Original Article

Integrating Deep Learning-Based Dose Distribution Prediction with Bayesian Networks for Decision Support in Radiotherapy for Upper Gastrointestinal cancer

Dong-Yun Kim¹,²,*, Bum-Sup Jang¹,*, Eunji Kim¹,³, Eui Kyu Chie¹,⁴

¹Department of Radiation Oncology, Seoul National University College of Medicine, Seoul,
²Department of Radiation Oncology, Chung-Ang University Hospital, Seoul,
³Department of Radiation Oncology, Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul,
⁴Institute of Radiation Medicine, Medical Research Center, Seoul National University, Seoul, Korea

*Dong-Yun Kim and Bum-Sup Jang contributed equally to this work.

Correspondence: Eui Kyu Chie

Department of Radiation Oncology, Seoul National University College of Medicine, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea

Tel: 82-2-2072-3705   Fax: 82-2-765-3317   E-mail: ekchie93@snu.ac.kr

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi:10.4143/crt.2024.333

1

Korean Cancer Association
This article is protected by copyright. All rights reserved.
Abstract

Purpose
Selecting the better techniques to harbor optimal motion management, either a stereotactic linear accelerator delivery using TrueBeam (TBX) or Magnetic Resonance (MR)-guided gated delivery using MRIdian (MRG), is time-consuming and costly. To address this challenge, we aimed to develop a decision-supporting algorithm based on a combination of deep learning-generated dose distributions and clinical data.

Materials and Methods
We retrospectively analyzed 65 patients with liver or pancreatic cancer who underwent both TBX and MRG simulations and planning process. We trained three-dimensional U-Net deep learning models to predict dose distributions and generated dose volume histograms (DVHs) for each system. We integrated predicted DVH metrics into a Bayesian network (BN) model incorporating clinical data.

Results
The MRG prediction model outperformed the TBX model, demonstrating statistically significant superiorities in predicting normalized dose to the PTV and liver. We developed a final BN prediction model integrating the predictive DVH metrics with patient factors like age, PTV size, and tumor location. This BN model an area under the receiver operating characteristic curve index of 83.56%. The decision tree derived from the BN model showed that the tumor location (abutting vs. apart of PTV to hollow viscus organs) was the most important factor to determine TBX or MRG.

Conclusion
We demonstrated a decision-supporting algorithm for selecting optimal RT plans in upper gastrointestinal cancers, incorporating both deep learning-based dose prediction and BN-based
treatment selection. This approach might streamline the decision-making process, saving resources and improving treatment outcomes for patients undergoing RT.

**Keywords**

Deep learning, Upper GI cancer, Linear accelerator-based treatment plan, MR-guided treatment plan, Bayesian network, Decision-supporting algorithm
Introduction

Advances in image-guided radiation therapy (RT), which involve precise localization of the neighboring organ at risk as well as treatment target prior to therapy, have greatly enhanced the accuracy of RT [1–3]. The RT planning process for patients with upper gastrointestinal cancer is recognized as a complex and time-intensive endeavor, not only due to the movement of the treatment target, but primarily due to the proximity of multiple critical organs at risk (OARs) [4–6]. Within this context, organs such as the duodenum and stomach are situated in a close proximity to the RT targets [4,5], while the liver and other surrounding organs can exhibit substantial motion during respiration [6–8]. To address the challenges posed by respiratory-induced movement and the physiological mobility of internal organs, various RT techniques have been employed. These techniques include gating, which administers RT at specific respiratory phase, internal targeted volume therapy, which encompasses the entire tumor movement, forced respiratory depression that reduces breathing volume through abdominal compression, and tracking, which involves tracking of the tumor motion [9,10].

The magnetic resonance (MR) RT machine boasts a significant advantage in the realm of RT due to its capacity for real-time magnetic resonance imaging (MRI) [11-13]. This unique feature provides a higher resolution for soft tissues compared to computed tomography (CT) or conventional X-ray imaging. Consequently, it simplifies and enhances the differentiation of tumors from surrounding normal tissues [4,14,15]. Within the institution, in addition to the aforementioned MR-guided gated delivery, which utilized MRIdian® (Viewray Inc, OH, MRG), the CT-based high dose rate stereotactic linear accelerator delivery using TrueBeam-STX™ coupled with Eclipse Treatment Planning System (Varian Inc. CA, TBX) method combined with forced respiratory depression was available. These approaches are chosen interchangeably based on expert judgment. However, in a real-world practice, superiority of
one technique over the other would not be readily noticeable, requiring simulation and
treatment planning to select the superior RT plan that offers clinical benefits. To conduct a
comparison of the dose distribution between the two plans, patients are inevitably subjected to
simulations for both plans and must endure the waiting period required for treatment plan
generation. Moreover, each plan is semi-supervised, necessitating the participation of several
specialists, including radiation oncologists and dosimetrists, which places a significant strain
on available resources.

