# A Deep Learning Model for V50%, V60%, and V66.7% Prediction in LINAC-based Treatment Planning of Single-Iso-Multiple-Targets (SIMT) Stereotactic Radiosurgery (SRS)

by

Mercedeh Khazaieli

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Chunhao Wang, Supervisor

_____
Yunfeng Cui

_____
Trey Mullikin

_____
Julie Raffi

_____
Fang-Fang Yin

Thesis submitted in partial fulfillment of
the requirements for the degree of Master of Science in the Graduate Program in Medical Physics in the Graduate School
of Duke University

2023

ABSTRACT

A Deep Learning Model for V50%, V60%, and V66.7% Prediction in LINAC-based Treatment Planning of Single-Iso-Multiple-Targets (SIMT) Stereotactic Radiosurgery (SRS)

by

Mercedeh Khazaieli

Graduate Program in Medical Physics
Duke University

Date:_____

Approved:

_____
Chunhao Wang, Supervisor

_____
Yunfeng Cui

_____
Trey Mullikin

_____
Julie Raffi

_____
Fang-Fang Yin

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree
of Master of Science in the Graduate Program in Medical Physics in the Graduate School
of
Duke University

2023

# Abstract

Brain metastases are a common complication of several types of cancer, with about 30-40% of cancer patients developing brain metastases during the course of their treatment. One treatment approach for brain metastases is LINAC-based single-isocenter multiple-target (SIMT) stereotactic radiosurgery (SRS). However, generating high-quality SIMT SRS treatment plans can be time-consuming and can lead to large variations in plan quality, which can affect patient outcomes. Therefore, predicting achievable dosimetric outcomes can be valuable for reducing plan quality variation and improving planning efficiency.

The amount of normal brain tissue receiving 12Gy during a single treatment is a known predictor of brain tissue toxicity, also known as radionecrosis. If the treatment involves 20Gy x 1fx, then the volume of normal brain tissue receiving 60% (V60%) of the treatment dose is equal to the volume receiving 12Gy. Similarly, if the treatment involves 18Gy x 1fx, then the volume of normal brain tissue receiving 66.7% (V66.7%) of the treatment dose is equal to the volume receiving 12Gy. The volume of normal brain tissue receiving 50% (V50%) of the treatment dose is a good measure of the gradient of the dose fall-off from the targets. A prediction of achievable V50% (gradient measure), V60% (assuming 20Gy x 1fx), and V66.7% (assuming 18Gy x 1fx) can assist physicians in the determination of fractionation schemes (i.e., single fx vs. multiple fx). Such

predictions can be used as guidelines for planners to generate a SIMT plan more rapidly with reduced dosimetric variability.

To this end, we developed deep learning networks for the prediction of V50%, V60%, and V66.7% based on each patient's target contours. A key technical innovation of this work is the spherical projection design: by projecting target contours on a spherical surface, the target contours in 3D collapse to a polar-azimuthal angular distribution map. This transformation is view-invariant and enables a dimensional reduction for deep learning input without losing volumetric information.

Our results indicate promising potential but there is a need for further work to improve the accuracy of our predictions.

# Contents

# List of Tables

# List of Figures

# 1. Introduction

## *1.1 What are brain metastases?*

Brain metastases arise from primary systemic tumors and are the most common malignancy encountered in the central nervous system (Achrol et al., 2019; Kotecha et al., 2018; Mehta et al.). Approximately 30-40% of cancer patients develop brain metastases at some point during the course of their treatment (Achrol et al., 2019; Kotecha et al., 2018; Mehta et al.). Brain metastases are a common complication of many types of cancer, including lung, breast, and melanoma (Schouten et al., 2002; Soffietti et al., 2002). Neurological manifestations of metastatic brain tumors vary and can range from focal symptoms, headaches, seizures, speech changes, or change in mental status (Noh & Walbert). These symptoms can have detrimental impacts on a patients' quality of life. Palliative care is essential to addressing patients with brain metastases and alleviating symptoms that can significantly impact their quality of life.

## *1.2 Current Treatment Options*

Current treatment options for brain metastases include surgery, radiation therapy, or a combination of both (Gállego Pérez-Larraya & Hildebrand). Chemotherapy has a limited role due to the presence of the blood-brain barrier and low response rates (Kotecha et al., 2018). Historically, whole brain radiation therapy (WBRT) has been utilized to treat multiple intracranial metastases and surgery is typically used for the treatment of solitary or symptomatic metastases (Palmer et al., 2020). Stereotactic

radiosurgery (SRS) is a type of radiation therapy that delivers high doses of targeted radiation to brain tumors, including brain metastases, while sparing normal brain tissue. Comparative studies have established SRS as the standard of care for post-operative radiation therapy as compared to observation or WBRT (Brown et al., 2017; Mahajan et al., 2017). SRS has become the minimally invasive standard of care for patients with brain metastases as it offers high local efficacy, preservation of cerebral functions and short hospitalization (Hartgerink et al., 2019; Lippitz et al.)

## *1.3 Single Isocenter Multiple Target (SIMT) Stereotactic Radiosurgery (SRS)*

Historically, an isocenter was defined for each target when treating with SRS in order to reduce the uncertainty from rotational setup errors (Slagowski & Wen, 2020). However, this limits the number of beam angles that may be used during treatment and can result in long treatment times (Slagowski & Wen, 2020). To improve treatment efficacy, treating multiple brain targets with SRS using a single isocenter has been investigated by several groups (Chang et al.; Clark et al.; Huang et al.; Slagowski & Wen, 2020). It is common practice to place the isocenter at the midpoint, or geometric center, of all the targets (Chang et al.).

