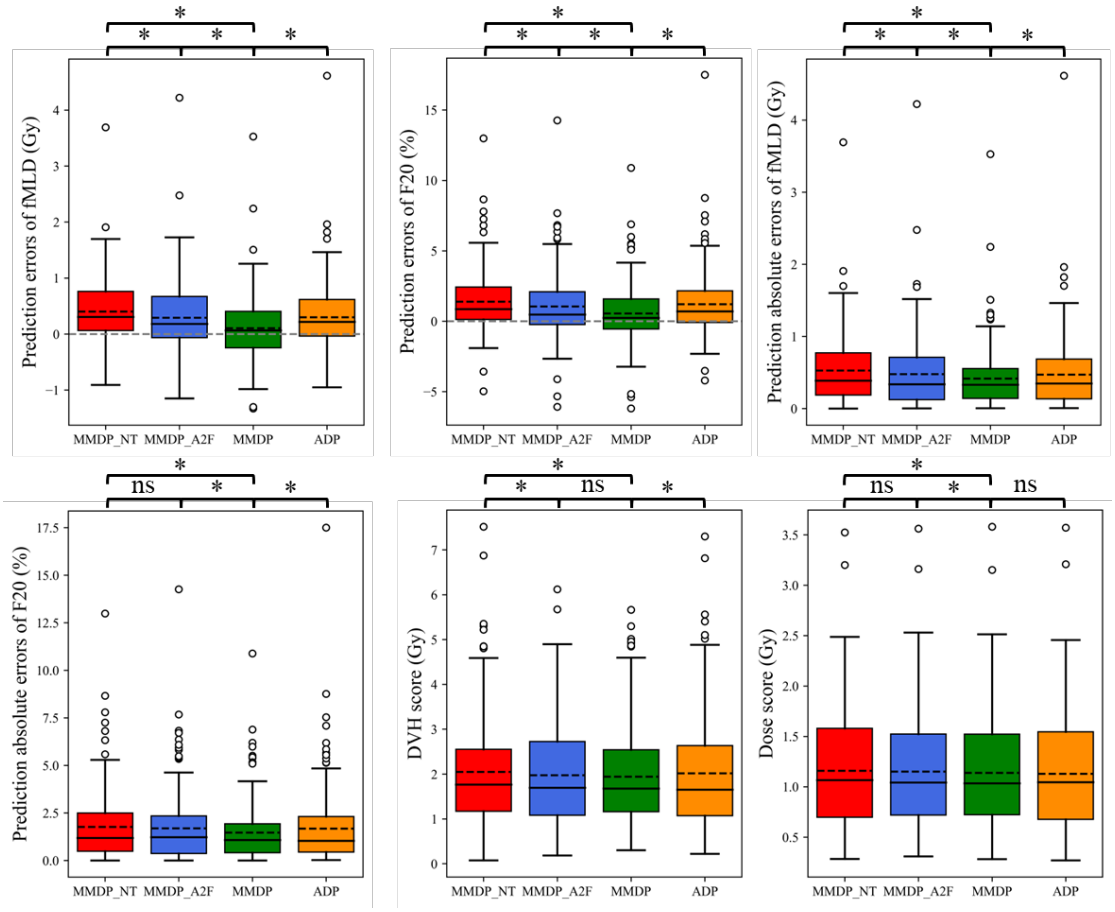


Figure 14. Boxplots summarizing the ablation study results. Prediction errors are defined as prediction results minus planning results. Boxes represent interquartile ranges. Dashed and solid lines within boxes indicate mean and median values respectively. Open circles stand for outlier values. Wilcoxon Signed-Rank test is utilized to calculate significance values, and the results are annotated by ns for $p > 0.05$ and * for $p \leq 0.05$. *Abbreviations*: MMDP_NT = MMDP model trained through naive training; MMDP_A2F = MMDP model trained through anatomy-to-function two-stage training; MMDP = MMDP model trained through instance-weighting anatomy-to-function two-stage training; ADP = Anatomy-guided dose prediction model.

Figure 14 compares the performance of the three ablation models against the MMDP model on the testing dataset. Compared with the MMDP_NT model, prediction errors for fMLD and F_{20} were significantly reduced from 0.40 ± 0.58 Gy and 1.38 ± 2.10% to 0.29 ± 0.62 Gy ($p < 0.01$) and 1.05 ± 2.26% ($p < 0.01$) respectively through anatomy-to-function training (MMDP_A2F). The MMDP model further reduced prediction errors for fMLD and F_{20} to 0.10 ± 0.57 Gy ($p < 0.01$) and 0.55 ± 2.02% ($p < 0.01$) respectively. Compared to the MMDP_NT model, the MMDP model significantly reduced prediction absolute errors for fMLD and F_{20} by 22.64% and 17.05% respectively. These results indicate that the instance-weighting anatomy-to-function training strategy can effectively improve the MMDP's prediction accuracy in function-painted dose distributions. The ADP model exhibited significant overestimations for fMLD (0.30 ± 0.60 Gy, $p < 0.01$) and F_{20} (1.21 ± 2.24%, $p < 0.01$) compared to the MMDP model. Regarding prediction absolute errors for fMLD and F_{20}, the MMDP model achieved an accuracy improvement by 12.77% and 12.57% respectively compared to the ADP model. These results indicate that the MMDP model can effectively extract complementary features from functional inputs, enhancing dose prediction accuracy.

4.3.3 Automatic planning performance evaluation

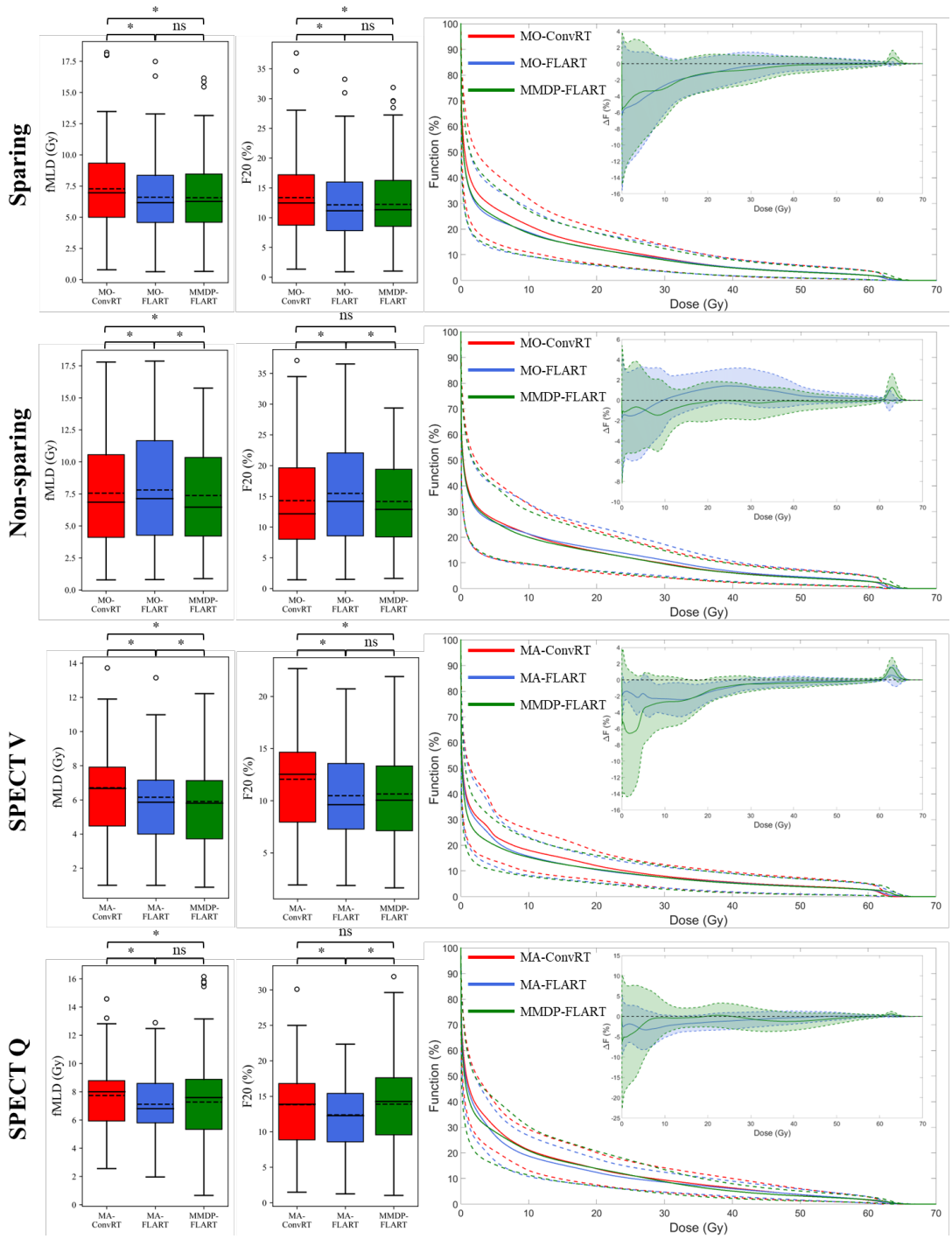


Figure 15. Boxplots and DFH curves summarizing the results of auto-planning performance evaluation on the testing cohort. The first two rows compare MMDP-FLART plans against corresponding MO-ConvRT and MO-FLART plans on the high-quality and low-quality MO-FLART plan datasets respectively. The last two rows

compare MMDP-FLART plans against corresponding MA-ConvRT and MA-FLART plans on the SPECT V and Q datasets respectively. Wilcoxon Signed-Rank test is utilized to calculate significance values, and the results are annotated by ns for $p > 0.05$ and * for $p \leq 0.05$. In the DFH plots, solid and dashed lines represent mean and standard deviation of DFH curves respectively. Insets elucidate the DFH difference between FLART plans and ConvRT plans.

Figure 15 compares MMDP-FLART plans with MO-ConvRT and MO-FLART plans in terms of DFH metrics on the testing cohort. MMDP-FLART plans achieved comparable fMLD and F_{20} compared to high-quality MO-FLART plans, while significantly reducing fMLD and F_{20} compared to low-quality MO-FLART plans. Compared to MA-ConvRT plans, MMDP-FLART effectively reduced fMLD from 7.28 ± 3.35 Gy on the high-quality dataset and 7.56 ± 3.91 Gy on the low-quality dataset to 6.57 ± 3.05 Gy ($p < 0.001$) and 7.37 ± 3.67 Gy ($p = 0.002$) respectively. Overall, MMDP-FLART plans demonstrate comparable or superior performance in reducing DFH metrics compared to MO-FLART plans.

Figure 15 also shows comparisons between MMDP-FLART plans against MA-ConvRT and MA-FLART plans on the SPECT V and Q datasets. MMDP-FLART plans showed a lower fMLD on the SPECT V dataset (5.91 ± 3.04 Gy versus 6.15 ± 3.03 Gy, $p < 0.001$) while a higher F_{20} on the SPECT Q dataset compared to MA-FLART plans (13.90 ± 6.82% versus 12.39 ± 5.37%, $p < 0.001$). Nevertheless, on

SPECT V and Q datasets, MMDP-FLART plans significantly reduced fMLD from 6.71 ± 3.20 Gy and 7.73 ± 2.75 Gy for MA-ConvRT plans to 5.91 ± 3.04 Gy ($p < 0.001$) and 7.27 ± 3.20 Gy ($p < 0.001$) respectively. Detailed comparisons between MMDP-FLART plans and manual plans in terms of clinically important DVH and DFH metrics are provided in Appendix 6.2. While MMDP-FLART plans exhibited inferior homogeneity and higher PTV D_1, they demonstrated better conformity and lower dose to esophagus and heart compared to MA-FLART plans.

Figure 16 illustrates the effectiveness of lung dose painting achieved by MMDP-FLART plans with MA-ConvRT plans as references. As shown in Figure 16(a-b), for both SPECT V and Q datasets, increasing the thresholds defining functional lung volumes resulted in a gradual decrease in the D_{mean} of functional lung volumes for MMDP-FLART plans, with the differences in D_{mean} between MA-ConvRT and MMDP-FLART plans increasing correspondingly. Figure 16(c-d) reveal that dose to HFL volumes were reduced and redirected to LFL volumes. Notably, the reductions in D_{mean} generally increased as the functional values of the lung volumes increased. These results demonstrate that the proposed MMDP-based auto-planning algorithm can achieve effective lung dose painting guided by either SPECT V or Q images.

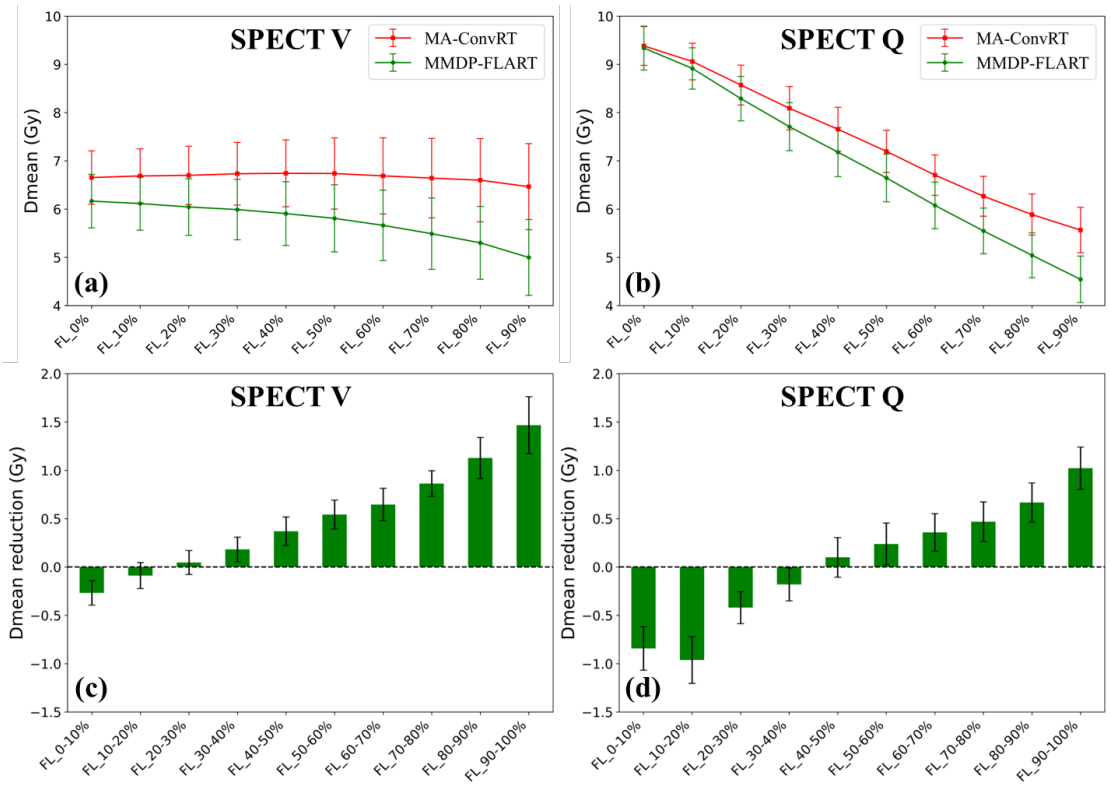


Figure 16. (a-b) Comparison between MA-ConvRT plans and MMDP-FLART plans in mean dose to different functional lung volumes on the SPECT V and SPECT Q datasets. (c-d) Statistical results of mean dose reduction to different functional lung volumes achieved by MMDP-FLART plans compared to MA-ConvRT plans on the SPECT V and SPECT Q datasets. Error bars indicate 95% confidence interval. *Abbreviations*: FL_X% = functional lung volume with functional values higher than X percentile. FL_X-Y% = functional lung volume with functional values ranging from X percentile to Y percentile. MA-ConvRT = manual conventional RT plans; MMDP-FLART = MMDP-based FLART plans.

4.4 Discussion

Substantial retrospective and prospective clinical studies for FLART have been conducted or are currently underway, aiming at validating the clinical benefits [18, 19, 25, 137] and establishing standardized protocols for FLART [11, 22, 102, 123]. While promising results have emerged [18, 19, 25], discrepancies among various studies persist [11, 22, 123, 137]. One potential reason for these controversies is the inter-planner and inter-institution variability in FLART planning. The clinical implementation of FLART is also hindered by the challenging nature of FLART planning induced by the lack of clinical experience and the intra-patient heterogeneity as well as inter-patient variability of lung function distributions [27]. An auto-planning tool for FLART could reduce the inter-planner and inter-institution variability as well as human burdens in FLART planning, assisting research and clinical applications of FLART. In the study of this section, we developed an auto-planning algorithm for FLART consisting of a novel MMDP model and a function-guided dose mimicking algorithm.

The MMDP model can extract complementary features from anatomical and functional information to predict function-guided dose distributions close to high-quality FLART plans. The prediction accuracy for conventional DVH metrics is comparable to, or even superior to, that of ADP models for conventional lung RT previously proposed by *Barragán-Montero et al.* [35] and *Dahiya et al*. [138] Specifically, the prediction absolute errors for lung D_{mean} and V_{20} are 0.36 ± 0.34 Gy and 1.45 ± 1.45% for MMDP, compared to 0.68 ± 0.56 Gy and 2.67 ± 2.87% for the

Barragán-Montero et al.'s model. The DVH and Dose scores are 1.94 Gy and 1.14 Gy for MMDP versus 1.92 Gy and 1.27 Gy for the *Dahiya et al.*'s model. Furthermore, the MMDP model has the advantage of extracting complementary functional features from functional information and promoting prediction accuracy for function-guided dose distributions as evidenced by ablation study results.

Inspired by studies [139-142] utilizing consistent and high-quality automatic plans generated by auto-planning tools to train deep learning-based dose prediction models, we applied the MO-based auto-planning method to generate MO-based plans, and selected sparing plans for MMDP model training. As a result, the MMDP model can predict DFH metrics close to sparing MO-FLART plans. When combined with the function-guided dose mimicking algorithm, the MMDP-FLART plans demonstrated quality comparable to sparing MO-FLART plans and significantly outperformed non-sparing MO-FLART plans in lung dose painting. The MMDP-based auto-planning algorithm also exhibits advantages in implementation complexity and planning efficiency. The MO-based auto-planning algorithm typically requires 0.5 to 1 hour to complete plan hyperparameter optimization for a single plan, even with parallel plan optimization, which is challenging to implement in clinical TPSs. In contrast, MMDP can predict dose distributions with sub-second speed and without demanding parallel plan optimization, making it a promising option for online adaptive radiotherapy planning and instantaneous treatment planning [141]. This strategy can also be applied to develop dose prediction models for other emerging RT techniques or for

cancers with low incidence rates.

Compared to MA-ConvRT plans, MMDP-FLART plans significantly reduced fMLD by 0.80 Gy (11.9%) and 0.46 Gy (6.0%) on the SPECT V and Q datasets respectively, which can potentially enhance lung function protection and reduce pulmonary side effects. Notably, MMDP-FLART plans demonstrate effective lung dose painting on both SPECT V and Q datasets. Most existing FLART clinical trials have adopted the contour-based FLART planning strategy yet different HFL definitions and different lung function mapping techniques [18, 23, 24, 137]. As shown in Figure 16, the lung dose painting strategy can effectively reduce doses to various HFL volumes with different thresholds. Furthermore, the degree of dose reduction is positively correlated with the HFL threshold, aligning with the principles underlying FLART. The MMDP model was trained using CTVI and showed good performance on SPECT V/Q datasets, demonstrating its good generalizability to various lung function mapping techniques.

One limitation of this study is that the performance and preference of the MMDP-based auto-planning algorithm are dependent on the quality and preference of training data. While MMDP-FLART plans exhibited lower fMLD on the SPECT V dataset, they showed higher F_{20} on the SPECT Q dataset compared to MA-FLART plans. Nevertheless, MMDP-FLART plans also showed lower doses to heart and esophagus compared to MA-FLART or even MA-ConvRT plans as presented at Appendix 6.2.

The discrepancy can be attributed to the preference difference between MO-FLART and manual planning. Further improvements to the MMDP model's lung function protection capabilities could be achieved by incorporating training data that prioritize better lung function protection, potentially generated through modified MO-based auto-planning or manual planning. In addition, we have demonstrated the feasibility and efficacy of the automatic lung dose painting algorithm based on the open-source TPS matRad. We will implement this algorithm on clinical TPSs for clinical evaluation and application in the future.

4.5 Conclusion

In this section, we developed a novel multi-modality learning-based automatic planning algorithm for FLART, building on the MO-based method. It comprises a multi-modality-guided dose prediction (MMDP) model and a function-guided dose mimicking algorithm. The MMDP model can extract complementary and useful features from anatomical and functional inputs to predict dose distributions close to high-quality voxel-based FLART plans. An instance-weighting anatomy-to-function training strategy is designed to enhance dose prediction accuracy. The function-guided dose mimicking algorithm can convert predicted dose distributions into FLART plans, demonstrating significant reductions in functional dose metrics and effective lung dose painting compared to manual conventional RT plans on SPECT V and Q datasets. The MMDP-based auto-planning algorithm outperforms the MO-based algorithm in planning quality and efficiency.

5 Towards clinical application: Clinical implementation and evaluation of MO-based and MMDP-based auto-planning algorithms for FLART

5.1 Introduction

In the preceding sections, two auto-planning algorithms for Functional Lung Avoidance Radiotherapy (FLART) were developed and their feasibility and effectiveness were preliminarily demonstrated using the open-source treatment planning system (TPS), matRad. Although matRad is widely utilized in research settings due to its open-source accessibility, it has not been extensively adopted in clinical practice. The initial evaluation of the proposed auto-planning algorithms was conducted using various quantitative dose metrics, as presented in Table 2, Table 17, and Table 18. However, while these metrics provide objective measures of plan quality, they do not fully encompass clinical acceptability or preferences.

To facilitate the clinical translation of the proposed auto-planning algorithms, this section focuses on their clinical implementation and evaluation. The proposed auto-planning algorithms will be integrated into clinical TPS. The clinical performance and clinical benefits of the auto-planning algorithms will be assessed through both quantitative dose metrics and qualitative blind review. Specifically, automatic FLART plans will be compared with manual FLART and conventional radiotherapy (ConvRT) plans in terms of quantitative dose metrics and normal tissue complication probabilities (NTCPs) for both anatomical and functional OARs, thereby evaluating plan quality and the effectiveness of lung function protection. Furthermore, the

clinical acceptability and preference for automatic versus manual FLART plans will be reviewed and ranked by experienced clinicians to comprehensively assess the clinical performance of the two auto-planning algorithms. The planning time associated with the two auto-planning algorithms and manual planning will be compared to evaluate planning efficiency. Additionally, the FLART planning results by a junior planner with and without the assistance of the developed AP-FLART system will be compared, aiming at evaluating the potential of the system in facilitating the broader clinical adoption of the emerging FLART technique.