To address these challenges, ongoing research focused on predicting dose distribution
using deep learning techniques, such as U-Nets adopting convolutional neural networks
(CNN), or generative adversarial networks (GAN) [16–20]. Deep learning models have the
potential to incorporate patient-specific factors and capture spatial dose correlations that go
beyond individual voxels within a simulation CT-based plan [16]. Much of these works have
been concentrated in specific tumor locations, such as head and neck, breast, and prostate
cancers, where the target position experiences relatively minimal changes due to various
internal motions including respiratory and physiological motions [17,18,21]. However, there is
a noticeable absence of relevant studies in cases where the prediction of RT dose distribution
is more critical due to significant physiological movements caused by breathing and internal
organ motion, as observed in upper abdominal tumors.

A Bayesian network (BN) is a statistical framework used to represent conditional
dependencies among variables by means of a directed acyclic graph (DAG). The nodes within
the DAG correspond to clinical variables, and the edges indicate conditional dependencies
between them. This framework allows for the accommodation of heterogeneity among clinical
and non-clinical variables and offers clinicians interpretable clinical probabilities.

Thus, the primary objective of this study was to develop a model capable of predicting
hypothetical dose distributions without the necessity of creating actual treatment plans using both systems. Further, we aimed to establish a decision-supporting algorithm using BN models that can be integrated with clinical data of individual patient, streamlining the decision-making process in RT planning.

Materials and methods

1. Data acquisition and preprocessing

This retrospective study included a total of 65 patients with primary liver cancer (n=40) or pancreatic cancer (n=25) treated with RT between 2016 and 2018. To compare the dose distribution between the two plans, we performed simulations and plan generation for TBX and MRG treatment plans for all patients. Both treatment plans typically involved maximum possible constraint of OARs, such as the duodenum, stomach, liver, and kidneys, to minimize RT exposure to these structures. Routine dose constraints were applied to ensure that the radiation dose to the OAR remained within safe limits. The two treatments were based on respiratory motion-controlled RT technique. TBX applied devices such as plates or wraps to apply forced respiratory depression, while MRG used gating. CT and MR images for simulation as well as the associated dose plans and contoured RT structures, were retrospectively collected as DICOM files.

In terms of CT/ MR images, we created a three-dimensional object per patient, consisting of three layers: CT/MR images, masks, and doses. The CT/MR images were normalized by subtracting the mean and dividing standard deviation. Outliers exceeding pixel values of 5 or -5 were assigned as 0, and the images were rescaled to range from 0 to 1. The reason for rescaling to the range (0, 1) was to ensure uniformity and enhance the quality of the images for further analysis. Masks were generated by extracting the contours of the body,
duodenum, right and left kidneys, liver, and stomach, and assigning the values 1, 2, 3, 4, and 5, respectively. Doses were registered to the PTV contour using an intensity-based automatic image registration method. Intensity-based automatic image registration is an iterative process which corresponds pixels in each image according to their relative intensity patterns [22]. Finally, all layered images were cropped using a bounding-box technique around the OARs. Assigned value was 0 for some OARs that were not delineated.

In addition, a medical review of the clinical data of the accrued patients was conducted. This data included patient age, gender, performance in European Cooperative Oncology Group (ECOG) score, Child-Pugh score, total RT dose, number of fractions, RT technique (Intensity Modulated Radiotherapy [IMRT] or Stereotactic Ablative Body Radiotherapy [SABR]), gross tumor volume (GTV) in cubic centimeters (cc), planning target volume (PTV) and tumor location (abutting vs. apart). Tumor location was categorized as "abutting" if the PTV was in contact with a neighboring hollow viscus or "apart" if there was spatial separation, based on the geographical relationship of the PTV to the OARs (S1 Fig.). Definition of abutting was derived from the concept of ‘abutment’ of pancreatic tumors [23]. All pre-processing, training and testing were performed by using MATLAB software 2022a.