Several studies have demonstrated the feasibility and efficacy of SIMT SRS for the treatment of multiple brain metastases (Clark et al.; Lau et al.; Palmer et al., 2020; Soliman et al., 2016). The SIMT technique delivers radiation to multiple targets

concurrently with shared fields or arcs, making it an efficient option for patients with multiple brain metastases.

The current treatment options for SRS include:

1. Gamma Knife Radiosurgery: This is a type of SRS that uses multiple focused Co-60 sources to deliver high doses of radiation to the brain. The gamma rays are focused onto the target using a series of collimators, resulting in a highly conformal dose distribution (Desai & Rich).
2. Linear Accelerator (LINAC)-Based Radiosurgery: This is another type of SRS that uses a LINAC to generate high energy X-rays that are then focused onto the target. The LINAC-based approach provides more flexibility and versatility compared to Gamma Knife, making it easier to treat complex tumors. LINAC-based radiosurgery utilizes volumetric modulated arc therapy (VMAT) technique which delivers treatment to multiple lesions using a single isocenter through the modulation of dose rate, gantry speed, and collimation (Otto, 2007).
3. CyberKnife Radiosurgery: This is a type of SRS that uses a LINAC mounted on a robotic arm to deliver high doses of radiation to the target. The robotic arm provides precise targeting and can track and adjust the radiation delivery in real-time, making it useful for patients with complex tumors or those who need to be treated while moving (Wowra et al.).

## *1.4 LINAC-based SIMT SRS*

LINAC based stereotactic radiosurgery is a promising treatment option for brain metastases as it allows for accurate patient positioning and a short treatment time (Hartgerink et al., 2019). The flexibility and versatility of LINAC-based SRS also make it easier to treat complex tumors. In addition, smaller clinics may not have access to specialized devices, and as brain metastases are highly prevalent, LINAC-based SRS may be the most accessible treatment option for many patients. LINAC-based SIMT SRS offers a convenient alternative and acceptable rates of radionecrosis for multiple intact or postoperative brain metastases (Palmer et al., 2020).

## *1.5 Limitations*

While LINAC-based SIMT SRS has shown promise in treating multiple brain metastases, it is important to note that there are limitations to this approach. Achieving high plan quality is a crucial goal in radiotherapy treatment planning. However, generating a high-quality SRS treatment plan, particularly for SIMT cases, can be a challenging process that requires extensive planning experience, multiple planning runs, and effective communication among planners, physicians, and other members of the radiation oncology team. Furthermore, the number, size, and spatial distribution of brain metastases can vary significantly between cases, leading to large variations in plan dosimetric quality. Patients with larger or more complex tumors may also be at a higher

risk for treatment-related adverse effects, and determining the optimal radiation dose and treatment regimen for each patient requires further investigation.

The amount of normal brain tissue receiving 12Gy during a single treatment is a known predictor of brain tissue toxicity, also known as radionecrosis (Blonigen et al.; Limon et al., 2017; Minniti et al., 2011; Slagowski & Wen, 2020). If the treatment involves 20Gy x 1fx, then the volume of normal brain tissue receiving 60% (V60%) of the treatment dose is equal to the volume receiving 12Gy. Similarly, if the treatment involves 18Gy x 1fx, then the volume of normal brain tissue receiving 66.7% (V66.7%) of the treatment dose is equal to the volume receiving 12Gy. The volume of normal brain tissue receiving 50% (V50%) of the treatment dose is a good measure of the gradient of the dose fall-off from the targets. These parameters (V50%, V60%, V66.7%) are routinely used to evaluate the dosimetric quality of SIMT SRS treatment plans but can be difficult and time-consuming to achieve.

The decision-making process for determining the fractionation scheme (i.e., 1fx vs 5fx) for SIMT SRS treatment plans can be time-consuming and difficult. Additionally, a quality control tool that addresses inter-planner variability in the quality of treatment plans represents an unmet clinical need (Eaton et al.; Nelms et al.). One potential solution is the development of a tool that can estimate, or predict, achievable dosimetric parameters, which can help reduce plan quality variation and improve planning efficiency.

## *1.6 A Solution: Deep Learning Models*

As previously noted, V50%, V60%, and V66.7% are important dosimetric evaluators in the treatment planning process for SIMT SRS. Studies have shown that SIMT SRS survival outcomes are increased with a lower volume of normal brain receiving greater than 12Gy (Limon et al., 2017). A prediction of achievable V60% and V66.7% (assuming 18Gy x 1fx) can assist physicians in the determination of fractionation schemes (i.e., single fx vs. multiple fx). Such predictions can be used as guidelines for planners to generate a SIMT plan more rapidly with reduced V60%, V66.7%, and V50% variability.

Convolutional neural networks (CNNs) have become popular in recent years for analyzing visual imagery due to their ability to extract and learn important local features and patterns from input images. CNNs are especially suited for image analysis since they reduce images to features that are valuable for accurate predictions. The dataset is typically divided into two subsets: a training set and a test set. The model is trained on the training set by exposing it to a large number of input projections along with their corresponding output labels. During training, the model learns to adjust its internal parameters, such as weights and biases, in order to minimize the difference between its predicted outputs and the ground truth labels. Once training is complete, the model is evaluated on the test set to measure its performance on unseen data. This process helps

to assess the model's ability to generalize its learned patterns and make accurate predictions on new inputs.