5.2 Method

5.2.1 Patient data and preprocessing

Table 4. Patient and plan data used in this study

Dataset	Dataset #1	Dataset #2	Dataset #3	Dataset #4
Source	GDP-HMM Dataset[143]	HNC Hospital & 4D-Lung Dataset[129]	Stanford University[130]	HNC Hospital
Patient number	753	142	21	62
Function map	None	CTVI	SPECT V	SPECT Q
Plan number	753 ConvRT plans	710 ConvRT plans & 508 FLART plans	21 pairs of ConvRT/FLART plans	62 pairs of ConvRT/FLART plans
Planning method	AIRTP-based auto-planning	MO-based auto-planning	Manual planning	
Utilization	MMDP model pretraining		MMDP model fine-tuning; Clinical evaluation	

Note: HNC Hospital = Henan Cancer Hospital; CTVI = CT-derived Ventilation Imaging; AIRTP = Automated Iterative RT Planning[143]

As summarized in Table 4, a total of four datasets comprising 2,137 plans from 978 patients were utilized for the development of the MMDP model and the clinical evaluation of both MMDP-based and MO-based auto-planning algorithms. The first dataset includes all lung IMRT plans from the GDP-HMM dataset, which were generated using the Automated Iterative Radiotherapy Planning (AIRTP)-based auto-planning algorithm [143]. The second dataset consists of planning CT images and contours for OARs and PTV, collected from Henan Cancer Hospital (122 patients) and the 4D-Lung dataset (20 patients). The super-voxel-based CTVI algorithm was employed to generate CTVI from the planning CT images. Conventional radiotherapy (ConvRT) and FLART plans were generated from planning CT images, contours, and

CTVI using the MO-based auto-planning algorithm. Further details regarding the second dataset are provided in Section 4.2.2.

The third dataset is a subset of the VAMPIRE dataset [58], consisting of planning CT images, contours, and SPECT ventilation images for 21 patients. The fourth dataset is prospectively collected from Henan Cancer Hospital containing planning CT images, contours, and SPECT perfusion images for 62 lung cancer patients underwent RT. For each patient from the third and fourth datasets, a senior physicist manually created a pair of ConvRT and FLART plans. Another senior physicist reviewed and approved all manual plans. The details regarding manual planning are presented at Appendix 6.3. Given the plan data from the first and second datasets are generated from auto-planning algorithms without clinical review, they are used for pretraining the MMDP model, enhancing its generalizability across different patients. The well-curated manual plans from the third and fourth datasets are used for fine-tuning the MMDP model and evaluating the clinical performance of the MMDP-based and MO-based auto-planning algorithms for FLART.

Figure 17 depicts the workflow of fusing lung function images into planning CT images for guiding FLART planning. For patients in the fourth dataset, rigid registration was performed to register the attenuation correction CT images from perfusion SPECT/CT scans with corresponding planning CT images based on MIM platform (MIM Software Inc., Cleveland, OH, USA). The resulting transformations

were directly utilized to align SPECT Q images with planning CT images. Trilinear interpolation was then performed to match the dimensions of SPECT Q images with those of planning CT images. The SPECT V images from the third dataset have been registered and interpolated to match the locations and dimensions of corresponding planning CT images in the VAMPIRE challenge. Following the VAMPIRE challenge [58], an iterative thresholding-based method [131] was adopted to mitigate the influence of spurious ventilation values from the SPECT V images. The registered SPECT images were normalized into percentile functional maps by replacing the functional value of each voxel with the corresponding cumulative distribution function value, ranging from the 0^{th} to 100^{th} percentile, as described in previous studies [19, 144, 145]. The whole lung, excluding the PTV, was divided into ten functional lung contours according to the percentile functional map, which were then used to guide FLART planning and approximate voxel-wise FLART planning. Each functional lung contour, denoted as "FL_i*10_(i+1)*10" with $i \in [0,1,2,\ldots,9]$ represents lung regions (excluding PTV) with functional values ranging from the $i*10^{th}$ percentile to $(i+1)*10^{th}$ percentile.

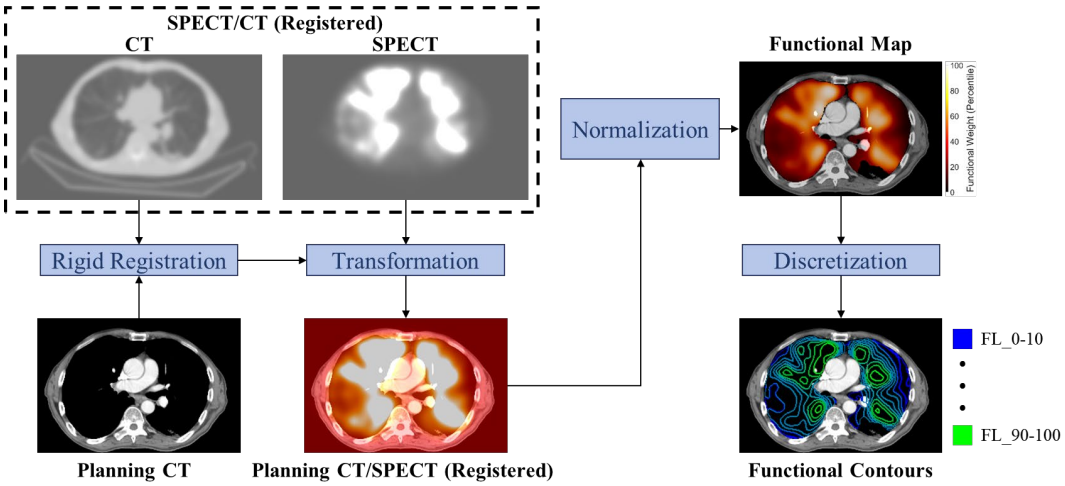


Figure 17. The flowchart of data preprocessing including the registration between SPECT and planning CT as well as the generation of functional contours.

5.2.2 Clinical implementation of MMDP-based auto-planning

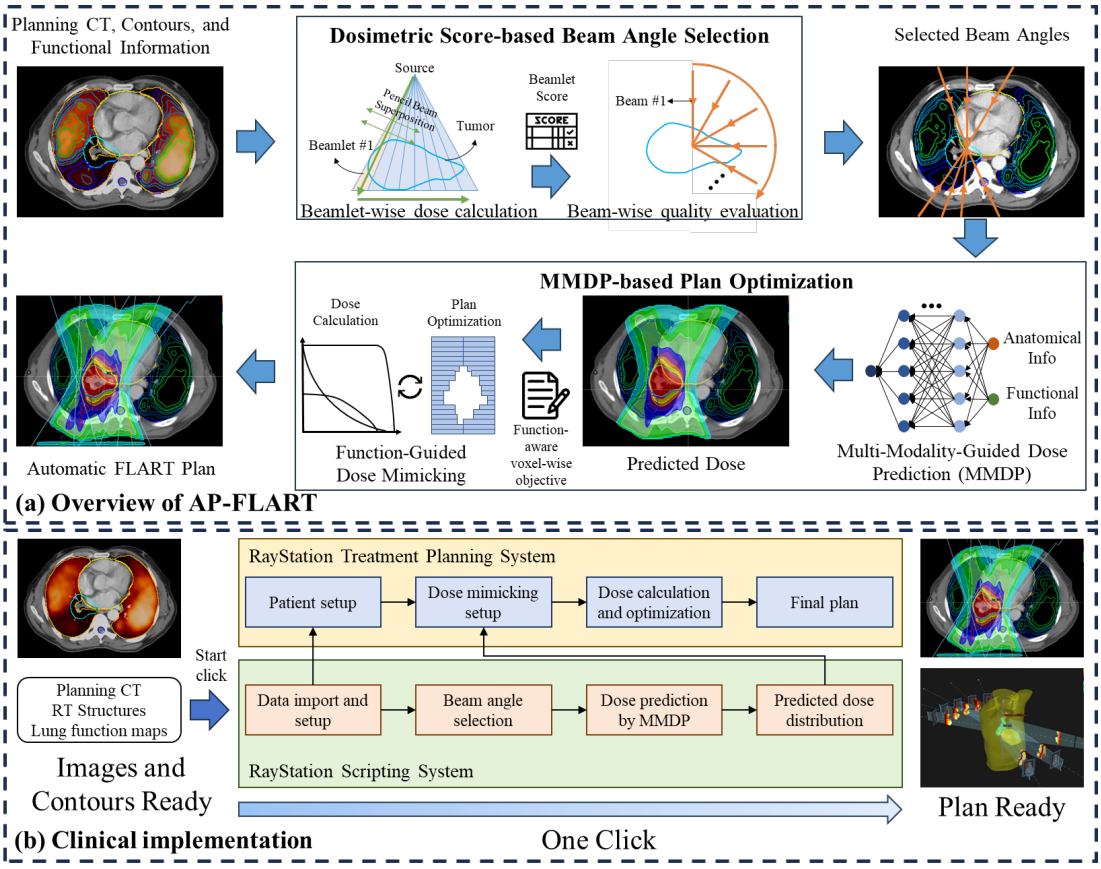


Figure 18. Overview and workflow of the multi-modality-guided dose prediction

(MMDP)-based auto-planning algorithm implemented in clinical treatment planning system RayStation (RaySearch Laboratories, Stockholm, Sweden).

As illustrated in Figure 18(a), the MMDP-based auto-planning algorithm comprises two primary components: a dosimetric score-based beam angle selection algorithm and an MMDP-based automatic plan optimization algorithm. The dosimetric score-based beam angle selection algorithm quantifies the quality of each candidate beam angle through beamlet-wise dose calculations, enabling the selection of beam angles that optimally cover the tumor while sparing both anatomical and functional healthy tissues. Further details regarding this algorithm are provided in Section 3.2.3. The MMDP model predicts optimal dose distributions for FLART based on patient-specific anatomical and functional data, as well as selected beam angles. Subsequently, a function-guided dose mimicking algorithm is employed to convert the predicted dose distributions into clinically deliverable FLART plans. The specifics of the MMDP-based automatic plan optimization process are described in Sections 4.2.4 and 4.2.5.

The MMDP-based auto-planning algorithm has been implemented within the clinical TPS RayStation for the purposes of clinical evaluation and application. Figure 18(b) depicts the workflow of the MMDP-based auto-planning algorithm and its integration with the RayStation TPS. Both the dosimetric score-based beam angle selection and the MMDP-based automatic plan optimization algorithms are implemented via the

RayStation scripting system, facilitating direct interaction with the dose calculation and plan optimization engines of RayStation. The required inputs for the auto-planning system include planning CT images, registered lung function images, and RT contours for OARs and PTV. Lung function contours are automatically generated and incorporated into the RT structure set for FLART planning. The preprocessed planning CT images, along with anatomical and functional contours, are automatically imported into the TPS to initiate plan generation. The workflow proceeds as follows: the beam angle selection algorithm is invoked to evaluate and select optimal beam angles; the MMDP algorithm is then used to predict the optimal dose distribution based on anatomical, functional, and beam configuration data. The predicted dose distribution is preprocessed and utilized to establish optimization objectives for dose mimicking. Subsequently, the RayStation plan optimization engine is called to generate deliverable FLART plans guided by these dose mimicking objectives. In summary, this fully automated planning system enables the generation of FLART plans from prepared images and contours with a single click. A demo for operating the MMDP-based auto-planning algorithm is available at the link provided.

While Section 4 has demonstrated the feasibility of MMDP-based auto-planning for FLART, this section details its integration into a clinical TPS and describes enhancements to its clinical performance through improvements in model architecture, utilization of higher-quality manual plans for model training, and refinement of dose mimicking settings. Further details are provided in Appendices 6.3.1 and 6.3.2.

5.2.3 Clinical implementation of MO-based auto-planning

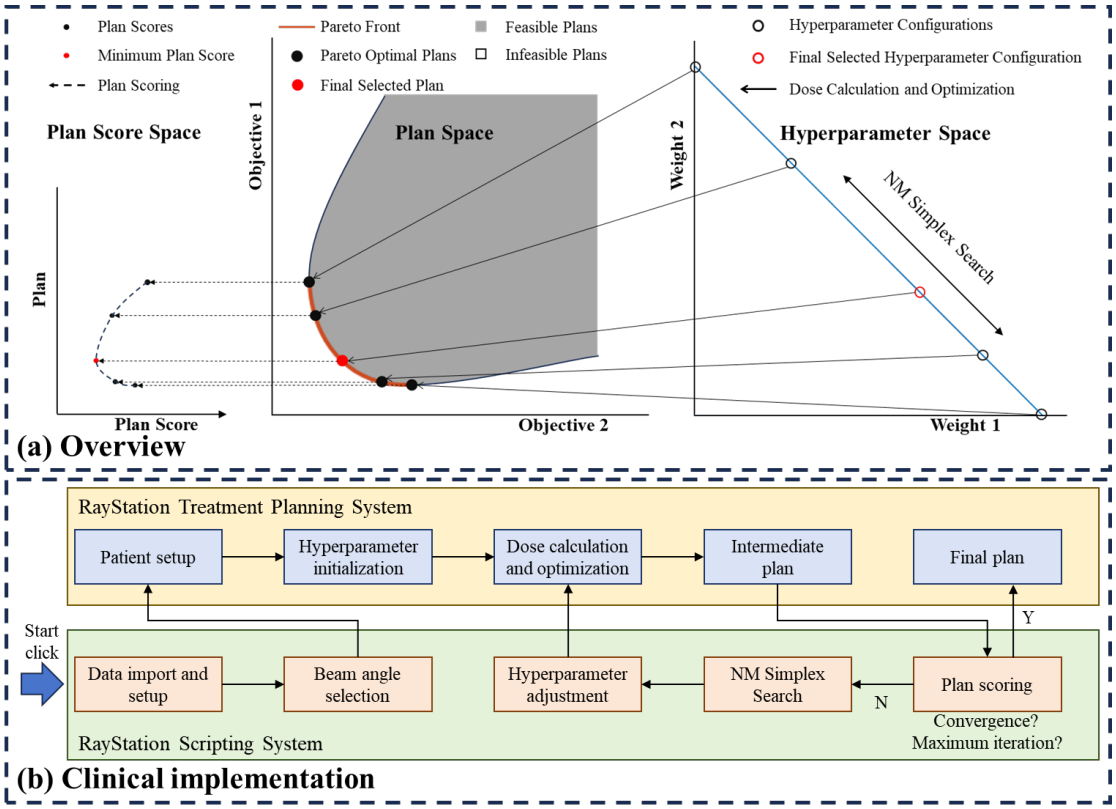


Figure 19. Overview and workflow of the meta-optimization (MO)-based auto-planning algorithm implemented in clinical treatment planning system RayStation

Figure 19(a) provides an overview of the MO-based auto-planning algorithm for FLART. The MO algorithm is designed to navigate the Pareto front, enabling the identification of treatment plans that optimally balance conventional and functional dose objectives. This balance is guided by a composite plan score that incorporates key conventional and functional plan quality metrics. Further details regarding the MO algorithm are provided in Section 3.2.4.

The MO-based auto-planning algorithm has been implemented in the RayStation TPS for clinical evaluation and application. Figure 19(b) depicts the workflow of the MO-based auto-planning algorithm and its integration with RayStation TPS. The processes for data preprocessing and beam angle selection are identical to those employed in the MMDP-based auto-planning algorithm. Following sequential data preprocessing, data import, and beam angle selection, the MO-based automatic plan optimization algorithm is invoked to iteratively evaluate intermediate plans and adjust the plan objective hyperparameters. The quality of each intermediate plan is assessed by calculating a plan score, which is defined by clinically significant conventional and functional plan quality metrics. Hyperparameters are iteratively adjusted with the objective of minimizing the plan score, utilizing the Nelder-Mead Simplex Search algorithm. The process of hyperparameter initialization/adjustment, intermediate plan optimization, and plan score calculation is repeated until the plan score converges or the maximum number of iterations is reached. To approximate voxel-wise FLART planning, ten functional lung contours were utilized, as detailed in Appendix 6.3.3.

For MMDP-based and MO-based auto-planning, the Collapse Cone dose engine (version 5.9) was employed for dose calculation. Plan generation was performed using a TrueBeam linear accelerator, with a dynamic IMRT technique and a beam energy of 6 MV, consistent with manual planning detailed in Appendix 6.3.4.

5.2.4 Clinical evaluation

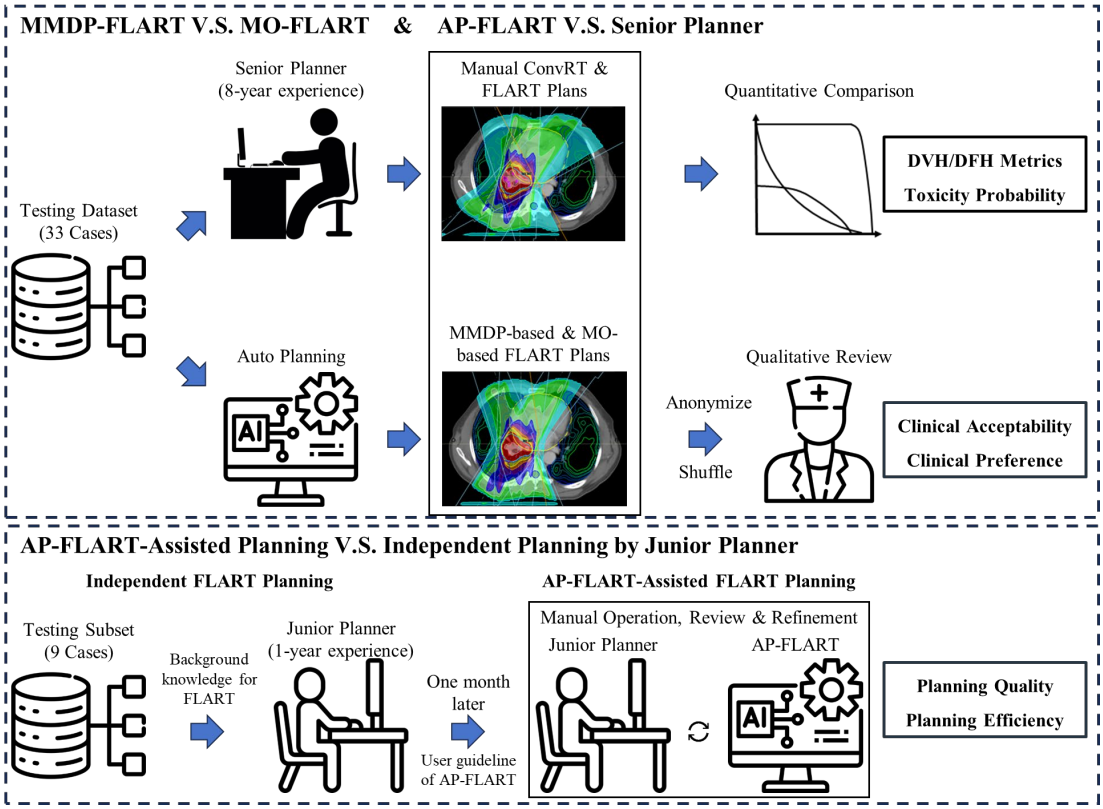


Figure 20. Overview of the clinical evaluation flowchart, illustrating paired comparisons between: (1) the MMDP-based and MO-based auto-planning systems; (2) the MMDP-based auto-planning system and an experienced senior planner (8 years of clinical experience); and (3) MMDP-FLART-assisted planning and independent planning by a junior planner (1 year of clinical experience).

As demonstrated by Figure 20, the clinical evaluation process comprised two distinct phases. The first phase compared the planning performance of the MMDP-based and MO-based auto-planning algorithms against that of a senior planner to assess the clinical performance and potential benefits of the proposed auto-planning systems. The second phase compared the outcomes of AP-FLART-assisted planning with those

of independent planning by a junior planner, aiming to determine whether the developed AP-FLART system enhances FLART planning quality and efficiency in a real-world clinical setting. All plans were normalized to ensure prescription dose cover 95% of PTV before qualitative and quantitative evaluation for fair comparison.

In the first evaluation phase, quantitative comparison and qualitative review were conducted to evaluate the clinical performance and clinical benefits of the MMDP-based and MO-based auto-planning algorithms. For each case in the testing dataset consisting of 33 cases, manual planning and auto-planning were performed to produce four plans: a manual ConvRT plan, a manual FLART plan, an MMDP FLART plan, and an MO FLART plan. The produced plans were compared based on DVH/DFH metrics, estimated toxicity probabilities and blind review by three senior clinicians.

The clinically important DVH and DFH metrics for the produced plans were calculated to quantitatively assess the dosimetric performance of the auto-planning algorithms. Furthermore, normal tissue complication probability (NTCP) models were adopted to estimate probabilities of radiation-induced toxicities for quantifying the lung function protection effectiveness and potential damage to other OARs. Table 5 lists all the evaluated radiation-induced toxicities and corresponding NTCP models. The functional NTCP model established by *Faught et al.* [146] was utilized for the estimation of radiation pneumonitis, while NTCP models recommended by the QUANTEC guidelines [147] were employed for the estimation of other toxicities.

It should be noted that the clinical benefit from FLART may vary among patients, depending on their individual anatomical and functional distributions [9]. Specifically, patients with homogeneous lung function [9], tumors located far from high-function regions, or small tumor volumes [18] may gain limited benefit from FLART, as it is either challenging to spare high-function regions or the dose to these regions is inherently low. Therefore, appropriate patient selection is essential to identify those who are most likely to benefit from FLART. To evaluate the potential clinical benefits of the proposed auto-planning algorithm in patients likely to benefit from FLART, "FLART-benefiting" patients within the testing cohort were identified based on manual planning results. Patients were classified as FLART-benefiting if the difference in the probability of radiation pneumonitis (Grade ≥ 2) between manual ConvRT and manual FLART plans exceeded 3 percentage points (pp). The potential clinical benefits of the auto-planning algorithm for these FLART-benefiting patients were also estimated using the NTCP models.

Table 5. NTCP models used for estimating probabilities of radiation-induced toxicities

OARs	Complication	NTCP model	Parameters
Lung	Radiation Pneumonitis (Grade ≥ 2)	HFL V20 LKB model [146]	TD50 = 35.9 %; m = 0.548
Esophagus	Acute Esophagitis (Grade ≥ 2)	LKB model [148]	TD50 = 47 Gy; n = 0.69; m = 0.36

Spinal Cord	Myelopathy	LKB model [149]	TD50 = 66.5 Gy; n = 0.05; m = 0.175
Heart	Cardiac Mortality	RS model [150]	TD50 = 70.3 Gy; k = 0.96; s = 1.0

Note: The functional NTCP model established by *Faught et al.* [146] was adopted for radiation pneumonitis estimation and the NTCP models recommended by the QUANTEC guideline [147] were used for estimating the other toxicities. LKB model = Lyman-Kutcher-Burman model; RS model = relative seriality model

Qualitative review was conducted by one medical physicist and two senior radiation oncologists to assess the clinical acceptability and preference of the automatic FLART plans. Specifically, for each testing case, three FLART plans (manual FLART, MMDP FLART, and MO FLART) were anonymized and shuffled. One senior physicist was invited to review the three FLART plans with the manual ConvRT plan as a reference. The physicist scored the clinical acceptability of each reviewed plan based on the criteria in Table 6. The physicist also ranked the three plans according to Table 7 to compare the clinical preference. Furthermore, two senior radiation oncologists were invited to independently and blindly score the clinical acceptability of manual and MMDP FLART plans and compare clinical preferences of the two FLART plans.

Table 6. Likert scale to score the clinical acceptability of treatment plans.

Score	Interpretation
5	Use as-is. Clinically acceptable. Plans could be used for treatment without change
4	Minor edits are not necessary. Stylistic changes preferred, but not clinically important. Current plan is clinically acceptable.
3	Minor edits are necessary. Minor edits are those that can be made in less time than starting from scratch or are expected to have minimal effect on treatment outcome.
2	Major edits. Necessary edits are required to ensure appropriate treatment and are sufficiently significant that the user would prefer to start from scratch.
1	Unusable. Quality of the plan is so bad it is unusable.