2. Training and internal validation of a 3D U-Net architecture

To learn and predict the dose distribution, we used the 3D U-Net layers provided by MATLAB software [24]. We removed the SoftMax layer and changed the terminal segmentation layers to regression output layers. Finally, we constructed a deep learning network consisting of 57 layers with 19 million learnable parameters. The overall network schema to construct the dose distribution model and two networks were created, one for MRG dose distribution (Fig. 1A) and one for TBX dose distribution (Fig. 1B). We trained the
networks on the input volume by dividing the images into small, random 3D patches. We used a patch size of $92 \times 92 \times 92$ voxels and a mini-batch size of 4 for the TBX deep learning model, and a patch size of $132 \times 132 \times 132$ voxels and a mini-batch size of 6 for the MRG deep learning model. These parameters were determined by the available GPU memory. This meant that while GPU capacity was limited to the same size, the different sizes of TBX-planned CT images and MRG-planned MR images resulted in different amounts of information being processed, leading to differences in patch size and mini-batch size between the two modalities.

Training was performed on a randomly selected 80% of the dataset, while maintaining the ratio of target location. The remaining 20% of the dataset was used for validation, resulting in a 5-fold training and validation method. For each fold, the initial learning rate was 0.1 and decreased by 10% for every 5 epochs, to a total of 200 epochs. These training parameters were determined empirically, considering available computing resources. These steps were performed for both TBX and MRG dose prediction model. The accuracy of the developed 3D U-Net model can be evaluated by comparing the results of the inference. The key result of the inference was the dose volume histogram (DVH) profiles, the differences between the ground truths and the predicted values. It was calculated as normalized maximum, and mean doses of the PTV and OAR’s. We used a calculated normalized dose (range, 0-1) to make the results more intuitive to interpret.

3. Bayesian network (BN) algorithm

In current analyses, we integrated the generated DVH profiles into the BN framework by incorporating clinical data (Fig. 1C). We employed the tree-augmented Markov blanket method (TAMB) and assessed the model performance using 5-fold cross-validation. We adopted TAMB due to following reasons. First, it utilizes a tree structure to make the
relationships between data more intuitively understandable. Also, it efficiently handles high-dimensional datasets while maintaining model parsimony. Additionally, the method offers robustness to noisy or incomplete data, thanks to its ability to incorporate information from the Markov blanket of each variable, thereby improving the overall reliability of the learned network [25]. We imputed missing values by using static imputation method, which performed the most satisfactory result after several attempts. Supervised learning was performed with a structural coefficient set at 0.4, to predict the selection of treatment modality. We evaluated model performance by measuring the mean area under the receiver operating characteristic curve (AUC) and calibration index. Additionally, we visualized the influence of each node on the mean value of the class variable while keeping the probability distribution of the other variables constant. All analyses were conducted using BayesiaLab 10.1 (Bayesia S.A.S, France).

Results

Our study included 65 patients with liver or pancreas cancers who underwent RT planning with both TBX and MRG systems. The median age was 64.8 years (range 40-78 years). Among patients, 40 patients (61.5%) had liver cancer, while 25 (38.5%) had pancreatic cancer. The primary aim of RT was salvage therapy for 45 patients (69.2%), followed by definitive therapy for 14 patients (21.5%). Most patients (92.3%) had a Child-Pugh classification of A, with the remaining 7.7% had a classification of B. As for the location of the tumor, 16 patients had tumors adjacent to hollow viscus such as bowel (‘abutting’) and 49 patients had tumors ‘apart’ from it. Regarding treatment delivery, 18 patients received MRG and 47 underwent TBX. SABR was the prescribed modality for 61.5% of patients, while 38.5% received IMRT. The mean dose for SABR was 48.8 Gy (range, 30-60 Gy) with a median
fractionation of 4 (range, 4-8). Similarly, the mean dose for IMRT was 50.8 Gy (range, 40-60 Gy) with a median fractionation of 10 (range, 10-30). The median PTV volume was 49.2 cc (range, 10-660 cc), with 49.2% of patients had a PTV volume below 35 cc and 50.8% above 35 cc. More detailed patient characteristics are provided in Table 1.

We developed a 3D U-Net deep learning model of each of TBX and MRG to predict 3D dose distribution. Fig. 2 showed how the developed model predicted on DVHs, specifically the difference between the ground truth and predicted dose. In Fig. 2, the solid lines are the ground truth values of PTV and OARs, and the dashed lines are the predicted dose by the model. Table 2 described the maximum and mean differences in normalized dose between ground truth and predicted values for the TBX and MRG dose prediction models. Overall, the difference of normalized dose of the MRG prediction model was smaller than that of the TBX. The normalized maximum dose differences for the PTV, duodenum, kidney, liver, and stomach were 0.06, 0.08, 0.10, 0.08, and 0.11, respectively, for the MRG prediction model, and 0.18, 0.14, 0.13, 0.19, and 0.12, respectively, for the TBX dose prediction model. And the normalized dose differences of mean value for the PTV, duodenum, kidney, liver, and stomach were 0.04, 0.03, 0.04, and 0.04, respectively, for the MRG model, and 0.11, 0.04, 0.05, 0.05, and 0.03, respectively, for the TBX model. For both models, the mean normalized dose differences were relatively smaller than the maximum dose differences, indicating better prediction regarding mean dose. Statistically significant differences between the two models were found in the maximum normalized metrics of the PTV (p < 0.001), duodenum (p=0.020), and liver (p<0.001). Similarly, the mean normalized dose differences of the PTV and liver were also significantly different (p<0.001 and p=0.031, respectively) (Fig. 3). Overall, the predicted normalized DVH metrics were better for the MRG prediction models.