Our research proposes a novel approach utilizing deep learning models to predict SIMT SRS dosimetric parameters based on target contours in spherical coordinates. Our objective is to develop a tool that can accurately predict three critical dosimetric evaluators (i.e., V50%, V60%, and V66.7%) based on the distribution of SIMT SRS targets. The prediction of achievable dosimetric parameters can provide reliable guidelines to assist physicians in determining optimal fractionation schemes. Additionally, this network can serve as a quality control tool for inter-planner variability of SIMT SRS cases. Ultimately, this tool has the potential to make SIMT SRS treatment planning more efficient, standardized, and reliable.

# 2. Methods

## *2.1 Dataset*

A total of 149 clinical SIMT plans made by a single planner, collected from January 2019 to April 2022, were retrospectively studied for this project. The data includes 89 single-fraction cases (18/20Gy×1fx), and 60 multi-fraction cases (5/5.5Gy×5fx). All patients were treated with a Volumetric Modulated Arc Therapy (VMAT) technique. Brain stem metastases and primary brain tumors were excluded from the study. The average number of brain metastases across the dataset was 7 (range: 2-53) with an average Planning Target Volume (PTV) volume of 12.30 cc. A distribution of the brain metastases across the dataset is shown in Fig 1.

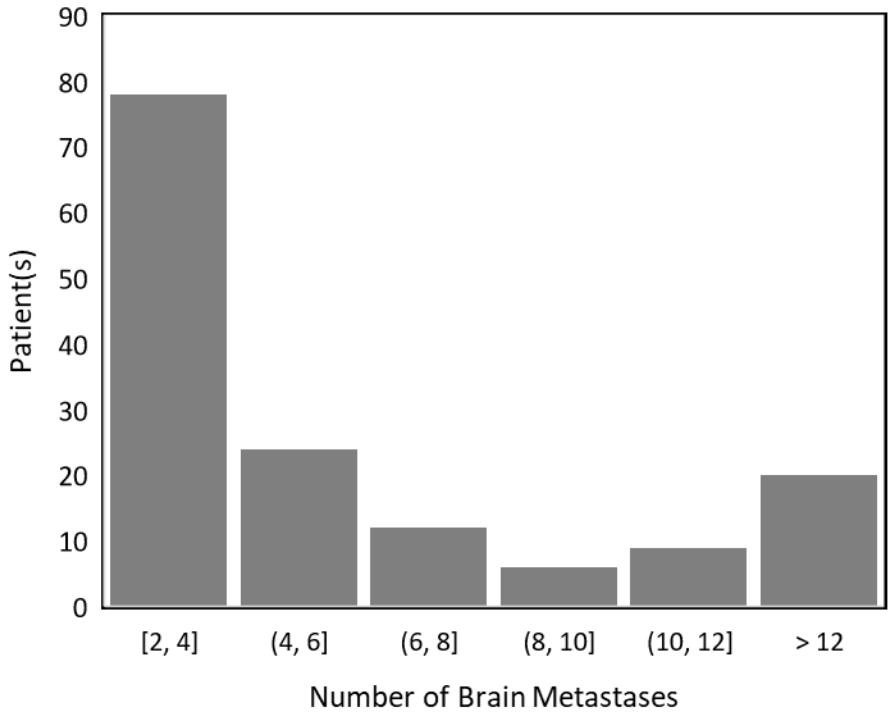


**Figure 1: Distribution of number of brain metastases across the patient dataset**

## *2.2 Data Processing: Spherical Projection*

The novelty in our study derives from the transformation of each patient's PTV contours from a traditional 3-dimensional CT scan to a 2-dimensional spherical projection subtended by two variables (i.e., azimuthal angle and polar angle). We will use patient CT scans with PTV contours to train our deep learning models. Due to the large data dimension, we need a larger neural network to handle the data. However, we are limited by the small sample size which means that we need to reduce the data dimension to prevent overfitting and improve model generalization. The spherical projection reduces the dimensionality of the dataset and produces images that are view-invariant while preserving important volumetric information. In order to achieve this, several steps were taken:

1) A binary mask (0's and 1's) of all the PTVs was created and the isocenter was defined as the geometric center of all the PTV's.
2) For a fixed polar and azimuthal angle, intensity values (either 0 or 1) were added radially every 0.20 cm for r = 10 cm extending from isocenter. The final intensity value (sum of 0's or 1's along radial trajectory) was stored in a 2-dimensional representation corresponding to the polar and azimuthal angle of the trajectory.
3) This process was repeated for 360 polar degrees and 180 azimuthal degrees, at 0.25-degree increments, about the geometric center.

4) For usage in a regression convolutional network, the 2D spherical projections were normalized from 0 to 1.