Note: This table was adopted from [151].

Table 7. Assessment table used for blindly comparing FLART plans generated from manual planning, MO-based auto-planning, and MMDP-based auto-planning.

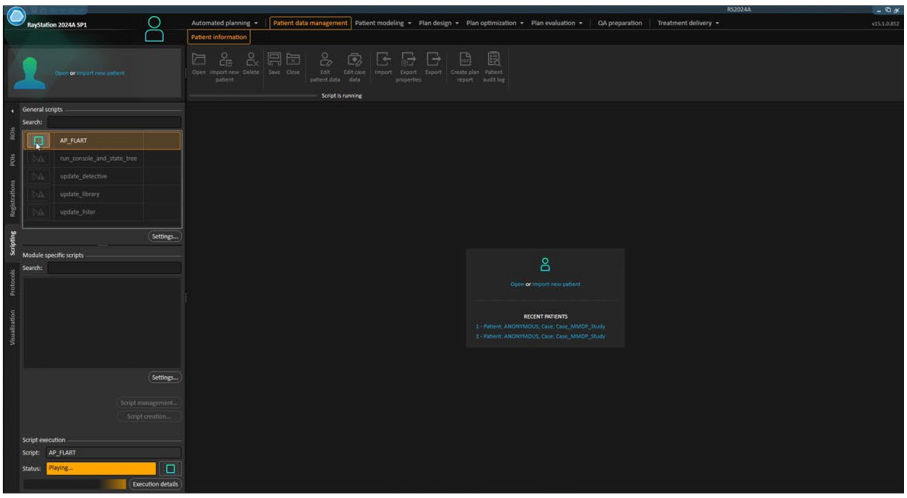
Evaluation metrics	Rank of three FLART plans		
	Plan#1	Plan#2	Plan#3
Lung function protection			
Overall quality			

To further assess the clinical utility of the developed auto-planning system in a real-world setting, a junior planner with no prior experience in FLART was invited for the second evaluation phase. Initially, the junior planner received foundational knowledge about FLART principles and planning strategies and was subsequently tasked with independently generating FLART plans for nine testing cases. The independent planning outcomes served as the baseline for evaluating the clinical value of the AP-FLART system. Following a one-month washout period to mitigate recall bias regarding the specific testing cases, the junior planner was instructed to generate FLART plans for the same cohort, this time utilizing the MMDP-based auto-planning system. Guided by a user guideline document for the AP-FLART system, the planner generated automated FLART plans, which they subsequently reviewed and refined to produce the final treatment plans. Finally, overall plan quality and planning efficiency were quantitatively compared between the independent manual workflow and the AP-FLART-assisted workflow.

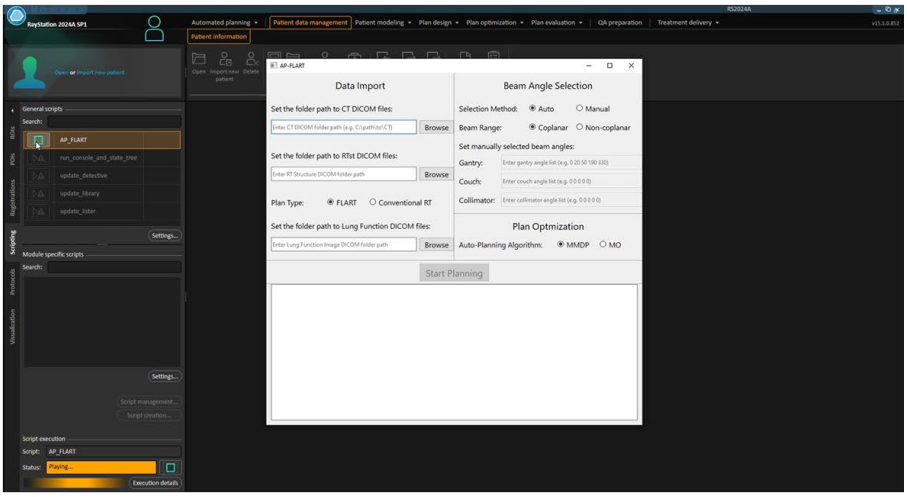
5.3 Results

5.3.1 Demo of our auto-planning system for FLART

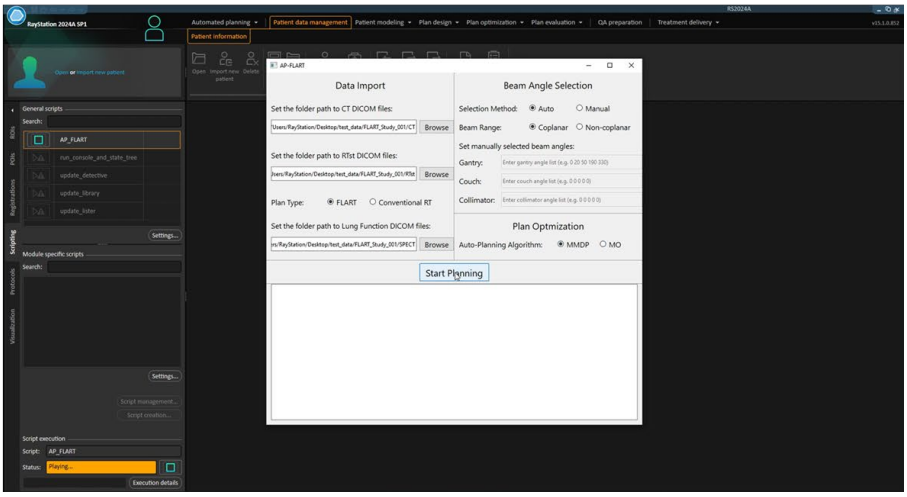
(a) Launch the AP-FLART system integrated in RayStation TPS



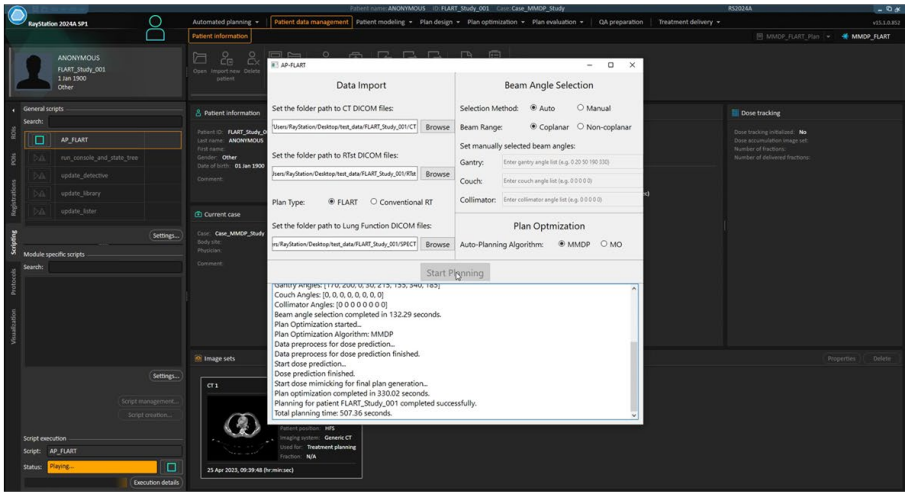
(b) User interface of AP-FLART



(c) Import data and start auto-planning



(d) Planning completed



(e) Produced FLART plans in RayStation

Figure 21. Demonstration of the workflow of the clinically ready auto-planning system for FLART (AP-FLART). A demo is available at the link provided.

As shown in Figure 21 and the attached demo, a clinically ready auto-planning system for FLART has been developed. It can achieve automatic beam angle selection and plan optimization for FLART. Two automatic plan optimization algorithms, one based on MMDP and another based on MO, are available. This system enables the generation of FLART plans with a single click, given planning CT images, RT contours, and registered lung function images. The plans are generated using the

clinically validated plan optimization and dose calculation engines of RayStation, ensuring both accuracy and safety. The integration with RayStation allows clinicians to review, revise, and approve the generated plans, and finally delivery for patient treatment. The clinical implementation enables the clinical evaluation and application of the proposed auto-planning algorithms.

5.3.2 Comparison of DVH and DFH metrics

Table 8 presents a comparison of MMDP FLART plans against manual ConvRT plans, manual FLART plans, and MO FLART plans on the testing dataset, focusing on clinically significant DVH and DFH metrics. Compared to manual ConvRT plans, FLART plans generated by manual planning, MMDP-based auto-planning, and MO-based auto-planning all effectively achieved lung dose painting and reduced functional dose metrics. Specifically, MMDP FLART plans significantly reduced the fMLD and HFL mean dose (D_{mean}) by 0.98 Gy (11.8%) and 1.25 Gy (15.1%), respectively. Additionally, F_{20} and HFL V_{20} were significantly reduced by 2.62 percentage points (pp) (16.6%) and 3.19 pp (21.1%), respectively. However, MMDP FLART plans exhibited higher spinal cord $D_{0.1cc}$ and heart V_{40}, as well as inferior PTV homogeneity, which are considered acceptable trade-offs for enhanced lung function protection.

When compared to manual FLART plans, MMDP FLART plans demonstrated similar DFH and DVH metrics, except for a higher spinal cord $D_{0.1cc}$ and slightly improved

PTV conformity. In comparison to MO FLART plans, MMDP FLART plans exhibited significantly superior lung dose painting performance, albeit at the expense of higher doses to the spinal cord, esophagus, and heart.

Table 8. Comparison between MMDP FLART plans against Manual ConvRT, Manual FLART, and MO FLART plans on the testing dataset. The mean and standard deviation of various DFH and DVH metrics are listed. The p-values by Wilcoxon Signed-Rank test are also shown.

	MMDP FLART	Manual ConvRT	p-value	Manual FLART	p-value	MO FLART	p-value
fMLD (Gy)	7.34 ± 2.92	8.32 ± 2.91	**<0.01**	7.36 ± 2.70	0.81	7.51 ± 3.28	0.19
F_{20} (%)	13.14 ± 6.03	15.76 ± 6.68	**<0.01**	12.78 ± 5.50	0.44	13.97 ± 6.90	**0.02**
F_5 (%)	31.35 ± 13.76	34.43 ± 12.35	**0.02**	31.53 ± 11.65	0.75	32.18 ± 15.94	0.46
HFL D_{mean} (Gy)	7.03 ± 3.55	8.28 ± 3.64	**<0.01**	7.04 ± 3.26	0.89	7.31 ± 3.95	**0.03**
HFL V_{20} (%)	11.94 ± 7.26	15.13 ± 8.29	**<0.01**	11.39 ± 6.57	0.11	13.02 ± 8.09	**<0.01**
HFL V_5 (%)	28.50 ± 15.13	32.46 ± 13.90	**0.01**	29.17 ± 13.00	0.34	29.52 ± 17.58	0.22
HI	7.66 ± 1.84	6.93 ± 1.42	**0.04**	7.90 ± 1.25	0.34	7.49 ± 2.51	0.85
CI	0.78 ± 0.11	0.79 ± 0.07	0.81	0.76 ± 0.07	0.07	0.76 ± 0.14	0.22
PTV D_{1cc} (Gy)	65.46 ± 1.07	64.88 ± 0.92	**<0.01**	65.62 ± 0.75	0.30	65.56 ± 1.84	0.82
Lungs D_{mean} (Gy)	8.15 ± 2.98	8.81 ± 2.87	**<0.01**	8.12 ± 2.80	0.71	8.18 ± 3.24	1.00
Lungs V_{20} (%)	15.17 ± 6.11	17.06 ± 6.53	**<0.01**	14.84 ± 5.83	0.35	15.71 ± 6.85	0.16
Lungs V_5 (%)	33.44 ± 13.35	35.71 ± 11.59	**0.03**	33.24 ± 11.23	0.89	33.92 ± 15.17	0.85
SpinalCord $D_{0.1cc}$ (Gy)	37.01 ± 10.58	34.40 ± 9.08	**<0.01**	35.53 ± 9.91	**0.01**	35.91 ± 10.47	**0.05**
Esophagus D_{mean} (Gy)	13.21 ± 7.88	12.82 ± 7.82	0.27	13.13 ± 8.05	0.81	11.88 ± 8.76	**<0.01**
Esophagus $D_{0.1cc}$ (Gy)	50.26 ± 18.94	49.15 ± 20.43	0.54	49.73 ± 20.26	0.54	43.69 ± 25.49	**0.01**
Heart D_{mean} (Gy)	8.34 ± 6.37	7.96 ± 5.93	0.14	8.27 ± 6.51	0.45	7.35 ± 5.82	**<0.01**
Heart V_{40} (%)	4.25 ± 4.11	3.47 ± 3.59	**<0.01**	4.13 ± 4.07	0.51	2.78 ± 3.01	**<0.01**

Abbreviations: fMLD = functionally weighted mean lung dose; F_x = portion of lung function irradiated by dose higher than x Gy; D_{mean} = mean dose; D_{max} = maximal dose; V_x = portion of structure volume irradiated by dose higher than x Gy; D_{xcc} = minimum dose to the "hottest" x cc of structure; HFL = high-function lung with functional weight higher than 66^{th} percentile.

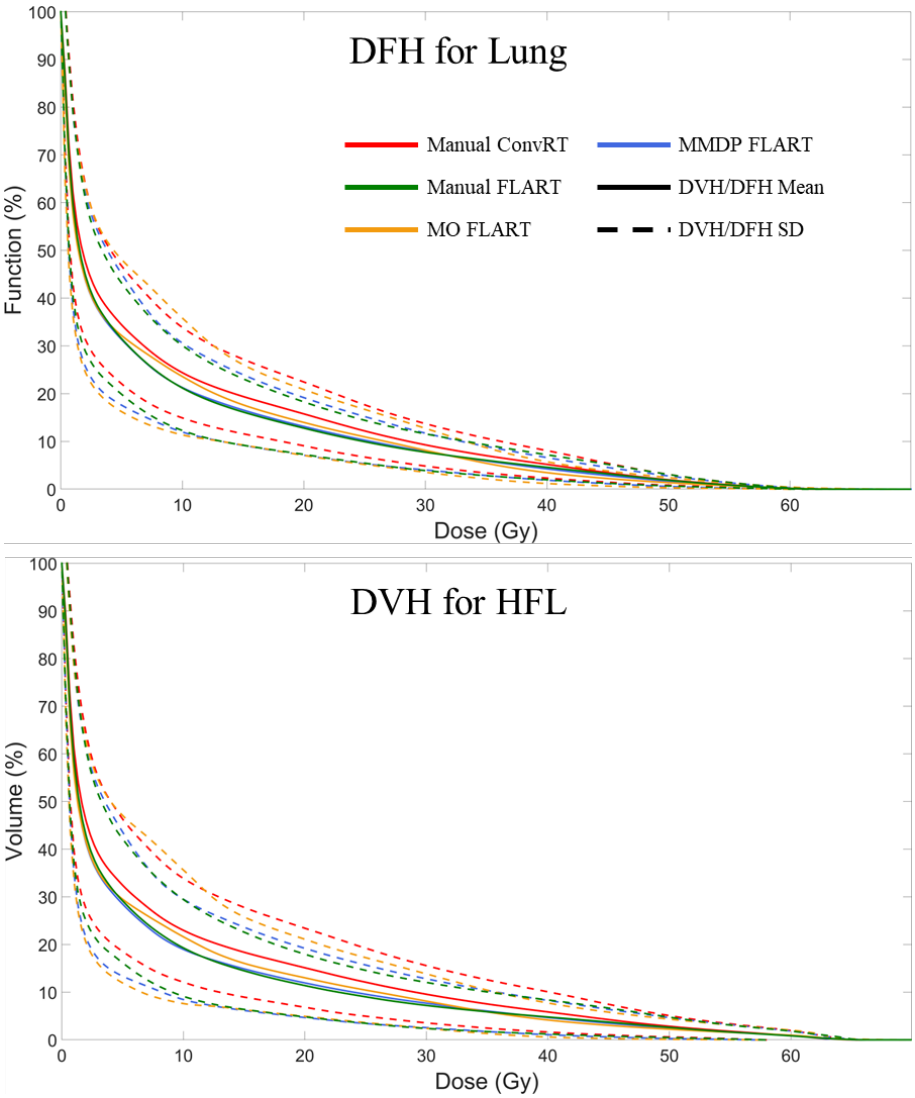


Figure 22. Comparison of DFH curves and HFL DVH curves for manual ConvRT plans, manual FLART plans, MO FLART plans, and MMDP FLART plans. The solid lines and dashed lines represent the means and the standard deviations over all testing cases.

Figure 22 compares the DFH curves and HFL DVH curves for manual ConvRT plans, manual FLART plans, MO FLART plans, and MMDP FLART plans. Compared to manual ConvRT plans, all FLART plans effectively achieved lung dose painting. Notably, MMDP FLART plans demonstrated lung function painting performance

comparable to manual FLART plans and significantly outperformed MO FLART plans.

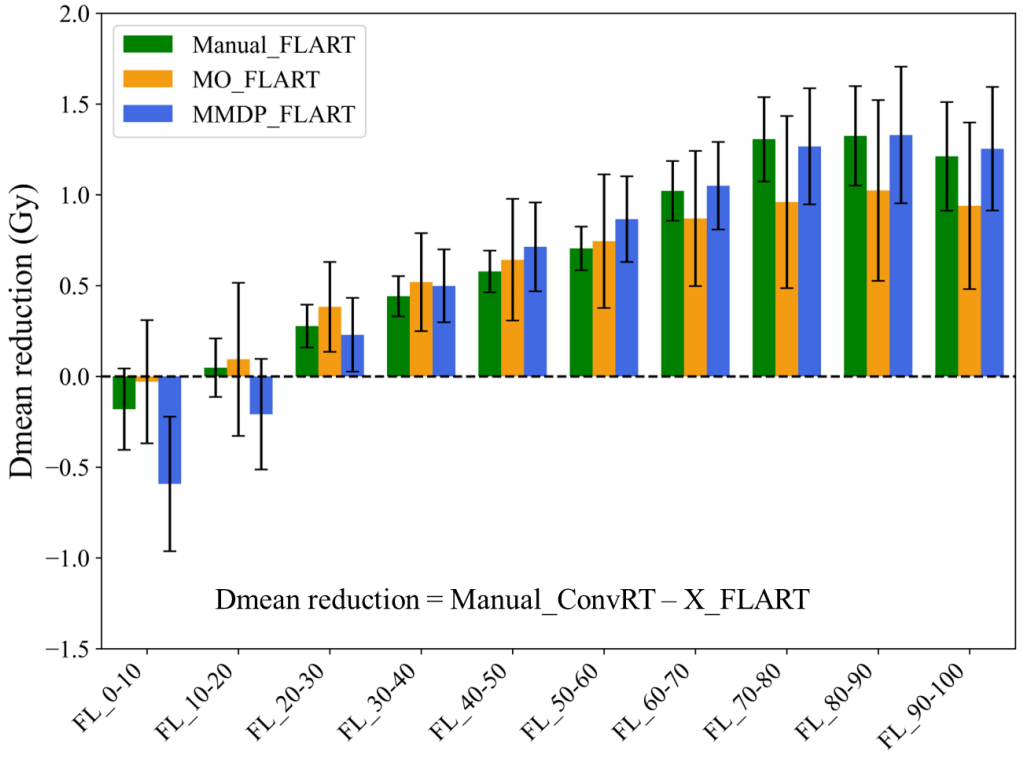


Figure 23. Statistical results of mean dose reduction to different functional lung volumes achieved by Manual FLART plans, MO FLART plans, and MMDP FLART plans compared to Manual ConvRT plans on the testing dataset. Error bars indicate 95% confidence interval. *Abbreviations*: FL_X-Y = functional lung volume with percentile functional weights ranging from X percentile to Y percentile.

Figure 23 illustrates the effectiveness of lung dose painting achieved by FLART plans generated via manual planning, MO-based auto-planning, and MMDP-based auto-planning. All FLART plans effectively reduced the dose to high-function lung regions

compared to manual ConvRT plans. Importantly, the reductions in D_{mean} generally increased with higher functional value, indicating effective lung dose painting. Compared to manual FLART and MO FLART plans, MMDP FLART plans redirected more dose to low-function lung regions (FL_0-20).

Figure 24 demonstrates the lung dose painting effectiveness of MMDP-based auto-planning for a representative testing case by comparing the dose distributions of its manual ConvRT plan and MMDP FLART plan. As shown in the dose difference map, the MMDP FLART plan effectively reduced doses to high-function lung regions and redirected them to low-function lung regions and other areas.

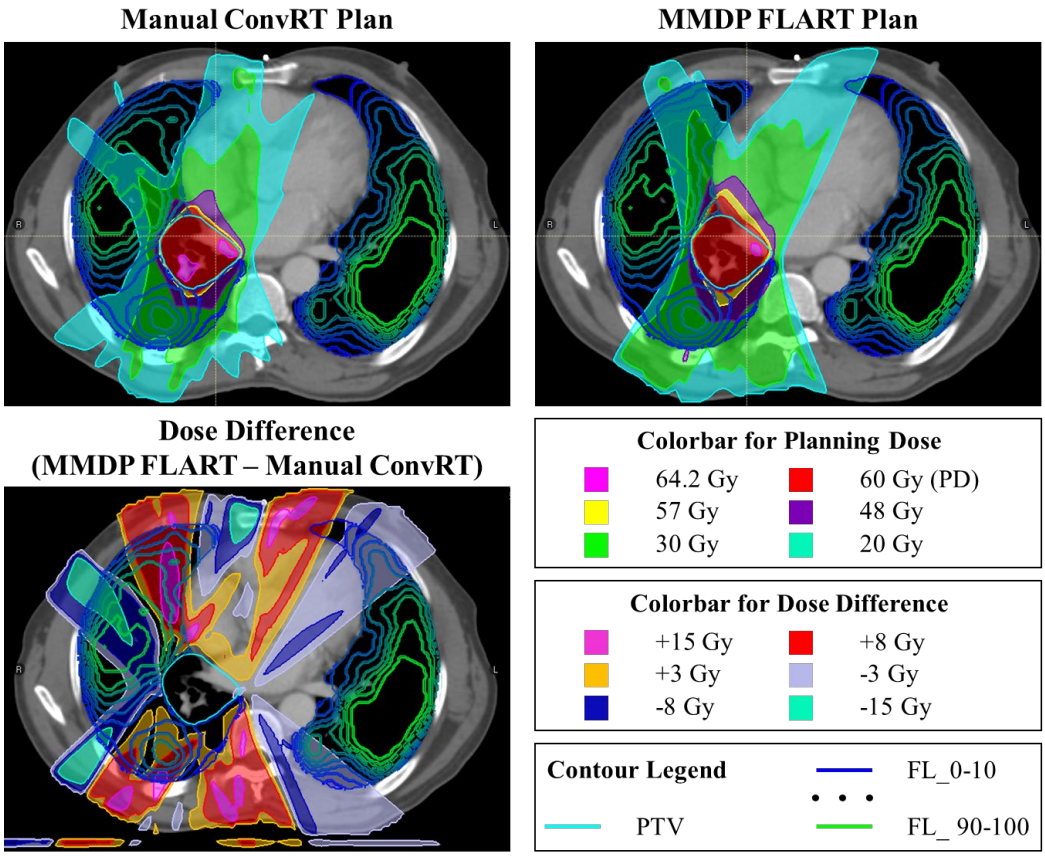


Figure 24. Comparison of dose distributions between a Manual ConvRT plan and an MMDP FLART plan for an example testing case.