Based on clinical and DVH profile data, we developed a final BN prediction model to
help clinicians decide whether to use TBX or MRG to treat patients. This BN model was established by using several clinical factors as well as predicted maximum and mean normalized doses. As illustrated in S2 Fig., the final BN model was composed of age, location, predicted maximum normalized dose of the stomach in the MRG plan, and the size of PTV. Five-fold validation of the BN model demonstrated excellent performance of predicting treatment modality selection (Table 3). The ROC indices for both TBX and MRG models were notably high, with MRG slightly outperforming TBX, indicating strong classification performance in distinguishing between the two treatment modalities. In terms of calibration, the MRG model exhibited better alignment between predicted probabilities and actual outcomes compared to MRG, as reflected by higher calibration indices. Despite similar levels of prediction error, as indicated by the binary log-loss values being identical for both models, both TBX and MRG demonstrated comparable overall precision and reliability, with roughly 83.08% of predictions being correct and approximately 82.49% consistency in predictions. Moreover, the overall ROC index, which provided an average assessment of classification ability across all classes, remain consistently high of 83.56% between TBX and MRG models.

Based on the factors illustrated in final BN model, the decision tree was generated and depicted in Fig. 4, summarizing the algorithms influencing treatment plan selection. The most important factor to determine TBX vs MRG was the tumor location (abutting vs. apart). And the 2nd most influential factor was maximum normalized dose to the stomach in MRG plan. Considering the score by TAMB method described in Fig. 4, PTV size and age might also influence treatment decisions, but their relevance was relatively low compared to aforementioned two factors. Based on the decision tree, the model strongly recommends MRG when: The tumor directly abuts hollow viscus and the maximum normalized dose to the stomach in the MRG is below 0.56 ([A] in Fig. 4). TBX is preferred when: First, the tumor
abuts hollow organs, the maximum normalized dose to the stomach in the MRG plan is above 0.56, the PTV is less than 65.7cc, and the patient is over 67 years old ([B] in Fig. 4). Second, the tumor is not in direct contact with hollow viscus organs and the maximum normalized dose to the stomach in the MRG plan is not below 0.2 ([C] in Fig. 4). Third, the tumor is not in direct contact with hollow organs, the maximum normalized dose to the stomach in the MRG plan is below 0.2, the PTV volume is greater than 65.7cc ([D] in Fig. 4).

Discussion

This study introduced a novel decision-supporting algorithm that facilitated the selection of an optimal treatment plan for patients with upper GI cancers undergoing RT. This model empowered physicians to choose swiftly between two available options, TBX vs. MRG. A 3D U-Net deep learning model was developed to predict the 3D dose distribution of both RT plans. Subsequently, DVH metrics derived from the 3D U-Net model were integrated with relevant clinical data to establish a final BN model to determine superiority of MRG versus TBX.

The MRG for GI cancers has become increasingly popular technique. This is due to the superior soft tissue visualization offered by MR imaging, leading to improved target delineation especially in GI cancers [26]. However, the widespread adoption of MR-guided RT for all patients with GI cancer remained limited by resource constraints including longer treatment times, limited availability in healthcare facilities, and high cost. Furthermore, the traditional workflow for RT planning of abdominal tumors necessitates the creation of two distinct treatments, one for MR based and another CT based, to select the better plan for each patient, potentially leading to treatment delays and disrupting continuum of patient care. This approach accounts for the complex nature of upper GI tumors, where targets lie in the close
proximity to surrounding hollow viscus and experience substantial positional fluctuations due to physiological movement [4-8]. Thus, immediate dose prediction for both MR based and CT based plan was indeed an unmet need.