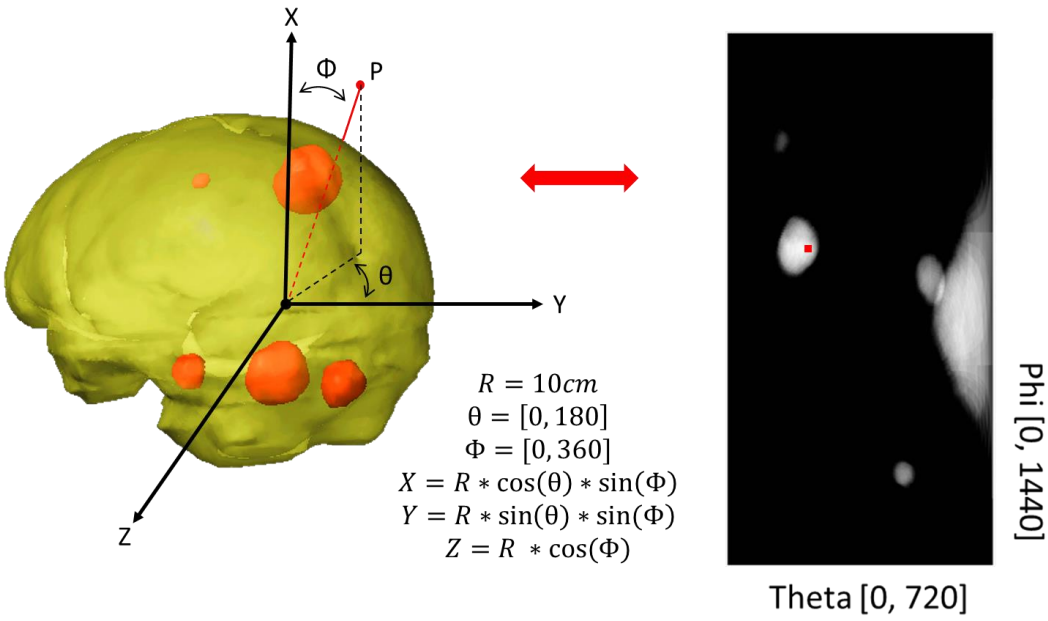


**Figure 2: Illustration of spherical projection algorithm. The representation of point P in cartesian coordinates is transformed to its corresponding spherical coordinate subtended by the polar (phi) and azimuthal (theta) angle.**

The spherical projection algorithm reduced the data size from ~500MB to ~500kB, which corresponds to a reduction in data dimensionality by three orders of magnitude. The target contours were projected onto a R = 10cm spherical surface via ray tracing based on the geometric center of the fitted sphere. The spherical surface was defined to have a radius of 10cm since the distance between the geometric center to each tumor should be below 10cm, per clinical guidelines. Thus, the volumetric information of the PTV distribution was transformed into a 2-dimensional representation in spherical

coordinates, as shown in Fig 2. The 2-dimensional spherical projections are view-invariant and preserve the volumetric information which we are interested in.

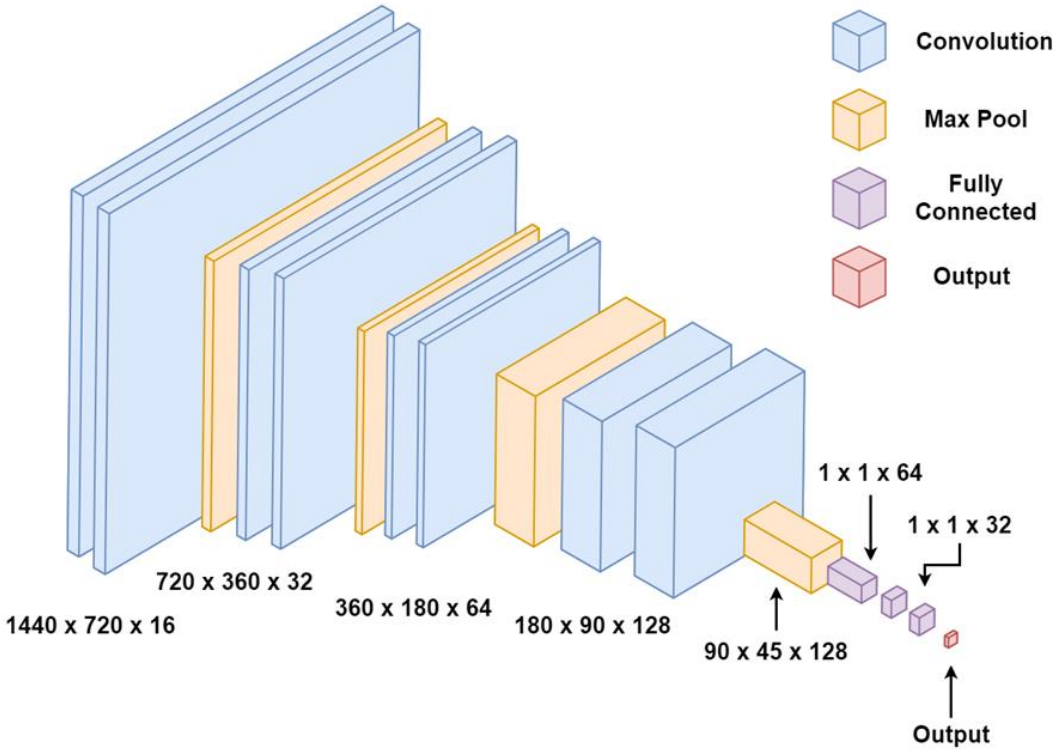


**Figure 3: CNN architecture consisting of 8 convolutional layers, 4 max-pooling layers, and a fully connected network.**