5.3.3 Comparison of estimated radiation-induced toxicity probability

Table 9 presents a comparison of the estimated probabilities of radiation-induced toxicities between manual ConvRT plans and MMDP FLART plans. Compared to the current clinical standard manual ConvRT plans, MMDP FLART plans reduced the probability of radiation pneumonitis (grade ≥ 2) by 3.78 pp (23%), while maintaining comparable probabilities for other side effects (with only minor increases within 0.30 pp).

For 17 FLART-benefiting patients from the testing cohort, the probabilities of radiation-induced toxicities were estimated and compared between manual ConvRT and MMDP FLART plans, as summarized in Table 10. The MMDP FLART plans reduced the probability of radiation pneumonitis (Grade ≥ 2) by 6.25 pp (27%) compared to manual ConvRT plans, while maintaining comparable probabilities for other side effects (with only minor increases, all within 0.25 pp). These findings demonstrate the potential clinical benefit and safety of the proposed auto-planning algorithm for FLART.

Table 9. Comparison of normal tissue complication probability (NTCP) between manual ConvRT and MMDP FLART plans. The mean NTCP over all testing cases are shown.

OARs	Complication	Manual ConvRT	MMDP FLART	Difference

Lung	Radiation Pneumonitis (Grade ≥ 2)	16.46%	12.68%	**3.78 pp (23%)**
Esophagus	Acute Esophagitis (Grade ≥ 2)	5.73%	6.03%	-0.30 pp
Spinal Cord	Myelopathy	0.13%	0.24%	-0.11 pp
Heart	Cardiac Mortality	0.02%	0.05%	-0.03 pp

Table 10. Comparison of normal tissue complication probability (NTCP) between manual ConvRT and MMDP FLART plans. The mean NTCP over all FLART-benefiting patients from the testing cohort are shown.

OARs	Complication	Manual ConvRT	MMDP FLART	Difference
Lung	Radiation Pneumonitis (Grade ≥ 2)	22.99%	16.74%	**6.25 pp (27%)**
Esophagus	Acute Esophagitis (Grade ≥ 2)	6.51%	6.76%	-0.25 pp
Spinal Cord	Myelopathy	0.14%	0.28%	-0.14 pp
Heart	Cardiac Mortality	0.01%	0.07%	-0.06 pp

Note: Patients were classified as FLART-benefiting if the difference in the probability of radiation pneumonitis (Grade ≥ 2) between manual ConvRT and manual FLART plans exceeded 3 percentage points (pp).

5.3.4 Clinical performance comparison: MMDP-FLART V.S. MO-FLART

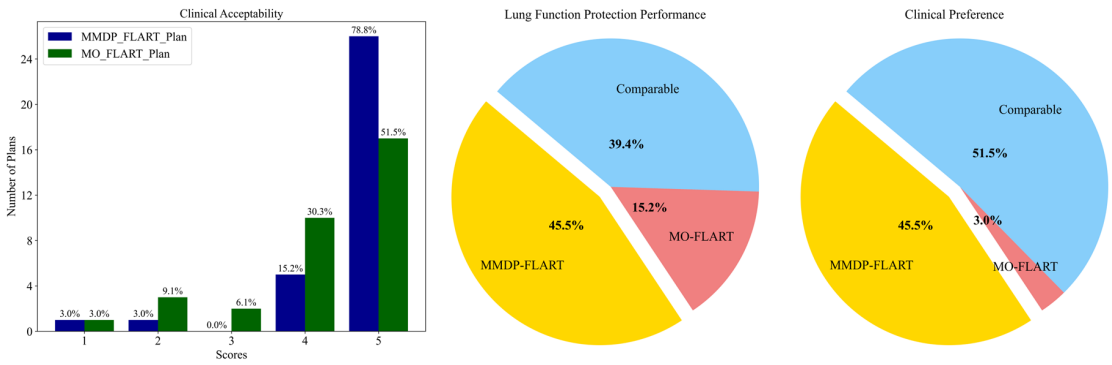


Figure 25. Qualitative blind review results for comparing MMDP FLART plans against MO FLART plans. The clinical acceptability scores for MMDP FLART plans and MO FLART plans are shown in the left bar plot. Higher scores indicate better

clinical acceptability. A score of 4 or 5 indicates the plan is regarded as clinically acceptable without any modifications. The lung function protection performance comparison results and the overall quality comparison results are shown in the middle pie chart and the right pie chart respectively.

Figure 25 summarizes the qualitative review results for comparing MMDP FLART plans against MO FLART plans, in which a senior physicist compared anonymized and shuffled MMDP FLART plans and MO FLART plans. As shown in Figure 25, 31 out of 33 (93.9%) MMDP FLART plans were deemed clinically acceptable without any modifications, outperforming MO FLART plans, of which 27 (81.8%) were considered clinically acceptable without modifications. Furthermore, Figure 25 demonstrates that MMDP FLART plans provided superior lung function protection compared to MO FLART plans, with 45.5% of MMDP FLART plans rated as better and 39.4% as comparable. In terms of overall plan quality, MMDP FLART plans were preferred, with 45.5% rated as better and 51.5% as comparable to MO FLART plans.

5.3.4 Clinical performance comparison: AP-FLART V.S. Senior Planner

Figure 26 presents the clinical acceptability scores for MMDP FLART plans as reviewed by three senior clinicians. On average, 87.9% of MMDP FLART plans were considered clinically acceptable (score $\geq$ 4) without any modifications. The comparison results between MMDP FLART plans against manual FLART plans in

terms of lung function protection performance and overall plan quality, as reviewed by three senior clinicians, are shown in Figure 27 and Figure 28 respectively. For lung function protection, 72.7% of MMDP FLART plans were rated as comparable (44.4%) or superior (28.3%) to manual FLART plans. Regarding overall plan quality, 30.3% of MMDP FLART plans were rated as superior to manual FLART plans, while 31.3% of manual FLART plans were considered of better quality. The results suggest that MMDP FLART plans exhibit clinical preference comparable to manual FLART plans meticulously created by a senior planner.

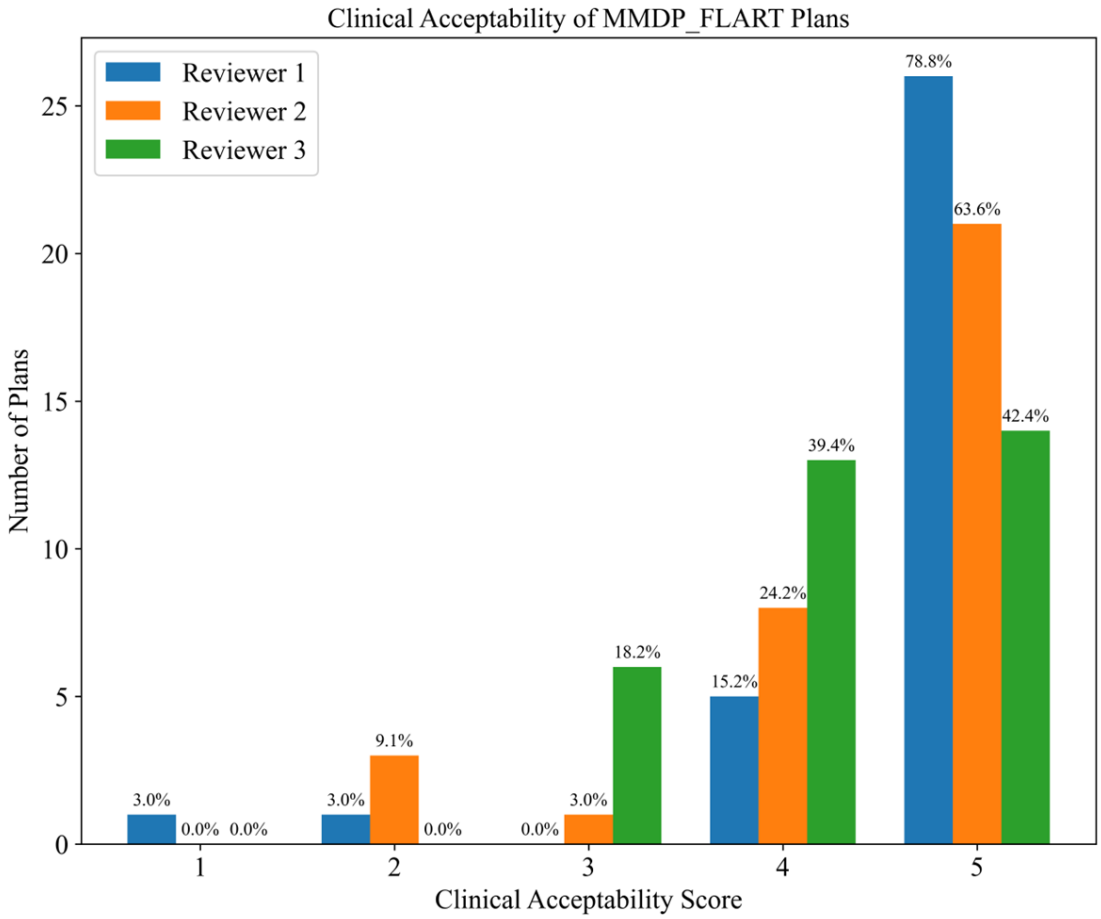


Figure 26. Clinical acceptability review results for MMDP FLART plans by one medical physicist (reviewer 1) and two radiation oncologists (reviewer 2 and 3).

Higher scores indicate better clinical acceptability. A score of 4 or 5 indicates the plan is regarded as clinically acceptable without any modifications.

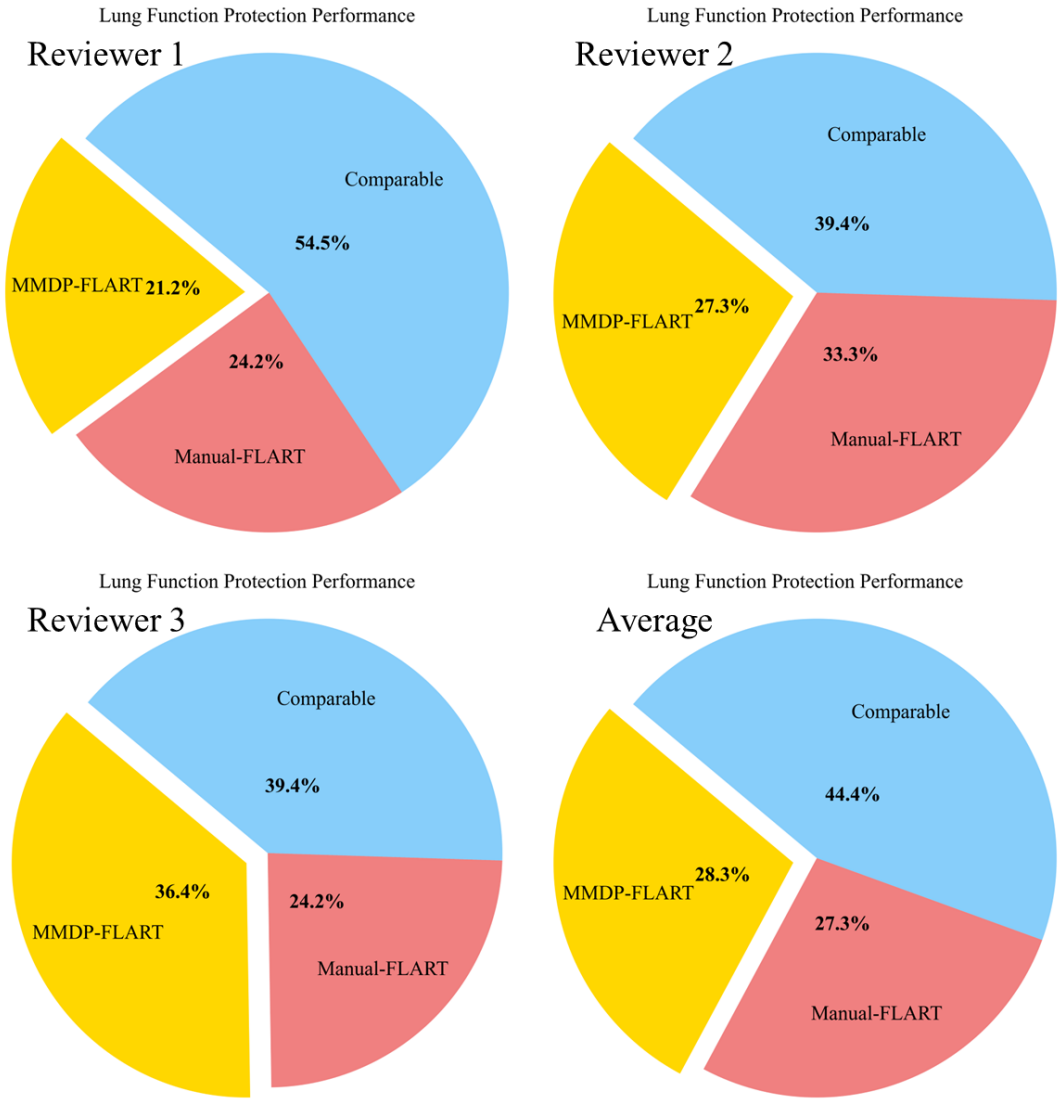


Figure 27. Comparison results between MMDP FLART plans and manual FLART plans in terms of lung function protection performance ranked by one medical physicist (reviewer 1) and two radiation oncologists (reviewer 2 and 3). The average results are also shown in the lower right pie chart.

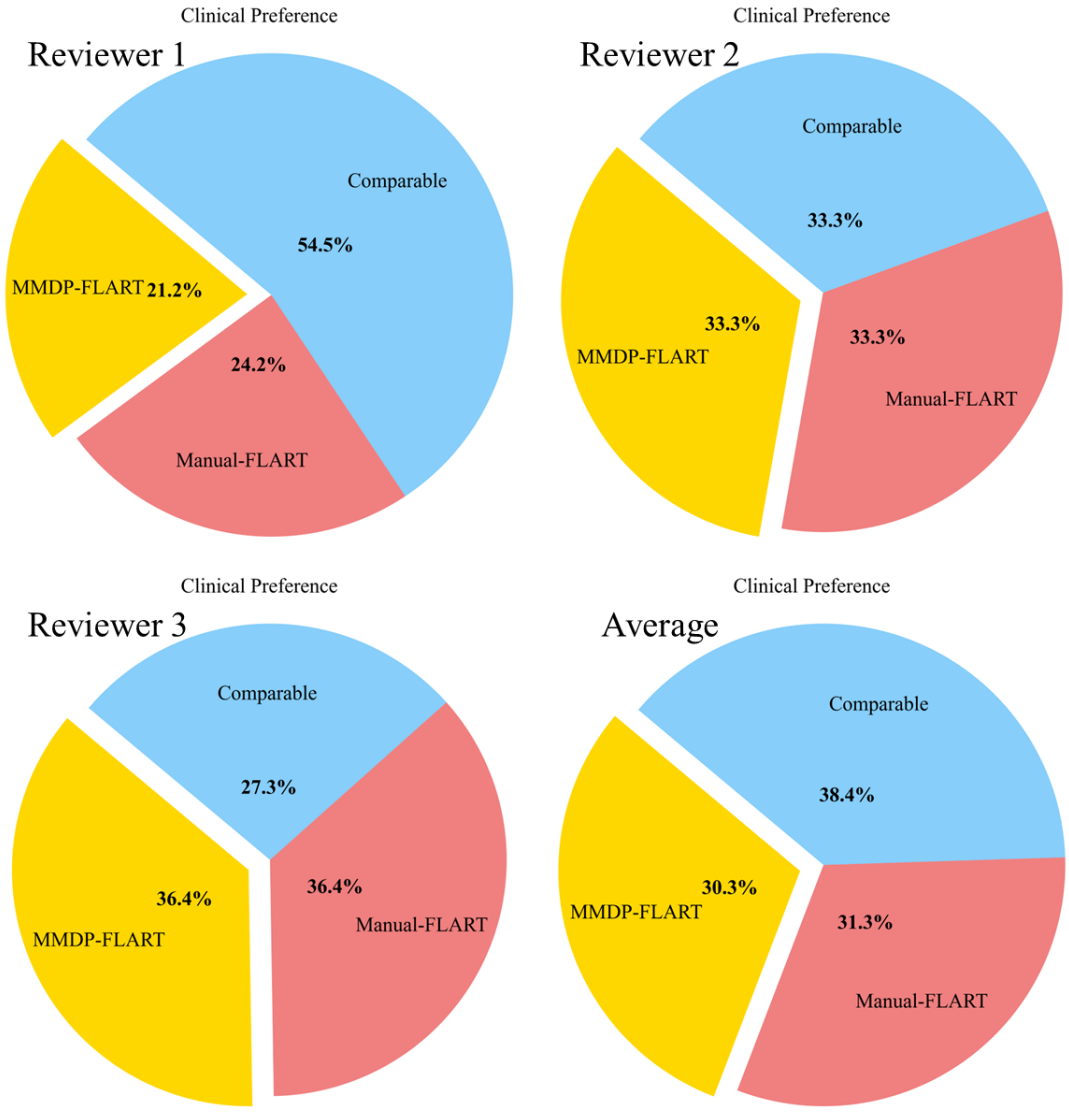


Figure 28. Comparison results between MMDP FLART plans and manual FLART plans in terms of overall plan quality ranked by one medical physicist (reviewer 1) and two radiation oncologists (reviewer 2 and 3). The average results are also shown in the lower right pie chart.

Table 11 compares the planning efficiency of manual FLART planning with the two auto-planning algorithms for FLART. The comparison was conducted on a RayStation workstation equipped with an Intel 8-core i9-10885H CPU (2.4 GHz / up to 5.3 GHz) and an NVIDIA Quadro RTX 5000 GPU. The MMDP-based auto-planning algorithm demonstrated the highest planning efficiency, achieving a reduction in planning time of approximately tenfold compared to MO-based auto-planning and over fifteenfold compared to manual FLART planning.

Table 11. Comparison of planning efficiency between manual FLART planning against MO-based and MMDP-based automatic FLART planning algorithms.

	Manual FLART planning	MO-based auto-planning	MMDP-based auto-planning
Dose grid		2.5×2.5×2.5 mm^3	
Planning time	~ 120 – 180 min	Beam angle selection 2 min	
		MO-based plan optimization 76 min	MMDP-based plan optimization 6 min
Total time	~ 120 – 180 min	78 min	8 min

Note: Planning time is defined as the interval from the availability of planning CT images, registered lung function images, and RT contours (including PTV and OARs) to the generation of a FLART plan. The auto-planning time refers to the average time required per plan across all test cases while manual planning time is estimated.

5.3.5 Clinical performance comparison: AP-FLART-Assisted Planning V.S. Independent Planning

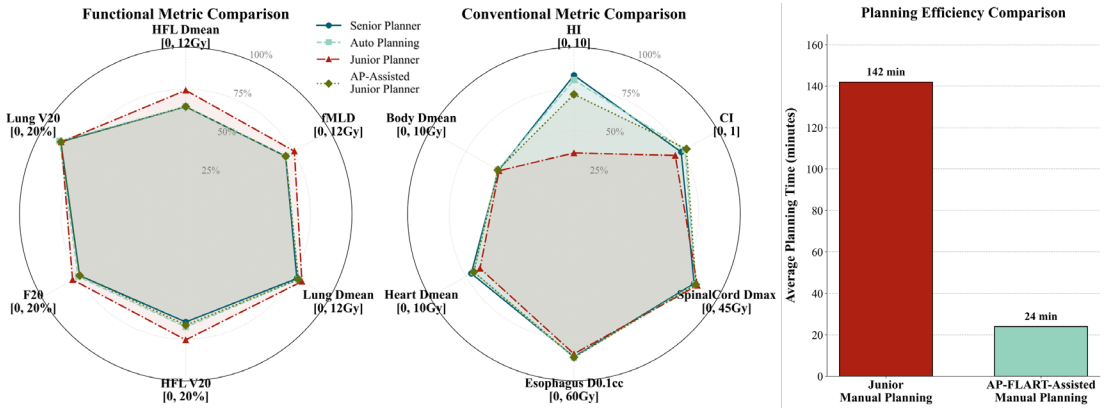


Figure 29. Comparison of functional and conventional plan quality metrics across FLART plans generated by the senior planner, the MMDP-based auto-planning system, the junior planner independently, and the junior planner utilizing AP-FLART assistance. Planning efficiency was also evaluated by comparing the junior planner's independent and AP-FLART-assisted workflows.

As shown in Figure 29, FLART plans generated independently by the junior planner exhibited inferior functional dose metrics, including HFL D_{mean}, HFL V_{20}, fMLD, and F_{20}, when compared to those produced by both the senior planner and the MMDP-based AP-FLART system. However, with the integration of the AP-FLART system, the junior planner was able to achieve functional dose metrics comparable to those of the senior planner. Furthermore, the average planning time was substantially reduced from 142 minutes for independent planning to 24 minutes for AP-FLART-assisted planning (comprising 9 minutes for auto-planning and 15 minutes for manual review and refinement). These findings demonstrate the clinical utility of the AP-FLART

system in significantly improving both the quality and efficiency of FLART planning for junior planners.

5.4 Discussion

FLART has demonstrated considerable promise in reducing radiation-induced lung toxicity. However, the substantial inter-patient variability and intra-lung heterogeneity of lung function distributions, coupled with limited clinical experience as an emerging radiotherapy technique, render manual FLART planning highly challenging. As a result, manual FLART planning is time-consuming and may yield inconsistent or suboptimal treatment plans. Therefore, the development of automatic planning algorithms for FLART is essential to enhance planning efficiency, consistency, and quality. In previous studies, we developed two auto-planning algorithms for FLART: one based on meta-optimization (MO) and another based on multi-modality-guided dose prediction (MMDP), and preliminarily validated their performance using an open-source TPS. In this section, we have implemented the proposed auto-planning algorithms on a clinical TPS and comprehensively evaluated their clinical performance and potential benefits, thereby advancing their clinical application.

Both the MMDP-based and MO-based auto-planning algorithms have been successfully integrated into the clinical TPS RayStation. As demonstrated in Figure 21 and the attached demo, a clinically ready auto-planning system for FLART has been developed, capable of automatic beam angle selection and plan optimization. This

system enables the generation of FLART plans with a single click, given planning CT images, RT contours, and registered lung function images. The plans are generated using the clinically validated plan optimization and dose calculation engines of RayStation, ensuring both accuracy and safety. The integration with RayStation allows clinicians to review, revise, and approve the generated plans, and finally delivery for patient treatment. The clinical implementation enables the clinical evaluation and application of the proposed auto-planning algorithms.