Several deep learning models have been proposed for dose prediction in RT planning. Ahn et al. proposed a U-Net-based dose prediction model trained on 50 volumetric modulated arc therapy (VMAT) plans for left breast cancer patients [18]. They reported that their model outperformed the RapidPlan dose prediction mode in terms of mean absolute error (MAE). Soomro et al. developed DeepDoseNet, a deep learning model for 3D dose prediction in RT using 340 head-and-neck datasets from the 2020 AAPM OpenKBP challenge. The model was trained with a loss function that combines DVH-based and MAE metrics, and it outperformed models trained with either DVH or MAE alone [26]. Lie et al. utilized a deep learning U-Net model to predict dose distribution based on the structures and corresponding 3D dose distribution of 130 patients with uterine cervical cancer. They adopted the loss function of GAN and reported that the predicted dose differences for OARs fulfilled safety requirements [27]. However, research on the use of deep learning models for RT for GI cancers remains limited, with few studies published to date. Notable exceptions include work on treatment planning for pancreatic SABR [28] and deep learning prediction of proton and photon dose distributions for pediatric abdominal tumors [29]. This scarcity is even more pronounced when considering the use of MR guided systems for dose prediction.

While the current model has shown acceptable performance in terms of predicting 3D dose distributions, the underlying U-Net architecture with its complex multi-layered structure poses challenges in terms of interpretability. This hinders clear understanding of the model's decision-making process, which is crucial for clinicians who rely on Bayesian reasoning to weigh the risks and benefits of different treatment modalities [30]. Indeed, BN have been
applied to predict the response to IMRT for prostate cancer [31], radiation-induced lung disease in lung cancer [32], and to support decision-making process for online CBCT-based IGRT for prostate cancer patients [33]. Thus, while most models focused only on the physical aspects of dose distribution, our current BN model were integrated with patient-specific clinical data such as age, tumor size, and location. This approach might enable more personalized and precise predictions, tailored to not only the dose distributions but also the patient's unique physiological profile. This integration of clinical data into results from the deep learning is a novel approach in RT planning, offering a more comprehensive and patient-centered decision-supporting tool.

Meanwhile, the results of the study found discrepancies between TBX and MRG prediction models. The MRG model predicted better than TBX, which could be due to two reasons. First, the MRIdian® has fixed angles of the three gantries, so the beam profiles of the actual RT plan are relatively constant [34]. Since step-and-shoot inverse IMRT is performed in MRG, the prediction model might fit better than TBX, which is planned with a variety of non-fixed beam angles [35]. Second, dose prescribing of PTV in TBX could be in a way of simultaneous integrated boost (SIB) with more than one dose, so there may have been limitations in dose alignment when preprocessing for 3D U-net learning. Based on this relative difference in prediction performance, we normalized ground truth and predicted doses of PTV and OARs' DVH metrics to a range of 0-1 to quantify these differences. Additionally, when conducting statistical analyses to identify factors influencing the decision of treatment modality (TBX vs. MRG) in BN, we included both MRG and TBX data individually to enhance predictive capability. Despite inherent differences in predictive ability between TBX and MRG data within the deep learning model itself, all datasets were integrated in the statistical analyses to identify four significant factors influencing the decision-making process. Therefore, we
believed this approach would overcome the limitations and allow for robust interpretation of the results.