## *2.3 Deep Learning: Convolutional Neural Network (CNN)*

Three convolutional neural networks (Fig 3) were independently trained to predict V50%, V60%, and V66.7% from the spherical projections as a regression problem. Convolutional neural networks were chosen since they are especially suited for analyzing image data, or visual imagery. The goal of CNNs is to reduce images to features that are valuable for accurate prediction. Our CNN architecture (Fig 3) includes

convolution layers, max-pooling layers, and a fully-connected layer. A model summary is shown in Fig 4. The convolutional layers use a set of kernels, or feature detectors, to extract and learn important local features and patterns from the input image. The max-pooling layers down sample the feature maps produced by the convolutional layers, thereby reducing the computational complexity of the network while preserving the important features. Max-pooling layers improve the stability of CNNs by reducing the number of parameters in the network and preventing overfitting. Overfitting occurs when the model learns from the training data too well, to the point that it starts to memorize the noise and errors in the training data, rather than capturing the underlying patterns and relationships. As a result, the overfit model performs poorly on new, unseen data, even though it may have a high accuracy on the training data. The purpose of the fully connected layer is to learn a mapping between the high-level features extracted by the convolutional layers and the output labels. The fully connected layer is responsible for producing the final output predictions based on the features extracted by the convolutional layers.

Since this is a regression problem, we utilize a 'relu' activation function. In a regression task, a common loss function is mean squared error (MSE), which measures the average squared difference between the predicted output and the true output values. The training process is achieved by minimizing the loss function and maximizing the coefficient of determination ($R^2$). The coefficient of determination, $R^2$, is an estimate of

how well the model can make predictions. Any $R^2$ between 0 and 1 indicates what percentage of the target variable can be explained by the input features. An $R^2$ of less than 0 indicates that the model is no better than one that constantly predicts the mean of the target variable.

The models were trained using a batch size of 10 and an epoch of 50. Batch size and epoch are two important hyperparameters that are related to the training process of a model. The batch size refers to the number of images that are processed in parallel through the network during one iteration of the training process. By processing the data in small batches at a time, we can reduce the memory requirements of the training process and use less memory during training. An epoch, on the other hand, refers to the number of times the model has iterated over the entire training dataset during the training process. Each epoch allows the model to see the entire training dataset, and by training over multiple epochs, the model can gradually improve its accuracy by adjusting its parameters based on the errors it makes on the training data. This process helps the model learn more complex patterns in the data and improve its accuracy over time.