The dosimetric and potential clinical benefits of the automatic FLART planning algorithms were assessed by comparing automatic FLART plans with the current clinical standard manual ConvRT plans. Compared to manual ConvRT plans, MMDP FLART plans effectively reduced the fMLD and HFL D_{mean} by 11.8% and 15.1%, respectively. Consequently, the probability of grade ≥ 2 radiation pneumonitis was reduced by 3.78 pp (23%), while the probability of other side effects remained comparable (minor increases within 0.30 pp). The estimated clinical gain of FLART planning in this study is relatively modest compared to previous studies [11, 146], likely due to smaller tumor volumes and the absence of patient selection. For example, in the study by *Faught et al.* [146] where patients with heterogeneous lung function distributions were selected, manual FLART plans reduced HFL V_{20} from 30.1% (manual ConvRT) to 26.8%, and the probability of grade ≥ 2 radiation pneumonitis from 38.4% to 32.2% (a reduction of 6.2 pp, 16.1%). In contrast, our study observed a reduction in HFL V_{20} from 15.1% (manual ConvRT) to 11.9% (MMDP FLART), and

a reduction in the probability of grade ≥ 2 radiation pneumonitis from 16.46% to 12.68% (a reduction of 3.78 pp, 23%). The relatively low HFL V_{20} in our cohort reflects the smaller tumor volumes. For selected FLART-benefiting patients, the MMDP FLART plans significantly reduced the probability of grade ≥ 2 radiation pneumonitis by 6.25 pp (27%), while maintaining comparable probabilities for other side effects (with only minor increases, all within 0.25 pp), demonstrating both the clinical benefits and safety of the MMDP-based auto-planning algorithm.

The clinical performance of the proposed auto-planning algorithms was also comprehensively evaluated and compared through quantitative comparisons and qualitative reviews. Compared to the MO-based auto-planning algorithm, the MMDP-based algorithm demonstrated superior lung function protection, higher planning efficiency, and better clinical acceptability. Expert review indicated that the majority (97%) of MMDP FLART plans were considered better or comparable to MO FLART plans. Blind review by three senior clinicians also revealed that 88% of MMDP FLART plans were clinically acceptable without modification, and 69% were rated as comparable or superior to manual FLART plans meticulously created by a senior planner, underscoring the clinical acceptability of the MMDP-based auto-planning algorithm. Additionally, the MMDP-based algorithm reduced planning time by over fifteenfold compared to manual FLART planning, significantly enhancing planning efficiency.

The FLART plans generated independently by the junior planner were inferior to the automated results, further validating the superior clinical performance of the AP-FLART system. Most notably, with the assistance of the AP-FLART system, the quality of the plans produced by the junior planner was substantially enhanced, achieving a level comparable to that of the senior planner. The required planning time was also significantly reduced by 83%. These findings demonstrate the potential of the AP-FLART system to facilitate the broader clinical adoption of the emerging FLART technique.

We have demonstrated feasibility and validated the clinical performance of the MMDP-based auto-planning algorithm for functional lung avoidance IMRT. In addition to the lung, functional images have also been used to modulate dose limitations for other OARs and dose prescriptions for tumors [152, 153]. In addition to IMRT, other advanced RT techniques such as Volumetric Modulated Arc Therapy (VMAT) [69], Tomotherapy [102], and proton therapy [154, 155] are also widely used in function-guided RT [23], which may achieve better dose painting effectiveness. With appropriate training data and minor adjustments, the MMDP model could be adapted to achieve dose prediction and assist in treatment planning and plan quality assurance for various multi-modality-guided RT techniques.

A limitation of this study is that the auto-planning algorithms for FLART were validated only through in-silico planning studies. Future work will involve

prospective in-vivo patient trials to further assess clinical performance and validate clinical benefits of the MMDP-based automatic FLART planning algorithm. Accurate co-registration between lung function imaging and planning CT images is critical for precise functional mapping and directly impacts the overall efficacy of FLART. Consistent with established methodologies [20, 58, 130], this study utilized rigid registration to align the attenuation correction CT with the planning CT. This approach minimizes artifactual anatomical distortions, thereby facilitating the accurate spatial transfer of functional data from the SPECT V/Q to the planning CT. Importantly, the proposed automated planning system is agnostic to the specific registration method utilized and can readily integrate functional images processed via alternative algorithms. Future research will explore advanced registration techniques to further enhance the spatial accuracy of functional lung mapping. Additionally, SPECT V/Q image-guided FLART requires additional SPECT/CT scans to obtain patient-specific lung function information, resulting in increased radiation exposure and additional time and financial costs for patients. As a next step, we aim to combine our auto-planning algorithm with CT-derived V/Q imaging techniques [58, 61, 65, 156, 157] to facilitate the broader clinical adoption of FLART.

5.5 Conclusion

In this section, we integrated the MO-based and MMDP-based auto-planning algorithms for FLART in clinical TPS and evaluated their clinical performance and clinical benefits. A fully automatic planning system capable of automatic beam angle selection and plan optimization for FLART has been clinically deployed. The MMDP-based auto-planning algorithm outperforms the MO-based auto-planning algorithm in lung function protection performance, clinical acceptability, and planning efficiency. The MMDP-based auto-planning algorithm demonstrates potential clinical benefits of reducing the probability of radiation pneumonitis (grade $\geq$ 2) by 6.25 pp (27%) compared to the current clinical standard. The clinical acceptability of the MMDP-based auto-planning algorithm has been validated by the blind review results that 88% of MMDP FLART plans are clinically acceptable without any modifications and 69% of MMDP FLART plans are comparable or even superior to manual FLART plans meticulously created by a senior planner. The MMDP-based auto-planning algorithm can significantly improve planning efficiency, reducing planning time by more than fifteenfold compared to manual FLART planning. Furthermore, assistance from the AP-FLART system substantially enhanced the quality of the plans generated by a junior planner (with one year of planning experience), elevating the plan quality to a level comparable to that of a senior planner (with eight years of planning experience).

In summary, an auto-planning system for FLART has been technically developed, clinically deployed, and comprehensively evaluated. It shows considerable promise in enhancing the efficiency, consistency, and quality of FLART planning, while also

reducing radiation-induced lung toxicity compared to the current clinical standard. A demo for the clinically ready auto-planning system is available at the link.

6 Appendices

6.1 Appendices for MO-based auto-planning

6.1.1 Parameter determination for beam angle selection

The dosimetric score-based beam angle selection method involves several parameters to be determined. They include the weights of different regions to be protected in the beam score function and the weight and regularization parameter of the beam span term. For conventional RT (ConvRT), the weights for spinal cord, esophagus, heart, and normal tissues were set following *Yuan et al.* [107]. The weights for lung and the beam span term were fine-tuned by using a grid search method to minimize a plan score function. This function differs from that in the plan hyperparameter optimization by substituting the clipped linear relationship between the score values and the quality metric values with an exponential linear function defined by the Equation (6.1) to better align with clinical plan evaluation preferences.

$$F_{QM} = \begin{cases} \exp\left(\frac{QM - QM_+}{QM_+ - QM_-}\right), & QM \geq QM_+ \\ \frac{QM - QM_-}{QM_+ - QM_-}, & QM_- < QM < QM_+ \\ \exp\left(\frac{QM - QM_-}{QM_+ - QM_-}\right) - 1, & QM \leq QM_- \end{cases} \tag{6.1}$$

As for FLART, all parameters excluding the lung weight were aligned with those used in ConvRT to maintain the relative importance of other OARs. The functional lung weight was fine-tuned to minimize functional dose metrics while preserving conventional dose metrics compared to ConvRT, which can be formulated as:

$$\min_{w} \quad d(w) = \frac{1}{N \bullet M} \sum_{i} \sum_{QM} \frac{QM_i^{FLART} - QM_i^{ConvRT}}{QM_i^{ConvRT}} \tag{6.2}$$

Where *d* stands for the relative differences between FLART and ConvRT plan quality metrics. *QM* represents quality metrics involved in the definition of the plan score function. QM_i^{FLART} and QM_i^{ConvRT} refer to the quality metric value of *i*th case in FLART and ConvRT plans respectively. *N* and *M* are the number of cases and quality metrics. *w* refers to the functional lung weight. The voxel size of dose grid was set to be 3×3×3 mm^3 for final plan generation to accelerate the parameter search process.

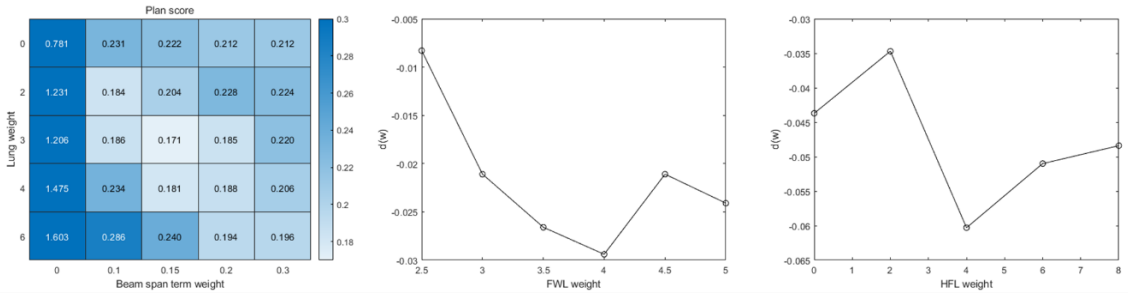


Figure 30. The grid search results for determining beam angle selection parameters.

Table 12. The determined parameters for beam angle selection

Term	Spinal Cord	Esophagus	Heart	Normal Tissue	Lung	FWL	HFL	Beam Span
ConvRT	10	0.3	0.3	0.15	3	0	0	0.15
vFLART	10	0.3	0.3	0.15	0	4	0	0.15
cFLART	10	0.3	0.3	0.15	2	0	4	0.15

Figure 30 presents the outcomes of the grid search optimization for beam angle selection parameters, with the optimal values detailed in Table 12. The chosen weight for the beam span term is lower than the value suggested by *Yuan et al*. [107]. It can be attributed to the use of solely coplanar beams in this study, in contrast to the inclusion of non-coplanar beams in *Yuan et al*., allowing for more varied beam distribution. The weight of FWL in vFLART is slightly higher than that of anatomical

lungs in ConvRT. The contralateral lung usually has higher function and lower doses in lung RT plans due to the relatively far distance from the tumor. Consequently, functionally weighted lung doses are usually lower than mean lung doses. A higher FWL weight can compensate for the lower dose values to maintain the relative importance of the anatomical/functional lung term in the beam score.

6.1.2 Nearest initialization for plan hyperparameter optimization

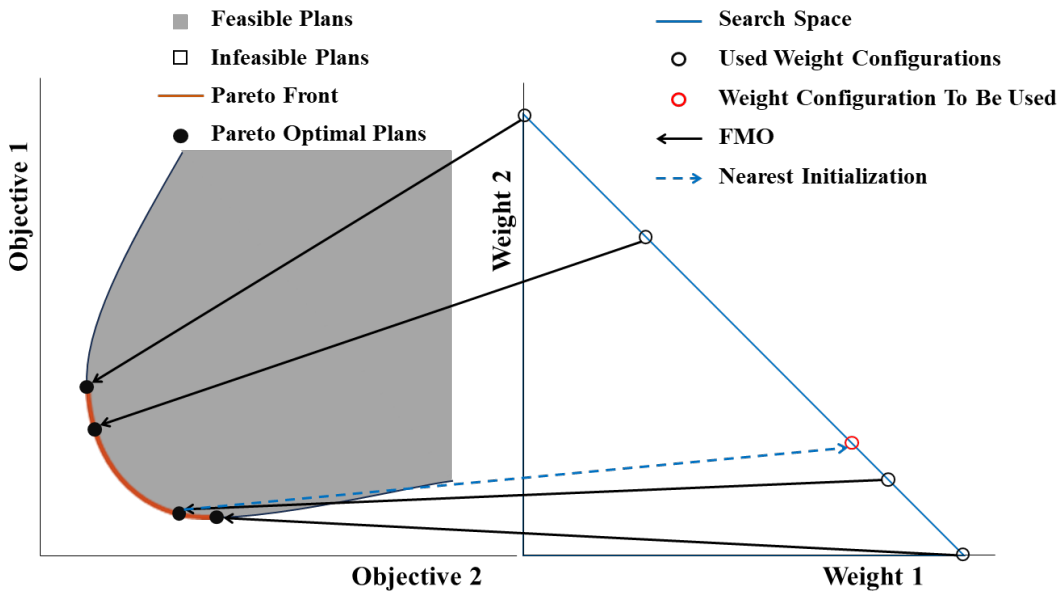


Figure 31. Illustration of the nearest initialization for the meta-optimization-based plan hyperparameter optimization.

The meta-optimization framework for plan hyperparameter optimization consists of two nested loops of optimization. For each iteration of the outer loop, dose objective weight configurations are updated and then fluence map optimization (FMO) is performed to generate corresponding Pareto optimal plans. In essence, FMO solves a

convex optimization problem using an iterative optimization algorithm, Interior Point OPTimizer (IPOPT). In the original meta-optimization framework, FMO initializes fluence maps with all one initialization. In this study, we propose a nearest initialization approach to accelerate the meta-optimization framework as illustrated by Figure 31. For a new dose objective weight configuration to be computed, the nearest computed weight configuration defined by Euclidean distance will be sought, and its FMO result will serve as the initialization point of the new configuration's FMO. Since the optimal fluence maps resulted from FMO depend continuously on the weight configurations, the nearest initialization approach shall reduce the iteration time for IPOPT to reach global optimal point compared to the original all one initialization method and thus accelerate the optimization process.

A comparison study was conducted to demonstrate the effectiveness of the nearest initialization approach. We generated two ConvRT plans for each collected case using the proposed AP-FLART with all one initialization and the nearest initialization respectively. The results showed that the mean iteration number per FMO was 121 versus 33. The computation time spent on FMO was 61.21 ± 18.15 min per plan when using all one initialization versus 22.55 ± 6.91 min per plan when using the nearest initialization, indicating a time reduction of 63%.

6.1.3 Dose mimicking for plan conversion

The dose mimicking algorithm aims to produce deliverable treatment plans that mimic or even improve given dose distribution. In this study, we utilize the voxel-based dose mimicking algorithm to generate matRad plans that can have similar or even better dose distribution compared to their Eclipse counterparts. The dose mimicking algorithm can be formulated as follows:

$$\min_{x} \quad \sum_{i \in PTV} \left(d_i^{cal} - d_i^{ref}\right)^2 + \sum_{i \in Body\ minus\ PTV} H\left(d_i^{cal} - d_i^{ref}\right)\left(d_i^{cal} - d_i^{ref}\right)^2 \quad (6.3)$$

$$s.t. \qquad x \geq 0$$

$$\overrightarrow{d^{cal}} = D\vec{x}$$

Where $\vec{x}$, D, and $\overrightarrow{d^{cal}}$ refer to the fluence map, the dose influence matrix, and the calculated dose distribution respectively. $\overrightarrow{d^{ref}}$ stands for the reference dose distribution, i.e., the planning dose distribution of manual Eclipse plans. H represents the Heaviside function. The IPOPT of matRad is applied to solve the above optimization problem to obtain optimal fluence map. Leaf sequencing and direct aperture optimization are then performed to generate final plans. The dose calculation and plan optimization algorithms used in the dose mimicking process are the same as those used in AP-FLART. The dose mimicking results are shown in Table 13.

Table 13. Comparison between original Eclipse plans and matRad plans converted through dose mimicking. The mean and standard deviation of the dose metrics over all collected cases are listed. The p-values by Wilcoxon Signed-Rank test are also shown.

	HI	CI	PTV D_1 (Gy)	HFL (Gy)	D_{mean}	HFL V_{20} (%)	HFL V_5 (%)	FWL (Gy)	fD_{mean}	FWL (%)	fV_{20}	FWL (%)	fV_5	Lung (Gy)	D_{mean}

Original plans	5.95±1.42	0.65±0.07	63.99±0.98	8.09±2.58	12.87±6.37	29.81±9.73	9.44±2.45	15.54±5.70	33.76±8.64	11.31±2.16
Converted plans	6.71±1.76	0.63±0.09	64.61±1.17	7.17±2.48	12.88±5.66	26.63±9.03	8.62±2.51	15.75±5.61	30.66±8.08	10.59±2.12
p-value	0.005	0.468	0.007	<0.001	1.000	0.003	<0.001	0.671	0.005	<0.001

	Lung V_{20} (%)	Lung V_5 (%)	SpinalCord D_{max} (Gy)	Esophagus D_{mean} (Gy)	Esophagus D_1 (Gy)	Heart D_{mean} (Gy)	Heart D_1 (Gy)	Heart V_{40} (%)	Body D_{mean} (Gy)
Original plans	19.45±4.46	38.00±7.00	39.30±6.12	16.94±8.37	54.09±15.61	9.82±7.47	45.39±17.72	6.31±6.85	5.66±1.78
Converted plans	19.73±4.44	35.05±5.76	36.63±3.95	13.32±6.39	50.93±17.26	7.37±5.95	39.27±17.33	1.92±2.19	4.91±1.48
p-value	0.671	0.010	0.003	<0.001	0.038	<0.001	<0.001	<0.001	<0.001

Compared to original Eclipse plans, mimicked matRad plans have significantly lower doses to all OARs including lung, spinal cord, esophagus, and heart. The lower OAR doses can be attributed to the compromise in PTV homogeneity and conformity, and the different dose calculation and plan optimization algorithms. For example, the pencil beam (PB) dose calculation algorithm tends to underestimate doses to OAR regions which are distant from beam paths for lung RT plans. This is because the PB algorithm lacks inhomogeneity correction for the lateral scattering and lung has a low density.

6.1.4 Detailed results of ablation studies

Table 14. The detailed results of the ablation studies. The mean and standard deviation of the dose metrics over all collected cases are listed. The p-values by Wilcoxon Signed-Rank test are also shown.

Dose metrics	ConvRT	cFLART	cFLARTBeam_ConvRT	p-value		ConvRTBeam_cFLART	p-value	
				VS ConvRT	VS cFLART		VS ConvRT	VS cFLART
HI	5.56±0.51	5.61±0.52	5.61±0.60	0.671	1.000	5.62±0.49	0.495	0.932
CI	0.76±0.06	0.72±0.06	0.74±0.08	0.081	0.099	0.74±0.07	0.369	0.054
PTV D_1 (Gy)	63.93±0.35	64.01±0.41	63.96±0.46	0.932	0.609	64.01±0.34	0.304	1.000
HFL D_{mean} (Gy)	7.21±2.28	6.08±2.44	6.84±2.32	0.099	<0.001	6.43±2.50	<0.001	0.007
HFL V_{20} (%)	13.01±5.55	11.00±5.75	12.80±5.92	0.304	<0.001	11.26±5.92	0.001	0.369
HFL V_5 (%)	32.64±10.50	25.98±9.32	28.11±7.70	0.018	0.002	29.66±11.59	0.010	0.008
FWL fMLD (Gy)	8.32±2.22	7.76±2.44	8.15±2.30	0.468	0.001	7.92±2.44	0.054	0.142
FWL F_{20} (%)	15.18±5.23	14.71±5.57	15.48±5.67	0.766	0.054	14.48±5.43	0.167	0.130
FWL F_5 (%)	35.92±9.27	30.69±9.16	32.47±7.56	0.024	0.003	33.51±10.81	0.060	0.024
Lung D_{mean} (Gy)	10.04±2.11	10.05±1.88	10.10±2.02	0.468	0.932	10.01±1.89	1.000	0.671
Lung V_{20} (%)	18.74±4.59	19.56±4.09	19.43±4.65	0.284	0.495	18.88±4.01	0.702	0.002
Lung V_5 (%)	39.79±8.61	36.55±7.74	37.83±7.66	0.074	0.014	38.30±8.31	0.181	0.060
SpinalCord D_{max} (Gy)	37.69±6.86	38.01±5.59	36.73±5.87	0.099	0.007	38.64±5.60	0.060	0.229
Esophagus D_{mean} (Gy)	13.29±8.38	12.05±7.80	12.80±8.18	0.099	0.142	12.27±7.90	0.054	0.417
Esophagus D_1 (Gy)	49.80±21.81	48.86±21.94	49.46±21.66	0.229	0.702	49.36±21.72	0.369	0.154
Heart D_{mean} (Gy)	6.37±6.21	6.69±6.38	6.34±6.24	0.181	0.495	6.17±5.85	0.229	0.016
Heart D_1 (Gy)	33.41±23.82	33.88±23.96	33.77±23.23	0.081	0.442	32.93±23.73	0.284	0.014
Heart V_{40} (%)	2.29±2.86	2.47±3.39	2.30±2.89	0.969	0.784	2.12±2.94	0.092	0.014
Body D_{mean} (Gy)	4.78±1.55	4.76±1.49	4.75±1.50	0.142	0.766	4.78±1.54	0.865	0.369
Dose metrics	ConvRT	vFLART	vFLARTBeam_ConvRT	p-value		ConvRTBeam_vFLART	p-value	
				VS ConvRT	VS vFLART		VS ConvRT	VS vFLART
HI	5.56±0.51	5.55±0.54	5.59±0.61	1.000	0.702	5.59±0.54	0.799	0.832
CI	0.76±0.06	0.71±0.08	0.73±0.08	0.027	0.081	0.75±0.07	0.229	0.005
PTV D_1 (Gy)	63.93±0.35	63.96±0.43	63.94±0.43	0.865	0.766	63.96±0.43	0.734	0.832
HFL D_{mean} (Gy)	7.21±2.28	6.18±2.39	6.66±2.38	0.009	<0.001	6.76±2.37	0.024	<0.001
HFL V_{20} (%)	13.01±5.55	11.14±5.60	12.43±5.73	0.081	0.001	11.55±5.95	0.027	0.167
HFL V_5 (%)	32.64±10.50	26.30±12.22	26.81±8.96	0.006	0.024	31.74±13.88	0.108	<0.001
FWL fMLD (Gy)	8.32±2.22	7.68±2.37	8.01±2.30	0.099	0.008	8.03±2.35	0.108	<0.001
FWL F_{20} (%)	15.18±5.23	14.28±5.55	15.15±5.40	0.702	0.030	14.22±5.66	0.119	0.671
FWL F_5 (%)	35.92±9.27	30.85±10.87	31.42±8.03	0.009	0.012	35.22±12.27	0.167	<0.001
Lung D_{mean} (Gy)	10.04±2.11	9.94±2.04	9.98±2.02	0.932	0.393	10.02±2.03	0.551	0.196
Lung V_{20} (%)	18.74±4.59	19.18±4.77	19.19±4.31	0.551	0.702	18.56±4.80	0.766	0.038
Lung V_5 (%)	39.79±8.61	36.26±7.75	36.77±7.25	0.081	0.021	39.36±8.58	0.167	0.006
SpinalCord D_{max} (Gy)	37.69±6.86	37.63±6.47	36.90±6.48	0.054	0.067	37.87±6.60	0.417	0.766
Esophagus D_{mean} (Gy)	13.29±8.38	12.91±8.07	12.89±8.18	0.246	0.671	13.19±8.13	0.284	0.167

Esophagus D_1 (Gy)	49.80±21.81	49.84±20.23	49.69±21.44	0.284	0.551	49.82±20.54	0.154	0.640
Heart D_{mean} (Gy)	6.37±6.21	7.03±6.33	6.49±6.43	0.142	0.734	6.52±5.86	0.196	0.027
Heart D_1 (Gy)	33.41±23.82	34.67±23.71	34.22±23.61	0.067	0.523	33.77±23.12	0.468	0.081
Heart V_{40} (%)	2.29±2.86	2.49±3.02	2.42±3.01	0.556	0.724	1.85±2.30	0.108	0.005
Body D_{mean} (Gy)	4.78±1.55	4.73±1.50	4.72±1.49	0.034	0.671	4.78±1.55	0.899	0.043

6.2 Appendices for MMDP-based auto-planning

6.2.1 Details of MMDP model architecture and training

The architectures of the MMDP model's encoders and decoder are similar to those of 3D U-net. The encoders and decoder consist of convolution layers (kernel size = 3, stride = 1, padding = 1), batch normalization layers, and leaky rectified linear unit layers (slope = 0.2), max pooling layers, and up sampling layers. The inputs to the anatomical encoder have four channels: relative electron density maps, binary PTV masks, OAR masks, and non-modulated dose maps. The relative electron density maps were derived from CT images based on the default conversion curve between HU values and relative electron densities of matRad since the reference MO-based plans were generated using this conversion curve. In the OAR masks, different regions of interest including lung, spinal cord, esophagus, heart, and normal tissue are assigned by different labels. The non-modulated dose maps were calculated using the pencil beam dose engine of matRad, providing beam setting information for dose prediction [35]. A residual connection was added to link the input non-modulated dose with the output predicted dose, aiming at facilitating model convergence [158]. The inputs to the functional encoder have two channels: lung function images and iso-function contour maps. In the iso-function contour maps, the lung function images are discretized, and a value x is assigned to lung voxels with functional values ranging from x-5 percentile to x+5 percentile (with $x \in [5,15,25,\ldots,95]$). All the input images have the same dimension of $160 \times 128 \times 96$ and the same resolution of $3 \times 3 \times 3$ mm^3.