The proposed model helps to generate more suitable plan only after simulation of CT/MR, instead of both MR- and CT-based manual plan generation. Conducting simulations only could reduce the time required for target and OAR contouring, as well as generating and comparing treatment plans for each image modality. This streamlined process might help to conserve resources by saving the manpower and time of radiation oncologists and dosimetrists, accelerating the decision-making process for treatment approaches. In addition, skipping plan generation once could save approximately 1.6 million won in insurance costs. Furthermore, BN model could leverage patient-specific clinical data including age, tumor size, and tumor location, to refine treatment plan selection. This personalized approach would contribute to optimal treatment decision by tailoring the chosen plan to the unique anatomical and physiological characteristics of individual patient. To apply decision tree derived from the BN into a real-world practice, the interpretation of TAMB method is crucial. In the TAMB, the sequence of factors in the tree structure primarily reflects the importance of each variable in constructing the tree. The topmost node in the tree structure is the root node. This node is positioned at the center of the entire tree and governs the relationships with all other nodes. From the root node, child nodes for each variable are represented in the tree structure. The sequence of these child nodes is usually determined by the importance of the corresponding variable. Variables deemed more important are placed higher in the order, indicating their greater influence in the model. Typically, the score used in TAMB is based on metrics like mutual information between variables. Consequently, a higher score suggests a stronger relationship between the variables [36]. Applying this to the decision tree in Fig. 4, tumor location was the most important factor in determining treatment modality. When the tumor was
abutting hollow viscus and the dose to the stomach was not overly high (maximum normalized dose of the stomach in the MR guided plan $\leq 0.56$), MRG was favored showing higher score (Fig. 4A vs. C). Conversely, if the tumor was not close to hollow viscus and the dose to the stomach was not particularly low (maximum normalized dose of the stomach in the MR guided plan $> 0.56$), TBX was chosen, even though the score of TAMB was much lower. Since MR guided delivery offers higher resolution for soft tissues combined with gated feature to minimize potential overlap [15], the proximity to hollow viscus might explain the preference in favor of MRG in such cases. Additionally, the wider physical penumbra of MRG, as employed model utilized Cobalt source, might be a factor in interpreting the dose to the stomach [37]. If the maximum dose to the stomach is relatively high, MRG could be less suitable for tumors abutting hollow viscus. On the other hand, the scores for PTV size and age were lower than those for tumor location and maximum normalized dose to the stomach. These two factors might also influence treatment decisions, but their relevance was relatively low. Notably, MRG were favored plan for patients with smaller tumors (PTV < 65.7cc) and older patients (> 67 years old). This might be explained by the fact that elderly patients are more likely to have irregular breathing or less adaptation to the shallow breathing required for volumetric access, and are even more susceptible to RT toxicity. Respiratory management approach as well as better soft tissue resolution of MR guidance allows for more accurate tumor contouring, which can help to reduce overdose to the neighboring organs at risk. This decision tree can lead to better treatment outcomes, reduced adverse events, and improved quality of life for patients undergoing RT for upper GI malignancies.

On the other hand, for liver cancer, MRG is generally expected to be favored for its superior resolution compared to CT. Typically, PTV of liver tumors exhibit spatial separation from hollow viscus organs. Contrary to conventional expectations, our study's decision tree
indicated a preference for TBX in cases where the tumor is located in a position categorized as ‘apart’. However, MRG emerged as the favored option under specific circumstances, such as when there is minimal radiation exposure to the stomach and the PTV size is less than 65.7 cc. This observation underscores the advantages of MRG in scenarios where the target volume is small and the impact on stomach radiation dose is minimal. In fact, it is important to understand that lesions in the right hepatic lobe are primarily related to the relationship with the bowel, while lesions in the left hepatic lobe are generally more important for the relationship with the stomach. Thus, the example of liver malignancy illustrates that treatment decisions should be based on a comprehensive consideration of multiple factors, including PTV size and other relevant clinical data, in addition to tumor location.

Although this study focused on upper GI cancers, the proposed framework has the potential to be more broadly applicable in RT planning for various tumor locations. The model developed in this study might be applied to the MR-LINAC system, which have gradually replaced the MR-cobalt system [38,39]. The MR-LINAC system is a newer type of MR-guided RT system that uses a LINAC as radiation source instead of Tri-Cobalt-60 employed in current analysis, resulting in narrower penumbra due to dosimetric difference. In addition, MR-guided system in current analysis used magnetic field strength of 0.35 Tesla, whereas magnetic field strength of 1.5 Tesla is available for system from MR-LINAC, leading to superior soft tissue contrast and spatial resolution. Exploring the model's potential to predict dose distribution for newer MR guided treatment system with LINAC source and/or higher magnetic field strength may expand its clinical utility. With a BN model trained on MR-LINAC images, treatment suggestion can be made as soon as simulation images are obtained. For example, a BN decision tree can be applied by performing a simulated CT/MR of MR-LINAC and then substituting the MRG predictive model values. In terms of efficient operation of the RT equipment, which is a
limited resource, MR-LINAC might be applied in cases where MRG is preferred, and conventional LINAC equipment could be selected in other cases of TBX preferred. Furthermore, the clinical implications of this study can be extended to compare Volumetric Modulated Arc Therapy (VMAT) and IMRT. TBX represents an Internal Target Volume (ITV)-based VMAT plan, whereas MRG represents a gating-based IMRT plan. Therefore, even without MR-guided RT, these concepts can provide valuable insights with extended clinical implications for decision-making in LINAC therapy.

The primary limitation of this study was the restricted patient cohort to develop 3D U-Net deep learning model. It was inevitable that only patients with treatment plans generated for both TBX and MRG were eligible for inclusion. However, to address this limitation, integrating online learning algorithms into the developed model would enable continuous refinement of its predictions based on real-time patient data, ultimately enhancing its accuracy over time.