The proposed CNNs were trained on augmented training data. The training process followed a 8:2 train/test split ratio, and a 10-fold cross validation was also assessed.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 1440, 720, 16)     160
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 1440, 720, 16)     2320
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 720, 360, 16)      0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 720, 360, 32)      4640
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 720, 360, 32)      9248
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 360, 180, 32)      0
_________________________________________________________________
conv2d_4 (Conv2D)            (None, 360, 180, 64)      18496
_________________________________________________________________
conv2d_5 (Conv2D)            (None, 360, 180, 64)      36928
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 180, 90, 64)       0
_________________________________________________________________
conv2d_6 (Conv2D)            (None, 180, 90, 128)      73856
_________________________________________________________________
conv2d_7 (Conv2D)            (None, 180, 90, 128)      147584
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 90, 45, 128)       0
_________________________________________________________________
global_average_pooling2d (Gl (None, 128)               0
_________________________________________________________________
dense (Dense)                (None, 64)                8256
_________________________________________________________________
dense_1 (Dense)              (None, 32)                2080
_________________________________________________________________
dense_2 (Dense)              (None, 32)                1056
_________________________________________________________________
dense_3 (Dense)              (None, 1)                 33
=================================================================
Total params: 304,657
Trainable params: 304,657
Non-trainable params: 0
_________________________________________________________________
```

**Figure 4: CNN Model Summary**

## *2.4 Data Augmentation*

Data augmentation is a technique used in convolutional neural networks (CNNs) to artificially increase the size of the training dataset by applying various transformations to the input images. Because we are limited by the small sample size (number of cases) we utilized standard data augmentation techniques to reduce the chance of overfitting, increase the diversity of the training data and improve the