In the first training stage, the loss function consists of mean absolute error (MAE) loss and DVH loss. The MAE loss is formulated by Equation (6.4). D_i^{pred} and D_i^{plan}, represent predicted and planning dose values to body voxel *i* respectively and the total number of voxels within body is n_{body}. The MAE loss was applied to provide voxel-wise guidance to induce the predicted dose distributions close to the ground truth. A DVH can be approximated by a series of moments with different orders, which are defined by Equation (6.5). n_s stands for the total number of voxels in structure *s* and *p* is the order of the moment. The DVH loss is defined as the mean square errors between the predicted and planning moments as formulated by Equation (6.6). We utilized *p* = (5,10) for spinal cord, *p* = (1,2) for other OARs including esophagus, heart, and lung, and *p* = (2,4,6) for PTV following the work by *Jhanwar et al.* [135]. The DVH loss was adopted to improve the prediction accuracy of clinically important DVH metrics. The total loss function of the first training stage was the weighted sum of the MAE loss and DVH loss. w_{DVH} was the relative weight of the DVH loss, set to 0.01.

$$L_{MAE} = \frac{1}{n_{body}} \sum_{i \in n_{body}} |D_i^{pred} - D_i^{plan}| \tag{6.4}$$

$$M_p^s = \left(\frac{1}{n_s} \sum_{i \in n_s} {D_i}^p \right)^{\frac{1}{p}} \tag{6.5}$$

$$L_{DVH} = \sum_{p \in P, s \in S} ||M_p^{spred} - M_p^{splan}||_2^2 \tag{6.6}$$

$$L_{ConvRT} = L_{MAE} + w_{DVH} L_{DVH} \tag{6.7}$$

In the second stage, in addition to the MAE loss and DVH loss, a functionally

weighted mean absolute error (fMAE) loss and a dose function histogram (DFH) loss were included as formulated by Equation (6.8-6.11). f_i represents the functional weight to lung voxel *i* and the total number of voxels within lung is n_{lung}. We utilized p = (1,2) for DFH loss.

$$L_{fMAE} = \frac{1}{n_{lung}} \sum_{i \in n_{lung}} f_i |D_i^{pred} - D_i^{plan}| \tag{6.8}$$

$$F_p^L = \left(\frac{1}{n_{lung}} \sum_{i \in n_{lung}} f_i {D_i}^p \right)^{\frac{1}{p}} \tag{6.9}$$

$$L_{DFH} = \sum_{p \in P} ||F_p^{L^{pred}} - F_p^{L^{plan}}||_2^2 \tag{6.10}$$

$$L_{FLART} = L_{MAE} + L_{fMAE} + w_{DVH}(L_{DVH} + L_{DFH}) \tag{6.11}$$

6.2.2 Manual planning

All manual plans were created using Eclipse TPS (version 16.1, Varian Medical Systems, Palo Alto, CA) with the analytical anisotropic algorithm dose engine (version 15.1.51) and the dynamic IMRT technique by an experienced physicist. For each patient, the physicist first created a conventional lung IMRT plan following institutional guidelines adapted based on the principles established by the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) [147] and the Radiation Therapy Oncology Group (RTOG) 0617 [6], as detailed in Table 15. A prescription dose of 60 Gy was applied for all cases. A FLART plan was then created by additionally considering lung function information with the conventional RT plan as a baseline. Given that most existing clinical trials for FLART have adopted the contour-based FLART strategy and the current clinical TPSs do not support voxel-wise FLART optimization, MA-FLART plans were made using the contour-based FLART planning strategy. A HFL contour was delineated from the lung function map through threshold segmentation with a threshold of 66^{th} percentile [111]. The FLART plan was generated with the aim of reducing HFL mean dose and V_{20} while avoiding severely damaging conventional dose objectives. To achieve this goal, the physicist meticulously adjusted beam angles, added functional dose objectives, and fine-tuned weight/dose/volume values of conventional and functional objectives. The final manual plans were reviewed and approved by another experienced physicist.

Table 15. Dose-volume constraints to organs at risk (OARs).

OARs	Value
Lung	$V_5 < 60\%$
	$V_{20} < 30\%$
	$D_{mean} < 20$ Gy
Spinal cord	$D_{max} < 45$ Gy
Esophagus	$D_{mean} < 34$Gy
	$D_{max} < 70$ Gy
Heart	$D_{mean} < 28$Gy
	$V_{40} < 33\%$

Abbreviations: V_x = percent volume receiving ≥ x Gy; D_{mean} = mean dose; D_{max} = maximum dose

To reduce unwanted discrepancies caused by using different TPSs, the Eclipse-based manual plans were converted into matRad plans through dose mimicking. The dose mimicking settings for plan conversion were the same as those for automatic plan generation, with the exception that a homogeneous lung function distribution was applied when converting MA-ConvRT plans. The dose metrics of original Eclipse plans and converted matRad plans are presented in Table 16.

Table 16. Comparison between original Eclipse manual plans against matRad plans converted through dose mimicking. The mean and standard deviation of various DFH and DVH metrics are listed.

	MA-ConvRT		MA-FLART	
DVH/DFH metrics	Eclipse	matRad	Eclipse	matRad
fMLD (Gy)	8.73 ± 3.26	7.34 ± 2.98	8.14 ± 3.23	6.74 ± 2.88
F_{20} (%)	14.90 ± 6.94	13.13 ± 6.10	13.08 ± 5.89	11.65 ± 5.31
HI	6.47 ± 1.59	5.36 ± 1.28	7.71 ± 1.86	5.72 ± 1.60
CI	0.72 ± 0.08	0.74 ± 0.08	0.67 ± 0.09	0.70 ± 0.09
PTV D_1 (Gy)	64.3 9± 1.09	63.67 ± 0.87	65.29 ± 1.25	63.92 ± 1.07
Lungs D_{mean} (Gy)	9.72 ± 3.37	8.32 ± 3.02	9.53 ± 3.47	8.12 ± 3.04
Lungs V_{20} (%)	17.23 ± 6.61	15.27 ± 5.81	16.37 ± 6.30	14.65 ± 5.61
Lungs V_5 (%)	35.15 ± 12.17	29.90 ± 9.86	33.83 ± 12.63	28.41 ± 10.22
SpinalCord D_{max} (Gy)	33.01 ± 9.81	30.93 ± 10.09	33.15 ± 10.74	30.50 ± 11.31
Esophagus D_{mean} (Gy)	12.27 ± 9.47	10.65 ± 8.89	12.38 ± 9.74	10.60 ± 9.04
Esophagus D_1 (Gy)	42.86 ± 21.16	41.17 ± 21.42	42.88 ± 21.18	40.96 ± 21.39

Heart D_{mean} (Gy)	7.11 ± 6.58	5.69 ± 5.74	7.50 ± 7.13	5.95 ± 6.11
Heart V_{40} (%)	3.08 ± 3.83	2.08 ± 2.70	3.83 ± 4.59	2.44 ± 3.04

Abbreviations: fMLD = functionally weighted mean lung dose; F_x = portion of lung function irradiated by dose higher than x Gy; D_{mean} = mean dose; D_{max} = maximal dose; V_x = portion of structure volume irradiated by dose higher than x Gy; D_x = minimum dose to the "hottest" x% of structure volume.

6.2.3 Comparison between MMDP-FLART plans against MA-FALRT/ConvRT plans

Table 17. Comparison between MMDP-FLART plans against MA-ConvRT and MA-FLART plans on the SPECT V dataset. The mean and standard deviation of various DFH and DVH metrics are listed. The p-values by Wilcoxon Signed-Rank test are also shown. MMDP-FLART plans outperform manual plans in terms of metrics marked in green while underperform manual plans in terms of metrics marked in blue.

DVH/DFH metrics	MMDP-FLART	MA-ConvRT	p-value	MA-FLART	p-value
fMLD (Gy)	5.91 ± 3.04	6.71 ± 3.20	<0.001	6.15 ± 3.03	<0.001
F_{20} (%)	10.65 ± 5.46	12.05 ± 5.62	<0.001	10.49 ± 5.00	0.487
HI	7.44 ± 1.21	4.61 ± 0.97	<0.001	5.55 ± 1.50	<0.001
CI	0.78 ± 0.08	0.79 ± 0.04	0.892	0.77 ± 0.03	0.001
PTV D_1 (Gy)	65.13 ± 0.81	63.13 ± 0.66	<0.001	63.78 ±1.01	<0.001
Lungs D_{mean} (Gy)	6.17 ± 2.86	6.65 ± 2.85	<0.001	6.39 ± 2.83	<0.001
Lungs V_{20} (%)	11.13 ± 5.02	11.91 ± 4.79	<0.001	11.13 ± 4.64	0.509
Lungs V_5 (%)	21.50 ± 9.87	24.09 ± 9.27	<0.001	23.02 ± 9.85	<0.001
SpinalCord D_{max} (Gy)	24.85 ± 13.79	22.91 ± 9.40	0.016	22.88 ± 9.09	0.023
Esophagus D_{mean} (Gy)	6.19 ± 9.95	6.99 ± 9.84	<0.001	7.02 ± 9.84	<0.001
Esophagus D_1 (Gy)	25.67 ± 19.89	27.86 ± 19.45	0.004	28.07 ± 18.61	<0.001
Heart D_{mean} (Gy)	3.84 ± 4.20	4.01 ± 4.71	0.055	4.20 ± 5.24	0.016
Heart V_{40} (%)	0.63 ± 1.08	0.90 ± 1.36	<0.001	1.17 ± 1.84	<0.001

Abbreviations: fMLD = functionally weighted mean lung dose; F_x = portion of lung function irradiated by dose higher than x Gy; D_{mean} = mean dose; D_{max} = maximal dose; V_x = portion of structure volume irradiated by dose higher than x Gy; D_x = minimum dose to the "hottest" x% of structure volume.

Table 18. Comparison between MMDP-FLART plans against MA-ConvRT and MA-FLART plans on the SPECT Q dataset. The mean and standard deviation of various DFH and DVH metrics are listed. The p-values by Wilcoxon Signed-Rank test are also shown. MMDP-FLART plans outperform manual plans in terms of metrics marked in green while underperform manual plans in terms of metrics marked in blue.

DVH/DFH metrics	MMDP-FLART	MA-ConvRT	p-value	MA-FLART	p-value
fMLD (Gy)	7.27 ± 3.20	7.73 ± 2.75	<0.001	7.12 ± 2.71	0.099
F_{20} (%)	13.90 ± 6.82	13.82 ± 6.28	0.889	12.39 ± 5.37	<0.001
HI	7.26 ± 1.34	5.83 ± 1.22	<0.001	5.82 ± 1.66	<0.001
CI	0.69 ± 0.12	0.71 ± 0.09	0.077	0.66 ± 0.10	<0.001
PTV D_1 (Gy)	65.11± 0.88	64.02 ± 0.81	<0.001	64.01 ± 1.09	<0.001
Lungs D_{mean} (Gy)	9.34 ± 2.96	9.38 ± 2.62	0.206	9.22 ± 2.64	0.164
Lungs V_{20} (%)	18.23 ± 6.17	17.41 ± 5.36	<0.001	16.89 ± 4.99	<0.001
Lungs V_5 (%)	33.12 ± 11.22	33.60 ± 8.32	0.239	31.84 ± 8.88	0.106
SpinalCord D_{max} (Gy)	35.23 ± 10.92	36.04 ± 6.58	0.193	35.35 ± 9.82	0.475
Esophagus D_{mean} (Gy)	11.47 ± 7.55	12.98 ± 7.34	<0.001	12.88 ± 7.67	<0.001

Esophagus D_1 (Gy)	47.38 ± 20.77	49.63 ± 18.05	<0.001	49.16 ± 18.85	0.003
Heart D_{mean} (Gy)	6.02 ± 5.22	6.76 ± 6.06	<0.001	7.06 ± 6.36	<0.001
Heart V_{40} (%)	2.37 ± 2.90	2.84 ± 3.04	<0.001	3.24 ± 3.36	<0.001

Abbreviations: fMLD = functionally weighted mean lung dose; F_x = portion of lung function irradiated by dose higher than x Gy; D_{mean} = mean dose; D_{max} = maximal dose; V_x = portion of structure volume irradiated by dose higher than x Gy; D_x = minimum dose to the "hottest" x% of structure volume.

6.3 Appendices for clinical implementation and evaluation

6.3.1 MMDP model architecture and training

MMDP model architecture

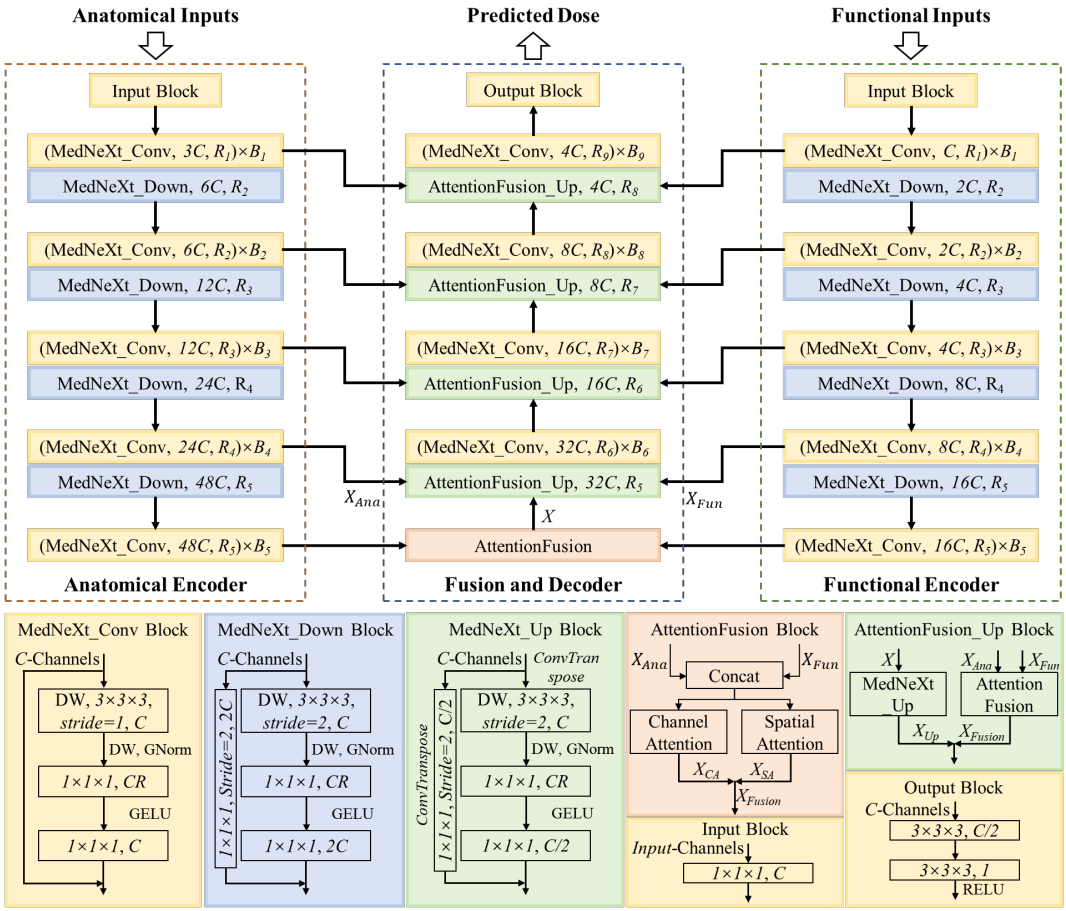


Figure 32. The architecture of the multi-modality-guided dose prediction (MMDP) model based on 3D MedNeXt [159]. C, R_i, and B_i represent the number of channels, the expansion ratio, the number of blocks respectively. The base configuration of 3D MedNeXt [159] is adopted with $C = 8$, $B_{all} = 2$, $R = [2,3,4,4,4,4,4,3,2]$.

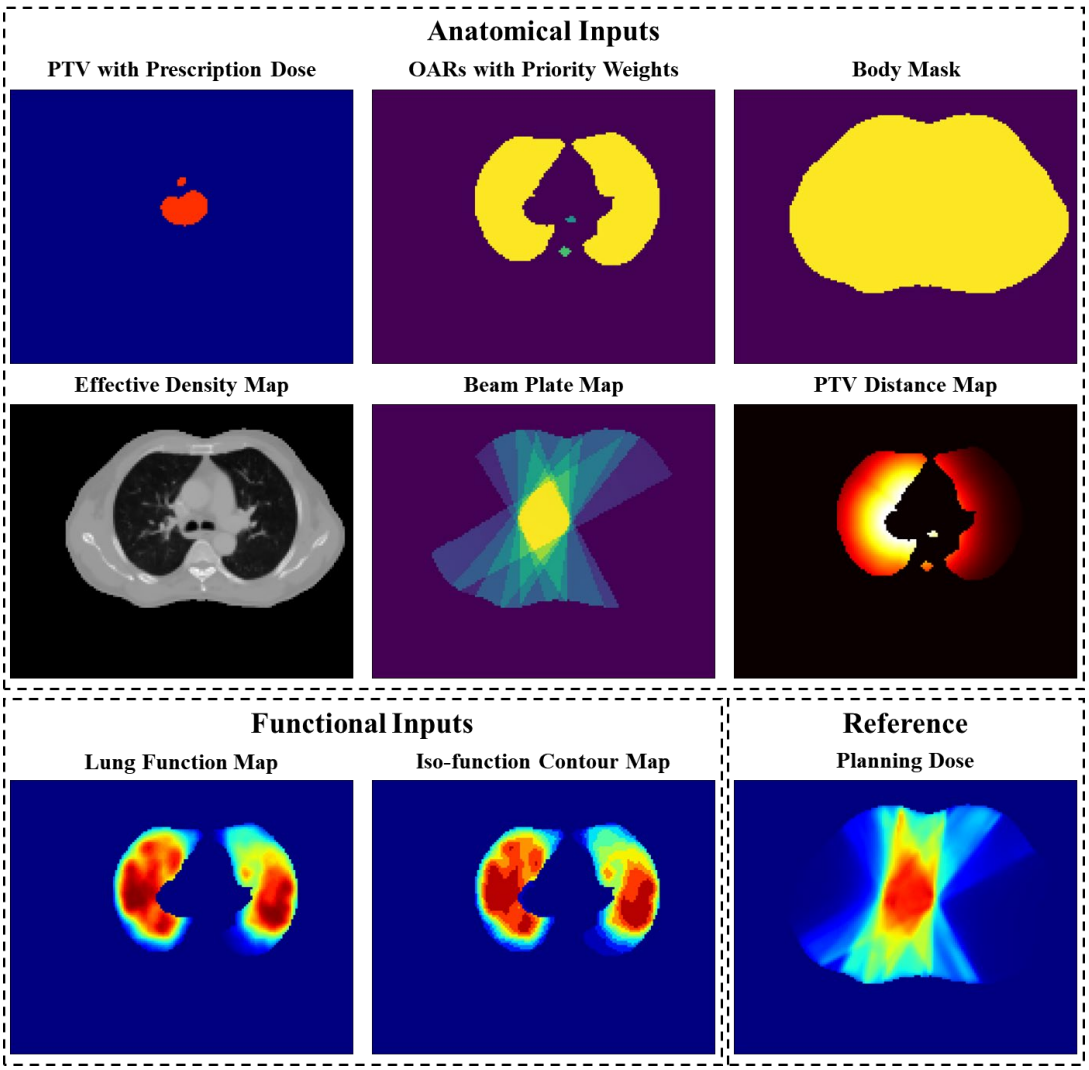


Figure 33. Demonstration of MMDP model inputs and references.

As illustrated in Figure 32, the MMDP model used in this section consists of two encoders, a decoder, and several attention-based fusion modules, which are similar to the MMDP model used in Section 4. The two encoders are designed to extract useful and complementary features from anatomical and functional inputs respectively. The extracted anatomical and functional features are fused and enhanced via the attention-based fusion modules consisting of channel attention blocks [133] and spatial attention blocks [134]. The decoder then utilizes the fused features to generate the optimal dose distributions. In contrast to the UNet-based MMDP model utilized in

Section 4, the current model is developed based on the 3D MedNeXt architecture [159], which has demonstrated superior performance compared to traditional UNet-based models in a variety of medical image segmentation and dose prediction tasks , making it a robust foundation for dose prediction in radiotherapy planning.

As depicted in Figure 33, the model inputs consist of anatomical inputs and functional inputs. The functional inputs contain percentile lung function maps and iso-function contour maps. The anatomical inputs contain six input channels, providing patient-specific anatomical information, beam setting information, physics-informed prior knowledge, and PTV distance map as follows:

1. **PTV with Prescription Dose:** In this first channel, voxels of PTV are assigned the prescription dose value, thereby encoding patient-specific prescription information and enabling dose prediction for plans with varying prescription dose settings.
2. **OARs with Priority Weights:** The second channel represents OARs, with each OAR voxel assigned a weight value ($W = 2^{1-P}$), where P denotes the clinical priority. The clinical priority (P) is set to be 1 for Lungs-PTV, 1.5 for SpinalCord, 2 for Esophagus, and 2 for Heart based on clinical preferences. It allows the input to encode clinical preferences regarding the clinical importance of different OARs.
3. **Binary Body Mask:** The third channel is a binary mask indicating the body volume, thereby specifying dose calculation and prediction space.