One more possible limitation of the study related to the definition of tumor location. Depending on the type and location of the PTV, the criteria for defining adjacent hollow viscus organs could vary. Additionally, inter-fractionally or intra-fractionally physiological movements could lead to positional differences, potentially altering distinctions between ‘abutting’ and ‘apart’ classifications. However, to mitigate this uncertainty, the tumor location was classified based on images obtained at the time of diagnosis and during simulations performed for treatment planning purposes. Furthermore, the reference organ was constant, as the hollow viscus within a 5 cm extension of the GTV was subject to the OARs. And there were no cases where the classification of abutting or apart actually changed due to physiologic motion. Meanwhile, discontinuation of the service due to the situation of the manufacturer have potentially outcasted the future use within the system. However, this study yielded valuable insights into the decision-making process of experts in radiation oncology dealing with challenging upper
abdominal cancers. By identifying the significance of tumor location and stomach dose in decision-making process, our findings hold relevance beyond the scope of MR-gated RT application. Therefore, within the framework of expanding expertise, substantial assistance on RT to upper GI maybe provided to practicing radiation oncologists.

We demonstrated a decision-supporting algorithm that enables clinicians to readily choose the optimal plan from two competing RT systems in upper abdominal malignancy. This is anticipated to save resources in terms of manpower, time, and cost.

**Ethical Statement**

This study was approved by the institutional review board of institution. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Because of the retrospective design of the analysis, requirement for obtaining informed consent of participants included in the study was exempted.

**Author Contributions**

Conceived and designed the analysis: Kim DY, Jang BS, Kim E, Chie EK.

Collected the data: Kim DY, Jang BS, Kim E, Chie EK.

Contributed data or analysis tools: Kim DY, Jang BS, Kim E, Chie EK.

Performed the analysis: Kim DY, Jang BS, Kim E.

Wrote the paper: Kim DY, Jang BS, Kim E, Chie EK.
ORCID iDs

Dong-Yun Kim: https://orcid.org/0000-0002-3196-9936
Eui Kyu Chie: https://orcid.org/0000-0003-2027-7472

Conflicts of Interest

Conflict of interest relevant to this article was not reported.

Acknowledgments

This study was supported by National Research Foundation of Korea (NRF) grants funded by the Ministry of Science and ICT, Republic of Korea (NRF-2017M2A2A7A02020639 and NRF-2017M2A2A7A02020640) and SNUH Research Fund (grant no 04-2020-2150) to Eui Kyu Chie, MD, PhD, SNUH Research Fund (Grant no. 04-2022-0260) to Bum-Sup Jang, MD, PhD.
References


Table 1. Baseline characteristics of patients undergoing radiation therapy for upper gastrointestinal cancers with two types of simulation and treatment plans

<table>
<thead>
<tr>
<th>Baseline characteristics</th>
<th>No.</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years (median[range])</td>
<td>64.8</td>
<td>[40-78]</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>44</td>
<td>(67.7)</td>
</tr>
<tr>
<td>Female</td>
<td>21</td>
<td>(32.3)</td>
</tr>
<tr>
<td>Type of cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liver</td>
<td>40</td>
<td>(61.5)</td>
</tr>
<tr>
<td>Pancreas</td>
<td>25</td>
<td>(38.5)</td>
</tr>
<tr>
<td>RT aim</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Radical</td>
<td>14</td>
<td>(21.5)</td>
</tr>
<tr>
<td>Salvage</td>
<td>45</td>
<td>(69.2)</td>
</tr>
<tr>
<td>Postoperative</td>
<td>3</td>
<td>(4.6)</td>
</tr>
<tr>
<td>Palliative</td>
<td>3</td>
<td>(4.6)</td>
</tr>
<tr>
<td>Child-Pugh classification</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>60</td>
<td>(92.3)</td>
</tr>
<tr>
<td>B</td>
<td>5</td>
<td>(7.7)</td>
</tr>
<tr>
<td>ECOG score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>57</td>
<td>(87.7)</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>(12.3)</td>
</tr>
<tr>
<td>Tumor location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abutting</td>
<td>16</td>
<td>(26.6)</td>
</tr>
<tr>
<td>Apart</td>
<td>49</td>
<td>(75.4)</td>
</tr>
<tr>
<td>Treatment type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBX</td>
<td>47</td>
<td>(72.3)</td>
</tr>
<tr>
<td>MRG</td>
<td>18</td>
<td>(27.7)</td>
</tr>
<tr>
<td>RT plan type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SABR</td>
<td>40</td>
<td>(61.5)</td>
</tr>
<tr>
<td>IMRT</td>
<td>25</td>
<td>(38.5)</td>
</tr>
<tr>
<td>SABR dose/fractionation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose (Gy, Mean ± SD)</td>
<td>48.8</td>
<td>± 8.3</td>
</tr>
<tr>
<td>Fractionation (range)</td>
<td>4 - 8</td>
<td></td>
</tr>
<tr>
<td>IMRT dose/fractionation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dose (Gy, Mean ± SD)</td>
<td>50.8</td>
<td>± 7.0</td>
</tr>
<tr>
<td>Fractionation (range)</td>
<td>10 - 30</td>
<td></td>
</tr>
<tr>
<td>PTV volume (cc)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 35</td>
<td>32</td>
<td>(49.2)</td>
</tr>
<tr>
<td>≥ 35</td>
<td>33</td>
<td>(50.8)</td>
</tr>
</tbody>
</table>