generalization performance of the model. We applied a 9-fold augmentation to the 2D spherical projections which included image flips, rotations, and various circular shifts. An illustration of the image augmentation on one patient is illustrated in Fig 5.

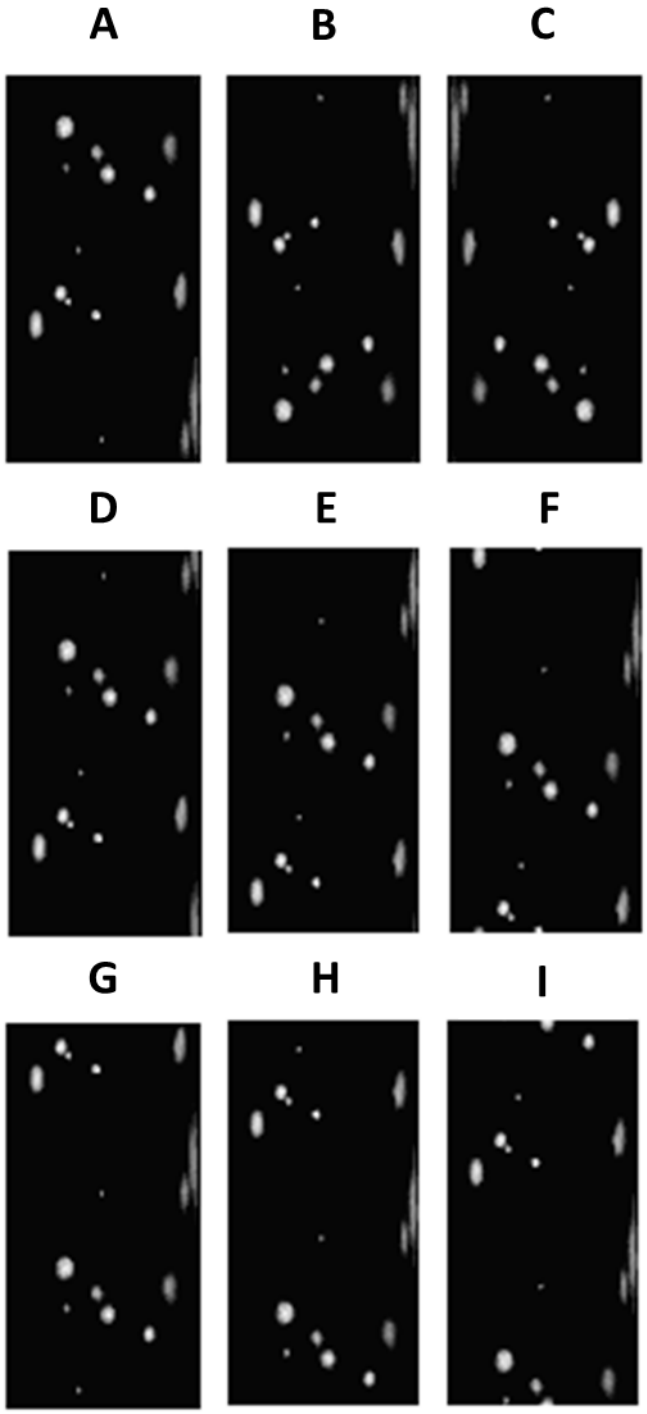


**Figure 5: Data Augmentation example: A) original image; B) horizontal flip, C) 180 degree rotation, D) 45 degree circular shift, E) 90 degree circular shift, F) 135 degree circular shift, G) 180 degree circular shift, H) 225 degree circular shift, I) 270 degree circular shift.**

## *2.5 Single test/train split*

Three CNNs were independently trained on the augmented training dataset in a single train/test split ratio of 8:2. The 2D spherical projections were used as the input to the CNN with the target variable being their corresponding dosimetric parameter. The distribution of the target variables in Fig. 6 ensures that the test dataset does not include out-of-distribution data, or test data that has a different distribution than the training set. Training process was evaluated through the loss and $R^2$ coefficient . Model performance was quantitatively evaluated through the median absolute difference in volume (cc) and median percent difference (%) between the predictions and ground truth values in the 30 test cases. The correlation between predicted and ground truth values (cc) in the 30 test cases was evaluated for each parameter. Additionally, predictions in single fractionated cases were compared to multi-fractionated cases

through box-and-whisker plots.

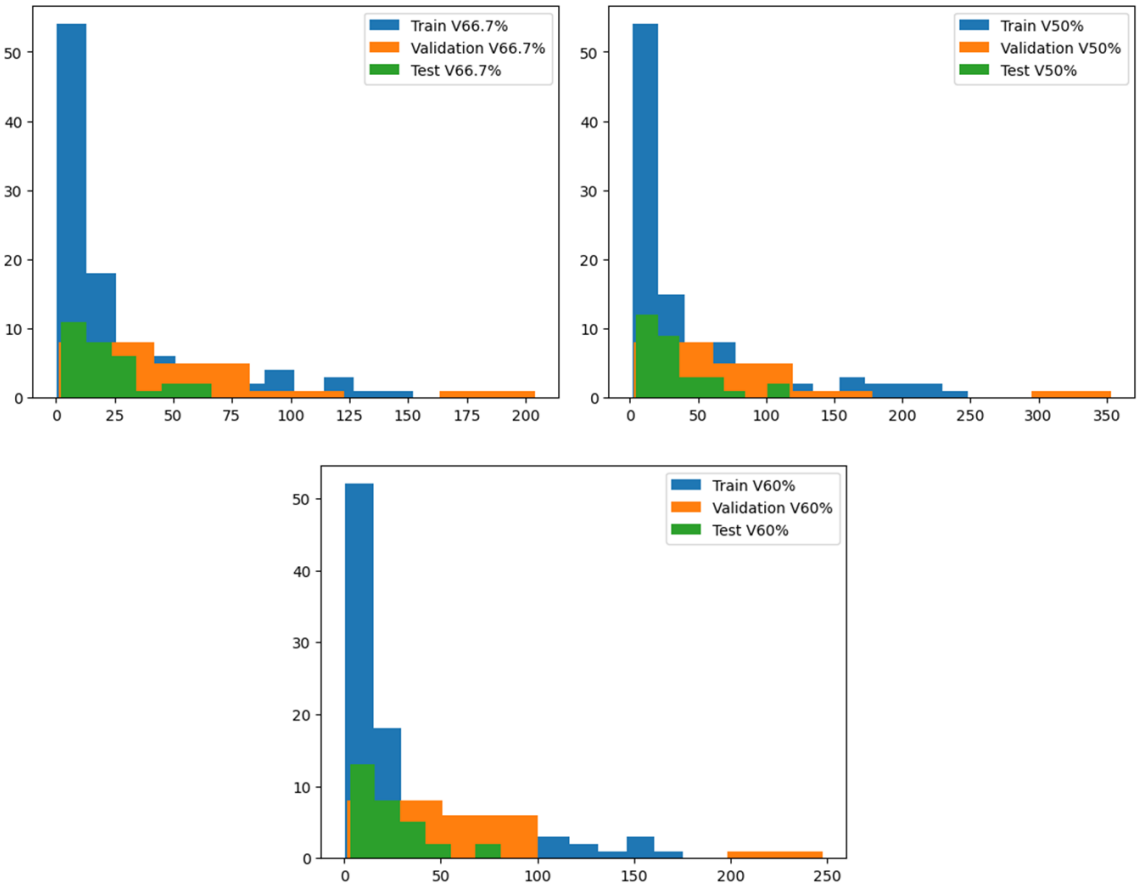


**Figure 6: Distribution of target variables in train/validation/test split**

## *2.6 10-fold Cross Validation*

Finally, we ran a 10-fold cross validation to estimate the generalizability of our model and try to arrive at an optimal solution. The 10-fold cross validation process is illustrated in Fig 7. In 10-fold cross-validation, the original dataset is randomly partitioned into 10 equal-sized subsets, or folds. The model is then trained on 9 of the folds and evaluated on the remaining fold, called the validation set. This process is

repeated 10 times, with each of the folds used as the validation set once. The performance of the model is then averaged across the 10 validation sets to obtain an estimate of its generalization performance. Given the limited dataset, a 10-fold cross validation can provide an estimate of the model's generalization performance on new data, which can help to ensure that the model is robust and reliable.