4. **Effective Density Map:** The fourth channel provides tissue heterogeneity information via effective density maps, which are converted from CT images using the same conversion tables between Hounsfield Unit (HU) values and mass densities or relative electron densities as those used for plan generation.
5. **Normalized Distance-Aware Beam Plate Map:** The fifth channel encodes patient-specific beam angle information, which is essential for predicting dose distributions for plans with different beam configurations [96]. The normalized distance-aware beam plate maps are defined as Eq. (6.12-6.13) where B_i represents beam plate value of voxel i. N is the total number of gantry angles for IMRT plans. D_{ij} refers to the distance of voxel i to source of beam j. SAD indicates the source to iso-center distance. Comparted to another common beam angle indicator – binary angle plate maps [96], where each beam angle is indicated by a binary plate with a fixed width, the normalized distance-aware beam plate maps have the advantages of masking complete beam path regions where most doses deposited, providing distance information from beam sources, and encoding the inverse-square law of beam intensity.

$$B_i = \frac{1}{N}\sum_{j}^{N} \delta_{ij} \frac{SAD^2}{{D_{ij}}^2} \tag{6.12}$$

$$\delta_{ij} = \begin{cases} 1, \ voxel\ i\ is\ in\ the\ path\ of\ beam\ j \\ 0, \ otherwise \end{cases} \tag{6.13}$$

6. **PTV Distance Map:** The fourth channel is a transformed distance map which encodes the minimum distance from each OAR voxel to the surface of the union of all PTVs [160]. The transformed distance map (T) is defined as Eq. (6.14) following the approach of *Yue et al.* [160]. D_i represents the minimum

distance from each OAR voxel to the surface of the union of all PTVs with a unit of mm. The hyperparameters *t*, *s*, and *c* are set as 1.3, 20, and 1.204 respectively as in *Yue et al.*[160]. The spatial relationship between OARs and the PTV is a critical feature influencing planning results and is commonly utilized in knowledge-based planning techniques[28]. Incorporating this distance information may enhance dose prediction accuracy[160].

$$T_i = \frac{t}{1 + \exp\left(\frac{D_i}{s} - c\right)} \tag{6.14}$$

All inputs have the same size of 160×128×96 with a consistent resolution of 2.5×2.5×2.5 mm^3 to cover most regions of interest.

MMDP model training

In contrast to the MMDP model described in Section 4, which was developed and evaluated using automatically generated ConvRT and FLART plans from a cohort of 196 patients without clinical review, the MMDP model presented in this section was trained with more and higher-quality plans to enhance its generalizability and clinical acceptability. Initially, the model was pretrained using a large dataset comprising 895 patients, including 1,463 automatic ConvRT plans and 508 automatic FLART plans, as detailed in Table 4. Subsequently, the model was fine-tuned and evaluated using a well-curated set of 83 pairs of manually generated ConvRT and FLART plans, which underwent clinical review. For the fine-tuning phase, 50 pairs of manual ConvRT and FLART plans were utilized, with 40 pairs allocated for model training and 10 pairs reserved for validation. The remaining 33 pairs were set aside for independent testing

of the model's performance. Model training employed both MAE and DVH losses, as elaborated in Appendix 6.2.

To further improve the model generalizability and robustness, on-the-fly data augmentation techniques—including random translation, rotation, and flipping—were applied during training. The Adam optimizer was used with a batch size of one. During pretraining, the model was trained for 400 epochs with an initial learning rate of 0.001, which gradually decreased according to a cosine annealing schedule. In the fine-tuning phase, the model was trained for 100 epochs with an initial learning rate of 0.0005, also following a cosine annealing schedule. The final MMDP model was selected based on its performance on the validation dataset. All model implementation and training were conducted using PyTorch on an NVIDIA RTX 4090 GPU.

6.3.2 Function-guided dose mimicking

The function-guided dose mimicking algorithm facilitates lung dose painting and enables the conversion of predicted dose distributions into deliverable FLART plans, as introduced in Section 4.2.5. Initially, the predicted dose maps were preprocessed to achieve clinically desirable dose distributions by increasing dose homogeneity within the PTV and reducing dose to OARs, as formulated in Equation (4.3) [136]. The dose preprocessing factors f_{PTV} and f_{OARs} were set to 0.1 and 0.05 respectively.

Following preprocessing, the refined dose maps were used to guide the setting of

optimization objectives for dose mimicking. Unlike the approach described in Section 4.2.5, which utilized voxel-wise lung function information to assist lung dose painting, the updated dose mimicking algorithm employed ten functional lung contours to approximate "voxel-wise" lung dose painting. This adjustment was necessary due to the clinical TPS RayStation's lack of support for voxel-wise objective weights. Additionally, several auxiliary conventional DVH objectives were incorporated to further enhance the clinical acceptability of the generated plans. The optimization functions utilized for dose mimicking are summarized in Table 19.

Table 19. Optimization objective settings for function-guided dose mimicking.

	ROI	ROI definition	Objective	Weight
Dose mimicking optimization objectives	PTV	PTV	$D^{cal} = D^{ref}$	800
	FL_i*10_(i+1)*10, $i \in [0,1,2,\ldots,9]$	Lung regions excluding PTV with functional weights ranging from i*10 percentile to (i+1)*10 percentile	$D^{cal} < D^{ref}$	(i*0.2+0.1)*600
	SpinalCord	SpinalCord	$D^{cal} < D^{ref}$	400
	Heart	Heart	$D^{cal} < D^{ref}$	200
	Esophagus	Esophagus	$D^{cal} < D^{ref}$	200
	NormalTissue	Body-PTV-OARs	$D^{cal} < D^{ref}$	100
Auxiliary optimization objectives	PTV	PTV	$D_{98} > 59Gy$	200
	PTV	PTV	$D_1 < 65.4Gy$	200
	PTV_internal_05	Inner wall of PTV with thickness of 5 mm	$D_{min} > 57Gy$	400
	Ring_02_07	PTV extended 7 mm minus PTV extended 2 mm	$D_{max} < 59Gy$	100
	Ring_07_12	PTV extended 12 mm minus PTV extended 7 mm	$D_{max} < 56Gy$	100
	Ring_12_17	PTV extended 17 mm minus PTV extended 12 mm	$D_{max} < 51Gy$	100
	Ring_17_22	PTV extended 22 mm minus PTV extended 17 mm	$D_{max} < 46Gy$	100
	Body-PTV22	Body minus PTV extended 22 mm	$D_{max} < 40Gy$	100
	SpinalCord	SpinalCord	$D_{max} < 43.5Gy$	Constraint

Note: Unlike the optimization function used in Eq. (4.4), the penalty value of an optimization function in RayStation is determined by the relative dose deviation. For example, the penalty value of an optimization function $D^{cal} < D^{ref}$ for ROI S with a weight of W_s is defined as $\frac{w_s}{n_s}\sum_{i\in S} H\left(D_i^{cal} - D_i^{ref}\right)\left(\frac{D_i^{cal}-D_i^{ref}}{D_i^{ref}}\right)^2$. Consequently, the weights listed in the table are multiplied by the square of the reference dose level when implemented in RayStation.

The Collapse Cone dose engine (version 5.9) was used for dose calculation. A TrueBeam (Varian Medical Systems, Palo Alto, CA, USA) machine was used for plan generation. Dynamic IMRT technique with beam energy of 6 MV was adopted, consistent with manual planning.

6.3.3 MO-based automatic plan optimization

When implementing the MO-based plan optimization algorithm in RayStation, the voxel-wise FLART planning optimization is approximated by using ten functional lung contours. Specifically, the plan optimization function defined by Eq. (3.7) is replaced by the following function:

$$\min_{x} \quad \frac{w_{PTV}}{N_{PTV}}\sum_{i\in PTV}\left(d_i - D_p\right)^2 + \sum_{S\in OARs}\frac{w_S}{N_s}\sum_{i\in S}H(d_i - D_s)(d_i - D_s)^2 + \sum_{\substack{fl=\text{FL\_}j*10\_(j+1)*10 \\ j\in[0,1,2,\ldots,9]}}\frac{w_{FL}(j*0.2+0.1)}{N_{fl}}\sum_{i\in fl}H(d_i - D_{FL})(d_i - D_{FL})^2 \quad (6.15)$$

$$s.t. \quad x \geq 0$$

$$w \geq 0$$

$$\vec{d} = D\vec{x}$$

Where $\vec{x}$, D, and $\vec{d}$ refer to the fluence map, the dose influence matrix, and the dose distribution respectively. d_i represents the dose value of voxel i. w and N are the weights of different objectives and the number of voxels in different regions respectively. D_p, D_s, and D_{FL} stand for reference dose of different objectives. H

represents the Heaviside function. The details of dose objectives are summarized in Table 1.

6.3.4 Manual planning

For each case in the third and fourth datasets listed in Table 4, a senior physicist manually generated a pair of ConvRT and FLART plans through a structured four-stage planning process: (1) manual ConvRT planning, (2) manual contour-based FLART planning, (3) dose mimicking-based voxel-wise FLART plan optimization, and (4) manual refinement.

The physicist first created a conventional lung IMRT plan following institutional guidelines adapted based on the principles established by the Quantitative Analysis of Normal Tissue Effects in the Clinic (QUANTEC) [147] and the Radiation Therapy Oncology Group (RTOG) 0617 [6], as detailed in Table 15. A prescription dose of 60 Gy was applied for all cases. Given that simultaneously considering multiple functional lung contours are challenging for human planners, a contour-based FLART plan was created by additionally considering a HFL contour with the conventional RT plan as a baseline. The HFL contour was delineated from the percentile lung function map through threshold segmentation with a threshold of 66^{th} percentile [111]. The contour-based FLART plan was generated with the aim of reducing HFL mean dose and V_{20} while avoiding severely damaging conventional dose objectives. To achieve this goal, the physicist meticulously adjusted beam angles, added functional dose

objectives, and fine-tuned weight/dose/volume values of conventional and functional objectives. In the third stage, function-guided dose mimicking was performed to convert the manually generated contour-based FLART plans into approximately voxel-wise FLART plans. The planning doses of the manual contour-based FLART plans were first preprocessed by reducing the doses to ten functional lung contours, creating optimization room for dose redirection from high-function lung regions to low-function lung regions and other less critical areas. Specifically, the doses to "FL_i*10_(i+1)*10" ($i \in [0,1,2,\ldots,9]$) were reduced by (i+1)*0.5%. The preprocessed doses were used to guide the setting of optimization objectives for dose mimicking, analogous to the approach used for MMDP-based auto-planning. Finally, the physicist conducted a thorough review and, if necessary, manual refinement of the "voxel-wise" FLART plan to ensure clinical acceptability. All manual plans underwent independent review and approval by a second experienced physicist. All manual plans were created using RayStation TPS with the Collapse Cone dose engine (version 5.9). A TrueBeam machine was used for plan generation. Dynamic IMRT technique with beam energy of 6 MV were used for all plans.

7 References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63.
2. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. The Lancet. 2021;398(10299):535-54.
3. Dela Cruz CS, Tanoue LT, Matthay RA. Lung cancer: epidemiology, etiology, and prevention. Clinics in chest medicine. 2011;32(4):605-44.
4. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. Non-Small Cell Lung Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2022;20(5):497-530.
5. Uzel EK, Figen M, Uzel Ö. Radiotherapy in Lung Cancer: Current and Future Role. Sisli Etfal Hastanesi tip bulteni. 2019;53(4):353-60.
6. Bradley JD, Paulus R, Komaki R, Masters G, Blumenschein G, Schild S, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. The Lancet Oncology. 2015;16(2):187-99.
7. De Ruysscher D, Faivre-Finn C, Nestle U, Hurkmans CW, Le Péchoux C, Price A, et al. European Organisation for Research and Treatment of Cancer recommendations for planning and delivery of high-dose, high-precision radiotherapy for lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(36):5301-10.
8. Morris S, Roques T, Ahmad S, Loo S. Practical radiotherapy planning: CRC Press; 2023.
9. Vinogradskiy Y, Schubert L, Diot Q, Waxweiller T, Koo P, Castillo R, et al. Regional Lung Function Profiles of Stage I and III Lung Cancer Patients: An Evaluation for Functional Avoidance Radiation Therapy. Int J Radiat Oncol Biol Phys. 2016;95(4):1273-80.
10. Marks LB, Munley MT, Spencer DP, Sherouse GW, Bentel GC, Hoppenworth J, et al. Quantification of radiation-induced regional lung injury with perfusion imaging. 1997;38(2):399-409.
11. Faught AM, Yamamoto T, Castillo R, Castillo E, Zhang J, Miften M, et al. Evaluating Which Dose-Function Metrics Are Most Critical for Functional-Guided Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2017;99(1):202-9.
12. Flakus MJ, Kent SP, Wallat EM, Wuschner AE, Tennant E, Yadav P, et al. Metrics of dose to highly ventilated lung are predictive of radiation-induced pneumonitis in lung cancer patients. Radiotherapy and Oncology. 2023;182.
13. Bucknell NW, Hardcastle N, Bressel M, Hofman MS, Kron T, Ball D, et al. Functional lung imaging in radiation therapy for lung cancer: A systematic review and meta-analysis. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2018;129(2):196-208.
14. Midroni J, Salunkhe R, Liu Z, Chow R, Boldt G, Palma D, et al. Incorporation of Functional Lung Imaging Into Radiation Therapy Planning in Patients With Lung Cancer: A Systematic Review and Meta-Analysis. International Journal of Radiation Oncology*Biology*Physics. 2024.
15. Chen Z, Li Z, Huang Y-H, Xiong T, Teng X, Li B, et al. A Review of Advances in Lung Function Imaging and Its Applications for Functional Lung Avoidance in Radiation Therapy. Seminars in Nuclear Medicine. 2025.

16. Doi Y, Kimura T, Nakashima T, Takeuchi Y, Takahashi I, Nishibuchi I, et al. Functional image guided radiation therapy planning in volumetric modulated arc therapy for patients with malignant pleural mesothelioma. Advances in radiation oncology. 2017;2(2):183-91.
17. Matrosic CK, Owen DR, Polan D, Sun Y, Jolly S, Schonewolf C, et al. Feasibility of function-guided lung treatment planning with parametric response mapping. Journal of Applied Clinical Medical Physics. 2021;22(11):80-9.
18. Baschnagel AM, Flakus MJ, Wallat EM, Wuschner AE, Chappell RJ, Bayliss RA, et al. A Phase 2 Randomized Clinical Trial Evaluating 4-Dimensional Computed Tomography Ventilation-Based Functional Lung Avoidance Radiation Therapy for Non-Small Cell Lung Cancer. International Journal of Radiation Oncology*Biology*Physics. 2024;119(5):1393-402.
19. Yamamoto T, Kabus S, Bal M, Keall PJ, Moran A, Wright C, et al. Four-Dimensional Computed Tomography Ventilation Image-Guided Lung Functional Avoidance Radiation Therapy: A Single-Arm Prospective Pilot Clinical Trial. International Journal of Radiation Oncology*Biology*Physics. 2023;115(5):1144-54.
20. Matuszak MM, Matrosic C, Jarema D, McShan DL, Stenmark MH, Owen D, et al. Priority-driven plan optimization in locally advanced lung patients based on perfusion SPECT imaging. Advances in radiation oncology. 2016;1(4):281-9.
21. Tahir BA, Bragg CM, Wild JM, Swinscoe JA, Lawless SE, Hart KA, et al. Impact of field number and beam angle on functional image-guided lung cancer radiotherapy planning. Physics in Medicine & Biology. 2017;62(17):7114.
22. Christian JA, Partridge M, Nioutsikou E, Cook G, McNair HA, Cronin B, et al. The incorporation of SPECT functional lung imaging into inverse radiotherapy planning for non-small cell lung cancer. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2005;77(3):271-7.
23. Bucknell NW, Hardcastle N, Woon B, Selbie L, Bressel M, Byrne K, et al. The HI-FIVE Trial: A Prospective Trial Using 4-Dimensional (68)Ga Ventilation-Perfusion Positron Emission Tomography-Computed Tomography for Functional Lung Avoidance in Locally Advanced Non-small Cell Lung Cancer. International Journal of Radiation Oncology*Biology*Physics. 2023;117(4):887-92.
24. Edwards DM, Schonewolf CA, Rice JD, Schipper M, Haken RKT, Matuszak M, et al. Phase II Trial Assessing Toxicity of Personalized Response-Based Radiation Treatment in Patients with Locally Advanced Non-Small Cell Lung Cancer. International Journal of Radiation Oncology*Biology*Physics. 2024.
25. Vinogradskiy Y, Castillo R, Castillo E, Schubert L, Jones BL, Faught A, et al. Results of a Multi-Institutional Phase 2 Clinical Trial for 4DCT-Ventilation Functional Avoidance Thoracic Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2022;112(4):986-95.
26. Yaremko BP, Capaldi DPI, Sheikh K, Palma DA, Warner A, Dar AR, et al. Functional Lung Avoidance for Individualized Radiation Therapy: Results of a Double-Masked, Randomized Controlled Trial. International Journal of Radiation Oncology*Biology*Physics. 2022;113(5):1072-84.
27. Faught AM, Olsen L, Schubert L, Rusthoven C, Castillo E, Castillo R, et al. Functional-guided radiotherapy using knowledge-based planning. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2018;129(3):494-8.
28. Ge Y, Wu QJ. Knowledge-based planning for intensity-modulated radiation therapy: A review of data-driven approaches. Medical Physics. 2019;46(6):2760-75.
29. Shwe MA, Tu SW, Fagan LM. Validating the knowledge base of a therapy planning system.

Methods of information in medicine. 1989;28(1):36-50.
30. Nguyen D, Lin MH, Sher D, Lu W, Jia X, Jiang S. Advances in Automated Treatment Planning. Semin Radiat Oncol. 2022;32(4):343-50.
31. Wang C, Zhu X, Hong JC, Zheng D. Artificial Intelligence in Radiotherapy Treatment Planning: Present and Future. 2019;18:1533033819873922.
32. Moore KL. Automated Radiotherapy Treatment Planning. Seminars in Radiation Oncology. 2019;29(3):209-18.
33. Monz M, Küfer KH, Bortfeld TR, Thieke C. Pareto navigation: algorithmic foundation of interactive multi-criteria IMRT planning. Phys Med Biol. 2008;53(4):985-98.
34. Lahanas M, Schreibmann E, Baltas D. Multiobjective inverse planning for intensity modulated radiotherapy with constraint-free gradient-based optimization algorithms. Phys Med Biol. 2003;48(17):2843-71.
35. Barragán-Montero AM, Nguyen D, Lu W, Lin M-H, Norouzi-Kandalan R, Geets X, et al. Three-dimensional dose prediction for lung IMRT patients with deep neural networks: robust learning from heterogeneous beam configurations. Medical Physics. 2019;46(8):3679-91.
36. Fan J, Wang J, Chen Z, Hu C, Zhang Z, Hu W. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. Medical Physics. 2019;46(1):370-81.
37. Kearney V, Chan JW, Haaf S, Descovich M, Solberg TD. DoseNet: a volumetric dose prediction algorithm using 3D fully-convolutional neural networks. Physics in Medicine & Biology. 2018;63(23):235022.
38. Wang W, Sheng Y, Wang C, Zhang J, Li X, Palta M, et al. Fluence Map Prediction Using Deep Learning Models – Direct Plan Generation for Pancreas Stereotactic Body Radiation Therapy. 2020;3.
39. Gao Y, Shen C, Jia X, Kyun Park Y. Implementation and evaluation of an intelligent automatic treatment planning robot for prostate cancer stereotactic body radiation therapy. Radiotherapy and Oncology. 2023;184:109685.
40. Zhang J, Wang C, Sheng Y, Palta M, Czito B, Willett C, et al. An Interpretable Planning Bot for Pancreas Stereotactic Body Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2021;109(4):1076-85.
41. Gintz D, Latifi K, Caudell J, Nelms B, Zhang G, Moros E, et al. Initial evaluation of automated treatment planning software. J Appl Clin Med Phys. 2016;17(3):331-46.
42. Nawa K, Haga A, Nomoto A, Sarmiento RA, Shiraishi K, Yamashita H, et al. Evaluation of a commercial automatic treatment planning system for prostate cancers. Medical Dosimetry. 2017;42(3):203-9.
43. Ahn SH, Kim E, Kim C, Cheon W, Kim M, Lee SB, et al. Deep learning method for prediction of patient-specific dose distribution in breast cancer. Radiation Oncology. 2021;16(1):154.
44. Liao M, Di S, Zhao Y, Liang W, Yang Z. FA-Net: A hierarchical feature fusion and interactive attention-based network for dose prediction in liver cancer patients. Artificial Intelligence in Medicine. 2024;156:102961.
45. Kandalan RN, Nguyen D, Rezaeian NH, Barragán-Montero AM, Breedveld S, Namuduri K, et al. Dose prediction with deep learning for prostate cancer radiation therapy: Model adaptation to different treatment planning practices. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology. 2020;153:228-35.
46. Zhang G, Jiang Z, Zhu J, Wang L. Dose prediction for cervical cancer VMAT patients with a full-

scale 3D-cGAN-based model and the comparison of different input data on the prediction results. Radiation Oncology. 2022;17(1):179.
47. van Genderingen J, Nguyen D, Knuth F, Nomer HAA, Incrocci L, Sharfo AWM, et al. Deep learning dose prediction to approach Erasmus-iCycle dosimetric plan quality within seconds for instantaneous treatment planning. Radiotherapy and Oncology.
48. Yuan S, Frey KA, Gross MD, Hayman JA, Arenberg D, Curtis JL, et al. Semiquantification and Classification of Local Pulmonary Function by V/Q Single Photon Emission Computed Tomography in Patients with Non-small Cell Lung Cancer: Potential Indication for Radiotherapy Planning. Journal of Thoracic Oncology. 2011;6(1):71-8.
49. Pommet R, Therain F. The use of Kr-81m in ventilation imaging. Clinical nuclear medicine. 1982;7(3):122-30.
50. Suga K, Kawakami Y, Zaki M, Yamashita T, Shimizu K, Matsunaga N. Clinical utility of co-registered respiratory-gated(99m)Tc-Technegas/MAA SPECT-CT images in the assessment of regional lung functional impairment in patients with lung cancer. European journal of nuclear medicine and molecular imaging. 2004;31(9):1280-90.
51. Li Z, Le Roux P-Y, Callahan J, Hardcastle N, Hofman MS, Siva S, et al. Quantitative assessment of ventilation-perfusion relationships with gallium-68 positron emission tomography/computed tomography imaging in lung cancer patients. Physics and Imaging in Radiation Oncology. 2022;22:8-12.
52. Lucia F, Bourhis D, Pinot F, Hamya M, Goasduff G, Blanc-Béguin F, et al. Prediction of Acute Radiation-Induced Lung Toxicity After Stereotactic Body Radiation Therapy Using Dose-Volume Parameters From Functional Mapping on Gallium 68 Perfusion Positron Emission Tomography/Computed Tomography. Int J Radiat Oncol Biol Phys . 2024;118(4):952-62.
53. Siva S, Thomas R, Callahan J, Hardcastle N, Pham D, Kron T, et al. High-resolution pulmonary ventilation and perfusion PET/CT allows for functionally adapted intensity modulated radiotherapy in lung cancer. Radiotherapy and Oncology. 2015;115(2):157-62.
54. Yuan ST, Frey KA, Gross MD, Hayman JA, Arenberg D, Curtis JL, et al. Semiquantification and classification of local pulmonary function by V/Q single photon emission computed tomography in patients with non-small cell lung cancer: potential indication for radiotherapy planning. J Thorac Oncol. 2011;6(1):71-8.
55. Tahir BA, Hughes PJC, Robinson SD, Marshall H, Stewart NJ, Norquay G, et al. Spatial Comparison of CT-Based Surrogates of Lung Ventilation With Hyperpolarized Helium-3 and Xenon-129 Gas MRI in Patients Undergoing Radiation Therapy. Int J Radiat Oncol Biol Phys. 2018;102(4):1276-86.
56. Mistry NN, Pollaro J, Song J, De Lin M, Johnson GA. Pulmonary perfusion imaging in the rodent lung using dynamic contrast-enhanced MRI. 2008;59(2):289-97.
57. Bauman G, Puderbach M, Deimling M, Jellus V, Chefd'hotel C, Dinkel J, et al. Non-contrast-enhanced perfusion and ventilation assessment of the human lung by means of fourier decomposition in proton MRI. Magnetic resonance in medicine. 2009;62(3):656-64.
58. Kipritidis J, Tahir BA, Cazoulat G, Hofman MS, Siva S, Callahan J, et al. The VAMPIRE challenge: A multi-institutional validation study of CT ventilation imaging. Medical Physics. 2019;46(3):1198-217.
59. Castillo E, Nair G, Turner-Lawrence D, Myziuk N, Emerson S, Al-Katib S, et al. Quantifying pulmonary perfusion from noncontrast computed tomography. Med Phys. 2021;48(4):1804-14.