ECOG, European Cooperative Oncology Group; TBX, Computed tomography (CT)-based high dose rate stereotactic linear accelerator delivery using TrueBeam-STX™; MRG, MR-guided gated delivery, which utilized MRIdian®; SD, standard deviation; SABR, stereotactic ablative radiotherapy; IMRT, Intensity modulated radiation therapy; PTV, planning target volume.
Table 2. The differences of normalized maximum and mean dose between ground truth and predicted dose of TBX and MRG model

<table>
<thead>
<tr>
<th></th>
<th>Maximum dose* difference</th>
<th>Mean dose* difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TBX</td>
<td>MRG</td>
</tr>
<tr>
<td>PTV</td>
<td>0.18</td>
<td>0.06</td>
</tr>
<tr>
<td>Duodenum</td>
<td>0.14</td>
<td>0.08</td>
</tr>
<tr>
<td>Kidney</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>Liver</td>
<td>0.19</td>
<td>0.08</td>
</tr>
<tr>
<td>Stomach</td>
<td>0.12</td>
<td>0.11</td>
</tr>
</tbody>
</table>

TBX, Computed tomography (CT)-based high dose rate stereotactic linear accelerator delivery using TrueBeam-STX™; MRG, MR-guided gated delivery, which utilized MRIdian®; PTV, planning target volume; *, normalized dose; P-value by T-test.
Table 3. Performance Metrics from 5-fold Cross-Validation from Bayesian Network Model

<table>
<thead>
<tr>
<th>Metrics</th>
<th>TBX</th>
<th>MRG</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROC Index</td>
<td>83.02%</td>
<td>85.25%</td>
</tr>
<tr>
<td>Calibration Index</td>
<td>76.05%</td>
<td>86.39%</td>
</tr>
<tr>
<td>Binary Log-Loss</td>
<td>0.893</td>
<td>0.893</td>
</tr>
<tr>
<td>Overall Precision</td>
<td>83.08%</td>
<td></td>
</tr>
<tr>
<td>Overall Reliability</td>
<td>82.49%</td>
<td></td>
</tr>
<tr>
<td>Overall ROC Index</td>
<td>83.56%</td>
<td></td>
</tr>
</tbody>
</table>

ROC, Receiver Operating Characteristic; MR, Magnetic Resonance; LINAC, linear accelerator.
Fig. 1. The overall network architecture to construct the dose distribution model of CT-based high dose rate stereotactic radiotherapy (TBX) (A), MR-guided radiotherapy (MRG) (B) and the Bayesian network model schema (C).
Fig. 2. The actuarial dose-volume histogram (DVHs) of ground truth and the predicted values by developed 3D U-Net deep learning model of each of CT-based high dose rate stereotactic radiotherapy (TBX) [A] and MR-guided radiotherapy (MRG) [B]. Solid lines depict in DVHs the ground truth values of PTV and OARs, and the dashed lines are the predicted dose by the model.
Fig. 3. Statistical significance for maximum (A) / mean (B) normalized dose difference between ground truth and predicted value for CT-based high dose rate stereotactic radiotherapy (TBX) and MR-guided radiotherapy (MRG) model.
Fig. 4. The decision tree based on Bayesian Network model to decide optimal RT plan modality (CT-based high dose rate stereotactic radiotherapy [TBX] vs. MR-guided radiotherapy [MRG]) based on clinical factors and generated dose-volume histogram profiles by a 3D U-Net deep learning. The score next to the clinical variables in each box indicates the likelihood of the evidence. The higher the score, the more likely the explanation is when ranked.
**S1 Fig.** Definition of tumor location (abutting [A] vs. apart [B]) is determined by geographic relationship between GTV/PTV (red line) and hollow organs (yellow line [A], brown line [B]).
S2 Fig. Bayesian network structure to estimate the modality (CT-based high dose rate stereotactic radiotherapy [TBX] vs. MR-guided radiotherapy [MRG]).