60. Huang Y-H, Li Z, Xiong T, Chen Z, Li B, Lou Z, et al. Constructing Surrogate Lung Ventilation Maps from 4DCT-derived Subregional Respiratory Dynamics. International Journal of Radiation Oncology, Biology, Physics.
61. Chen Z, Huang Y-H, Kong F-M, Ho WY, Ren G, Cai J. A super-voxel-based method for generating surrogate lung ventilation images from CT. Frontiers in Physiology. 2023;14.
62. Huang Y-H, Teng X, Zhang J, Chen Z, Ma Z, Ren G, et al. Respiratory Invariant Textures From Static Computed Tomography Scans for Explainable Lung Function Characterization. Journal of Thoracic Imaging. 2023;38(5).
63. Ren G, Lam S-k, Zhang J, Xiao H, Cheung AL-y, Ho W-y, et al. Investigation of a Novel Deep Learning-Based Computed Tomography Perfusion Mapping Framework for Functional Lung Avoidance Radiotherapy. Frontiers in oncology. 2021;11(466).
64. Ren G, Li B, Lam S-k, Xiao H, Huang Y-H, Cheung AL-y, et al. A transfer learning framework for deep learning-based CT-to-perfusion mapping on lung cancer patients. 2022;12:883516.
65. Ren G, Zhang J, Li T, Xiao H, Cheung LY, Ho WY, et al. Deep Learning-Based Computed Tomography Perfusion Mapping (DL-CTPM) for Pulmonary CT-to-Perfusion Translation. International Journal of Radiation Oncology*Biology*Physics. 2021;110(5):1508-18.
66. Liu Z, Miao J, Huang P, Wang W, Wang X, Zhai Y, et al. A deep learning method for producing ventilation images from 4DCT: First comparison with technegas SPECT ventilation. 2020;47(3):1249-57.
67. Xiong T, Zeng G, Chen Z, Huang Y-H, Li B, Zhou D, et al. Automatic planning for functional lung avoidance radiotherapy based on function-guided beam angle selection and plan optimization. Physics in Medicine & Biology. 2024;69(15):155007.
68. Parker JA, Coleman RE, Grady E, Royal HD, Siegel BA, Stabin MG, et al. SNM Practice Guideline for Lung Scintigraphy 4.0. 2012;40(1):57-65.
69. Yamamoto T, Kabus S, Bal M, Keall P, Benedict S, Daly M. The first patient treatment of computed tomography ventilation functional image-guided radiotherapy for lung cancer. Radiotherapy and Oncology. 2016;118(2):227-31.
70. Gooding MJ, Aluwini S, Guerrero Urbano T, McQuinlan Y, Om D, Staal FHE, et al. Fully automated radiotherapy treatment planning: A scan to plan challenge. Radiotherapy and Oncology. 2024:110513.
71. Maass K, Aravkin A, Kim M. A hyperparameter-tuning approach to automated inverse planning. 2022;49(5):3405-15.
72. Chanyavanich V, Das SK, Lee WR, Lo JY. Knowledge-based IMRT treatment planning for prostate cancer. Med Phys. 2011;38(5):2515-22.
73. Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, Jacques R, et al. Data-Driven Approach to Generating Achievable Dose–Volume Histogram Objectives in Intensity-Modulated Radiotherapy Planning. International Journal of Radiation Oncology*Biology*Physics. 2011;79(4):1241-7.
74. Zhang X, Li X, Quan EM, Pan X, Li Y. A methodology for automatic intensity-modulated radiation treatment planning for lung cancer. Phys Med Biol. 2011;56(13):3873-93.
75. Wu B, McNutt T, Zahurak M, Simari P, Pang D, Taylor R, et al. Fully automated simultaneous integrated boosted-intensity modulated radiation therapy treatment planning is feasible for head-and-neck cancer: a prospective clinical study. Int J Radiat Oncol Biol Phys. 2012;84(5):e647-53.
76. Zhu X, Ge Y, Li T, Thongphiew D, Yin F-F, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. 2011;38(2):719-26.

77. Fogliata A, Wang P-M, Belosi F, Clivio A, Nicolini G, Vanetti E, et al. Assessment of a model based optimization engine for volumetric modulated arc therapy for patients with advanced hepatocellular cancer. Radiation Oncology. 2014;9(1):236.
78. Amit G, Purdie TG, Levinshtein A, Hope AJ, Lindsay P, Marshall A, et al. Automatic learning-based beam angle selection for thoracic IMRT. Med Phys. 2015;42(4):1992-2005.
79. Boutilier JJ, Lee T, Craig T, Sharpe MB, Chan TCY. Models for predicting objective function weights in prostate cancer IMRT. 2015;42(4):1586-95.
80. Shiraishi S, Moore KL. Knowledge-based prediction of three-dimensional dose distributions for external beam radiotherapy. 2016;43(1):378-87.
81. Chung CV, Khan MS, Olanrewaju A, Pham M, Nguyen QT, Patel T, et al. Knowledge-based planning for fully automated radiation therapy treatment planning of 10 different cancer sites. Radiotherapy and Oncology.
82. Harms J, Zhang J, Kayode O, Wolf J, Tian S, McCall N, et al. Implementation of a Knowledge-Based Treatment Planning Model for Cardiac-Sparing Lung Radiation Therapy. Advances in radiation oncology. 2021;6(6):100745.
83. Momin S, Wolf J, Roper J, Lei Y, Liu T, Bradley JD, et al. Enhanced cardiac substructure sparing through knowledge-based treatment planning for non-small cell lung cancer radiotherapy. 2022;12.
84. Breedveld S, Craft D, van Haveren R, Heijmen B. Multi-criteria optimization and decision-making in radiotherapy. European Journal of Operational Research. 2019;277(1):1-19.
85. Craft D, Bortfeld T. How many plans are needed in an IMRT multi-objective plan database? Phys Med Biol. 2008;53(11):2785-96.
86. Kyroudi A, Petersson K, Ghandour S, Pachoud M, Matzinger O, Ozsahin M, et al. Discrepancies between selected Pareto optimal plans and final deliverable plans in radiotherapy multi-criteria optimization. Radiotherapy and Oncology. 2016;120(2):346-8.
87. Jee KW, McShan DL, Fraass BA. Lexicographic ordering: intuitive multicriteria optimization for IMRT. Phys Med Biol. 2007;52(7):1845-61.
88. Zarepisheh M, Hong L, Zhou Y, Oh JH, Mechalakos JG, Hunt MA, et al. Automated intensity modulated treatment planning: The expedited constrained hierarchical optimization (ECHO) system. Med Phys. 2019;46(7):2944-54.
89. Wang Q, Wang R, Liu J, Jiang F, Yue H, Du Y, et al. High-dimensional automated radiation therapy treatment planning via Bayesian optimization. 2023;50(6):3773-87.
90. Huang C, Nomura Y, Yang Y, Xing L. Meta-optimization for fully automated radiation therapy treatment planning. Physics in Medicine & Biology. 2022;67(5).
91. Holdsworth C, Kim M, Liao J, Phillips M. The use of a multiobjective evolutionary algorithm to increase flexibility in the search for better IMRT plans. Med Phys. 2012;39(4):2261-74.
92. Yang D, Wu X, Li X, Mansfield R, Xie Y, Wu Q, et al. Automated treatment planning with deep reinforcement learning for head-and-neck (HN) cancer intensity modulated radiation therapy (IMRT). Physics in Medicine & Biology. 2024.
93. Wang W, Sheng Y, Palta M, Czito B, Willett C, Yin FF, et al. Transfer learning for fluence map prediction in adrenal stereotactic body radiation therapy. Phys Med Biol. 2021;66(24).
94. Li Y, Cai W, Xiao F, Zhou X, Cai J, Zhou L, et al. Simultaneous dose distribution and fluence prediction for nasopharyngeal carcinoma IMRT. Radiation Oncology. 2023;18(1):110.
95. Teng L, Wang B, Xu X, Zhang J, Mei L, Feng Q, et al. Beam-wise dose composition learning for head and neck cancer dose prediction in radiotherapy. Medical Image Analysis. 2024;92:103045.

96. Gao R, Lou B, Xu Z, Comaniciu D, Kamen A. Flexible-Cm GAN: Towards Precise 3D Dose Prediction in Radiotherapy. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR): IEEE Computer Society; 2023. p. 715-25.
97. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-49.
98. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. Lancet (London, England). 2021;398(10299):535-54.
99. Vinod SK, Hau E. Radiotherapy treatment for lung cancer: Current status and future directions. Respirology. 2020;25:61-71.
100. Hanania AN, Mainwaring W, Ghebre YT, Hanania NA, Ludwig M. Radiation-Induced Lung Injury: Assessment and Management. Chest. 2019;156(1):150-62.
101. Ireland RH, Tahir BA, Wild JM, Lee CE, Hatton MQ. Functional Image-guided Radiotherapy Planning for Normal Lung Avoidance. Clinical Oncology. 2016;28(11):695-707.
102. Li S, Liu J, Gao S, Yin Y, Zhang L, Han Y, et al. CT ventilation image-guided helical Tomotherapy at sparing functional lungs for locally advanced lung cancer: analysis of dose-function metrics and the impact on pulmonary toxicity. Radiation Oncology. 2023;18(1):6.
103. McGuire SM, Zhou S, Marks LB, Dewhirst M, Yin FF, Das SK. A methodology for using SPECT to reduce intensity-modulated radiation therapy (IMRT) dose to functioning lung. Int J Radiat Oncol Biol Phys. 2006;66(5):1543-52.
104. Zhou PX, Zhang SX. Functional lung imaging in thoracic tumor radiotherapy: Application and progress. Frontiers in oncology. 2022;12:908345.
105. McGuire SM, Marks LB, Yin FF, Das SK. A methodology for selecting the beam arrangement to reduce the intensity-modulated radiation therapy (IMRT) dose to the SPECT-defined functioning lung. Physics in Medicine & Biology. 2010;55(2):403.
106. Bedford JL, Ahmed M. Functional lung avoidance in radiotherapy using optimisation of biologically effective dose with non-coplanar beam orientations. Physics and Imaging in Radiation Oncology. 2023;28:100518.
107. Yuan L, Zhu W, Ge Y, Jiang Y, Sheng Y, Yin F-F, et al. Lung IMRT planning with automatic determination of beam angle configurations. Physics in Medicine & Biology. 2018;63(13):135024.
108. Bria WF, Kanarek DJ, Kazemi H. Prediction of postoperative pulmonary function following thoracic operations. Value of ventilation-perfusion scanning. The Journal of thoracic and cardiovascular surgery. 1983;86(2):186-92.
109. Hoover DA, Reid RH, Wong E, Stitt L, Sabondjian E, Rodrigues GB, et al. SPECT-based functional lung imaging for the prediction of radiation pneumonitis: A clinical and dosimetric correlation. J Med Imaging Radiat Oncol. 2014;58(2):214-22.
110. Khalil AA, Hau E, Gebski V, Grau C, Gee H, Nyeng TB, et al. Personal innovative approach in radiation therapy of lung cancer- functional lung avoidance SPECT-guided (ASPECT) radiation therapy: a study protocol for phase II randomised double-blind clinical trial. BMC cancer. 2021;21(1):940.
111. Yamamoto T, Kabus S, von Berg J, Lorenz C, Keall PJ. Impact of four-dimensional computed tomography pulmonary ventilation imaging-based functional avoidance for lung cancer radiotherapy. Int J Radiat Oncol Biol Phys. 2011;79(1):279-88.
112. Meyer J, Hummel SM, Cho PS, Austin-Seymour MM, Phillips MH. Automatic selection of non-

coplanar beam directions for three-dimensional conformal radiotherapy. Br J Radiol. 2005;78(928):316-27.
113. Amit G, Purdie TG, Levinshtein A, Hope AJ, Lindsay P, Marshall A, et al. Automatic learning-based beam angle selection for thoracic IMRT. Medical physics. 2015;42(4):1992-2005.
114. Yuan L, Wu QJ, Yin F, Li Y, Sheng Y, Kelsey CR, et al. Standardized beam bouquets for lung IMRT planning. Physics in medicine and biology. 2015;60(5):1831-43.
115. Lee D, Wiswall M. A Parallel Implementation of the Simplex Function Minimization Routine. Computational Economics. 2007;30(2):171-87.
116. Nelder JA, Mead R. A Simplex Method for Function Minimization. The Computer Journal. 1965;7(4):308-13.
117. Wieser H-P, Cisternas E, Wahl N, Ulrich S, Stadler A, Mescher H, et al. Development of the open-source dose calculation and optimization toolkit matRad. Medical Physics. 2017;44(6):2556-68.
118. Bortfeld T, Schlegel W, Rhein B. Decomposition of pencil beam kernels for fast dose calculations in three-dimensional treatment planning. Medical physics. 1993;20(2):311-8.
119. Siochi RAC. Minimizing static intensity modulation delivery time using an intensity solid paradigm. International Journal of Radiation Oncology*Biology*Physics. 1999;43(3):671-80.
120. McIntosh C, Welch M, McNiven A, Jaffray DA, Purdie TG. Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method. Physics in Medicine & Biology. 2017;62(15):5926.
121. Huang TC, Hsiao CY, Chien CR, Liang JA, Shih TC, Zhang GG. IMRT treatment plans and functional planning with functional lung imaging from 4D-CT for thoracic cancer patients. Radiation oncology (London, England). 2013;8:3.
122. Munawar I, Yaremko BP, Craig J, Oliver M, Gaede S, Rodrigues G, et al. Intensity modulated radiotherapy of non-small-cell lung cancer incorporating SPECT ventilation imaging. Medical Physics. 2010;37(4):1863-72.
123. Lavrenkov K, Singh S, Christian JA, Partridge M, Nioutsikou E, Cook G, et al. Effective avoidance of a functional spect-perfused lung using intensity modulated radiotherapy (IMRT) for non-small cell lung cancer (NSCLC): An update of a planning study. Radiotherapy and Oncology. 2009;91(3):349-52.
124. Babier A, Zhang B, Mahmood R, Moore KL, Purdie TG, McNiven AL, et al. OpenKBP: The open-access knowledge-based planning grand challenge and dataset. Medical Physics. 2021;48(9):5549-61.
125. Nguyen D, Jia X, Sher D, Lin M-H, Iqbal Z, Liu H, et al. 3D radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected U-net deep learning architecture. Physics in Medicine & Biology. 2019;64(6):065020.
126. Wang E, Abdallah H, Snir J, Chong J, Palma DA, Mattonen SA, et al. Predicting the 3-Dimensional Dose Distribution of Multi-Lesion Lung Stereotactic Ablative Radiotherapy with Generative Adversarial Networks. International Journal of Radiation Oncology*Biology*Physics. 2024.
127. Xiong T, Zeng G, Chen Z, Huang Y-H, Li B, Ma Z, et al. Automatic lung dose painting for functional lung avoidance radiotherapy through multi-modality-guided dose prediction. Physics in Medicine & Biology. 2026;71(1):015037.
128. Roman NO, Shepherd W, Mukhopadhyay N, Hugo GD, Weiss E. Interfractional positional variability of fiducial markers and primary tumors in locally advanced non-small-cell lung cancer during audiovisual biofeedback radiotherapy. International Journal of Radiation

Oncology*Biology*Physics. 2012;83(5):1566-72.
129. Balik S, Weiss E, Jan N, Roman N, Sleeman WC, Fatyga M, et al. Evaluation of 4-dimensional computed tomography to 4-dimensional cone-beam computed tomography deformable image registration for lung cancer adaptive radiation therapy. International Journal of Radiation Oncology*Biology*Physics. 2013;86(2):372-9.
130. Yamamoto T, Kabus S, Lorenz C, Mittra E, Hong JC, Chung M, et al. Pulmonary Ventilation Imaging Based on 4-Dimensional Computed Tomography: Comparison With Pulmonary Function Tests and SPECT Ventilation Images. International Journal of Radiation Oncology*Biology*Physics. 2014;90(2):414-22.
131. Kipritidis J, Siva S, Hofman MS, Callahan J, Hicks RJ, Keall PJ. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles. Medical Physics. 2014;41(1):011910.
132. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, editors. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016; 2016 2016//; Cham: Springer International Publishing.
133. Hu J, Shen L, Sun G, editors. Squeeze-and-excitation networks. Proceedings of the IEEE conference on computer vision and pattern recognition; 2018.
134. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.
135. Jhanwar G, Dahiya N, Ghahremani P, Zarepisheh M, Nadeem S. Domain knowledge driven 3D dose prediction using moment-based loss function. Physics in Medicine & Biology. 2022;67(18):185017.
136. van Bruggen IG, van Dijk M, Brinkman-Akker MJ, Löfman F, Langendijk JA, Both S, et al. Clinical implementation of deep learning robust IMPT planning in oropharyngeal cancer patients: A blinded clinical study. Radiotherapy and Oncology.
137. Lombardo J, Castillo E, Castillo R, Miller R, Jones B, Miften M, et al. Prospective Trial of Functional Lung Avoidance Radiation Therapy for Lung Cancer: Quality of Life Report. International Journal of Radiation Oncology*Biology*Physics. 2024;120(2):593-602.
138. Dahiya N, Jhanwar G, Yezzi A, Zarepisheh M, Nadeem SJapa. Deep Learning 3D Dose Prediction for Conventional Lung IMRT Using Consistent/Unbiased Automated Plans. 2021.
139. Dahiya N, Jhanwar G, Yezzi AJ, Zarepisheh M, Nadeem S. Deep Learning 3D Dose Prediction for Conventional Lung IMRT Using Consistent/Unbiased Automated Plans. 2021;abs/2106.03705.
140. Gronberg MP, Beadle BM, Garden AS, Skinner H, Gay S, Netherton T, et al. Deep Learning-Based Dose Prediction for Automated, Individualized Quality Assurance of Head and Neck Radiation Therapy Plans. Pract Radiat Oncol. 2023;13(3):e282-e91.
141. van Genderingen J, Nguyen D, Knuth F, Nomer HAA, Incrocci L, Sharfo AWM, et al. Deep learning dose prediction to approach Erasmus-iCycle dosimetric plan quality within seconds for instantaneous treatment planning. Radiotherapy and Oncology. 2025;203:110662.
142. Gao R, Diallo M, Liu H, Magliari A, Sackett J, Verbakel WFAR, et al. Automating High Quality RT Planning at Scale. 2025;abs/2501.11803.
143. Gao R, Diallo M, Liu H, Magliari A, Sackett J, Verbakel W, et al. Automating High Quality RT Planning at Scale. 2025.
144. Kida S, Bal M, Kabus S, Negahdar M, Shan X, Loo BW, et al. CT ventilation functional image-based IMRT treatment plans are comparable to SPECT ventilation functional image-based plans.

Radiotherapy and Oncology. 2016;118(3):521-7.
145. Vinogradskiy Y, Castillo R, Castillo E, Tucker SL, Liao Z, Guerrero T, et al. Use of 4-dimensional computed tomography-based ventilation imaging to correlate lung dose and function with clinical outcomes. Int J Radiat Oncol Biol Phys. 2013;86(2):366-71.
146. Faught AM, Miyasaka Y, Kadoya N, Castillo R, Castillo E, Vinogradskiy Y, et al. Evaluating the Toxicity Reduction With Computed Tomographic Ventilation Functional Avoidance Radiation Therapy. International Journal of Radiation Oncology*Biology*Physics. 2017;99(2):325-33.
147. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. International Journal of Radiation Oncology, Biology, Physics. 2010;76(3):S3-S9.
148. Werner-Wasik M, Yorke E, Deasy J, Nam J, Marks LB. Radiation Dose-Volume Effects in the Esophagus. International Journal of Radiation Oncology, Biology, Physics. 2010;76(3):S86-S93.
149. Kirkpatrick JP, van der Kogel AJ, Schultheiss TE. Radiation Dose-Volume Effects in the Spinal Cord. International Journal of Radiation Oncology, Biology, Physics. 2010;76(3):S42-S9.
150. Gagliardi G, Constine LS, Moiseenko V, Correa C, Pierce LJ, Allen AM, et al. Radiation Dose-Volume Effects in the Heart. Int J Radiat Oncol Biol Phys . 2010;76(3):S77-S85.
151. Rhee DJ, Jhingran A, Huang K, Netherton TJ, Fakie N, White I, et al. Clinical acceptability of fully automated external beam radiotherapy for cervical cancer with three different beam delivery techniques. Medical Physics. 2022;49(9):5742-51.
152. Schaub SK, Apisarnthanarax S, Price RG, Nyflot MJ, Chapman TR, Matesan M, et al. Functional Liver Imaging and Dosimetry to Predict Hepatotoxicity Risk in Cirrhotic Patients With Primary Liver Cancer. International Journal of Radiation Oncology, Biology, Physics. 2018;102(4):1339-48.
153. Lei X, Cristian C, Sandeep H, Arthur LB, Elfar A, Daniel S. Inverse planning for functional image-guided intensity-modulated radiation therapy. Physics in Medicine & Biology. 2002;47(20):3567.
154. Dougherty JM, Castillo E, Castillo R, Faught AM, Pepin M, Park SS, et al. Functional avoidance-based intensity modulated proton therapy with 4DCT derived ventilation imaging for lung cancer. 2021;22(7):276-85.
155. Chen Z, Xiong T, Huang Y-H, Li B, Ma Z, Ge H, et al. Functional lung avoidance proton radiation therapy planning using anatomy-wise computed tomography-derived lung ventilation imaging. Radiation Oncology. 2026.
156. Huang Y-H, Li Z, Xiong T, Chen Z, Li B, Lou Z, et al. Constructing Surrogate Lung Ventilation Maps from 4DCT-derived Subregional Respiratory Dynamics. International Journal of Radiation Oncology*Biology*Physics. 2024.
157. Wang M, Liu X, Li H, Zhao M, Xiong T, Huang D, et al. Simultaneous synthesis of perfusion and ventilation images from CT using a dual-decoder residual attention network for lung disease diagnosis. Journal of Applied Clinical Medical Physics. 2026;27(3):e70498.
158. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Conf on Computer Vision and Pattern Recognition2016.
159. Roy S, Koehler G, Ulrich C, Baumgartner M, Petersen J, Isensee F, et al. MedNeXt: Transformer-driven Scaling of ConvNets for Medical Image Segmentation. 2023;abs/2303.09975.
160. Yue M, Xue X, Wang Z, Lambo RL, Zhao W, Xie Y, et al. Dose prediction via distance-guided deep learning: Initial development for nasopharyngeal carcinoma radiotherapy. Radiotherapy and Oncology. 2022;170:198-204.