For the three independent CNN's undergoing the 10-fold cross validation, the mean loss and mean $R^2$ metrics were collected. The resulting models were evaluated through the median absolute difference in volume (cc) and median percent difference (%) between the predictions and ground truth values in the 30 test cases. The correlation between predicted and ground truth values (cc) in the 30 test cases was evaluated for each parameter. Additionally, predictions in single fractionated cases were compared to multi-fractionated cases through box-and-whisker plots.

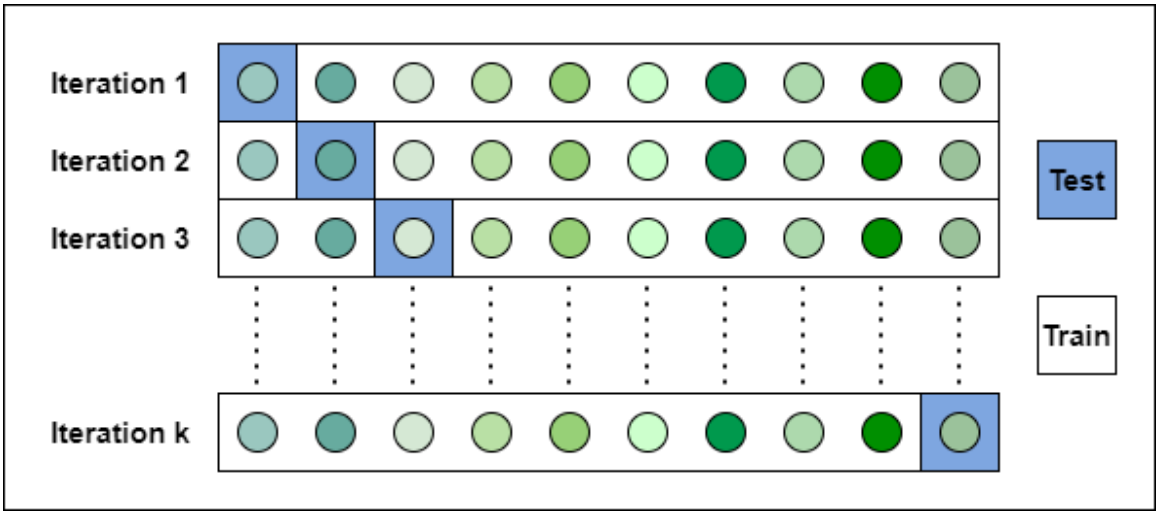


**Figure 7: Illustration of 10-fold cross validation (k=10)**

# 3. Results

A total of 6 models were trained on augmented training data: 3 using a single train/test split & 3 using a 10-fold cross validation. Model predictions on the 30 test cases were compared to the ground truth values based on the median absolute difference (cc) and median percent difference (%). The correlation between predicted and ground truth values (cc) in the 30 test cases was evaluated for each parameter within each model. Finally, prediction in single fractionated cases was compared to multi-fractionated cases for each model. Quantitative results for the 6 models are shown in Tables 1 and 2.

**Table 1: median absolute difference (cc) between predictions and ground truth in 30 test cases**

| Parameter | Single train/test split | 10-fold cross validation |
|---|---|---|
| V50% | 10.05 | 7.97 |
| V60% | 10.15 | 5.52 |
| V66.7% | 7.11 | 4.81 |

**Table 2: median percent difference (%) between predictions and ground truth in 30 test cases**

| Parameter | Single train/test split | 10-fold cross validation |
|---|---|---|
| V50% | 47.01 | 31.11 |
| V60% | 36.26 | 31.11 |
| V66.7% | 35.06 | 41.35 |

### *3.1 Standard single train/test split*

Training loss and $R^2$ coefficients during the training process for V50%, V60%, and V66.7% models are shown in Fig 8, 9, and 10, respectively.

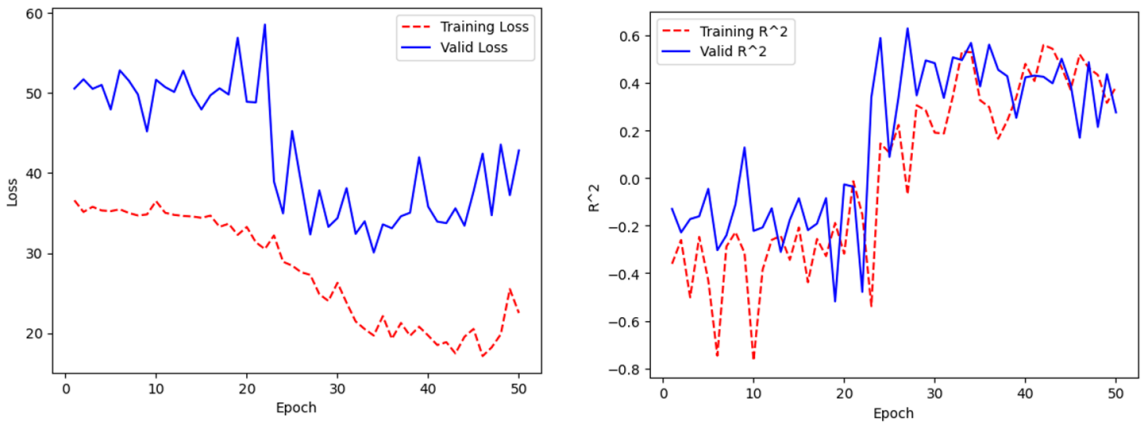


**Figure 8: Training loss and $R^2$ for V50% using standard single train/test split**

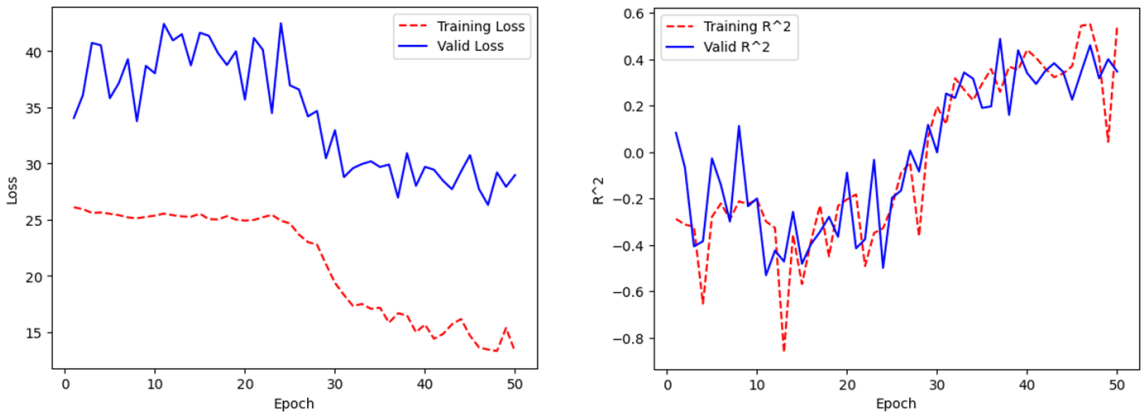


**Figure 9: Training loss and $R^2$ for V60% using standard single train/test split**

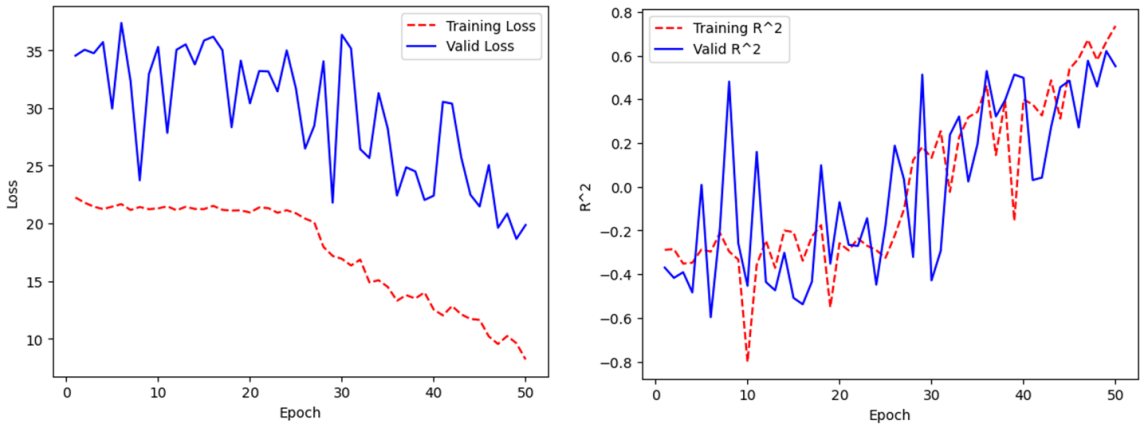


**Figure 10: Training loss and $R^2$ for V66.7% using standard single train/test split**

A box-and-whisker plot comparison between single and multi-fraction case predictions for V50%, V60%, and V66.7% models is shown in Fig 11, 12, and 13, respectively.

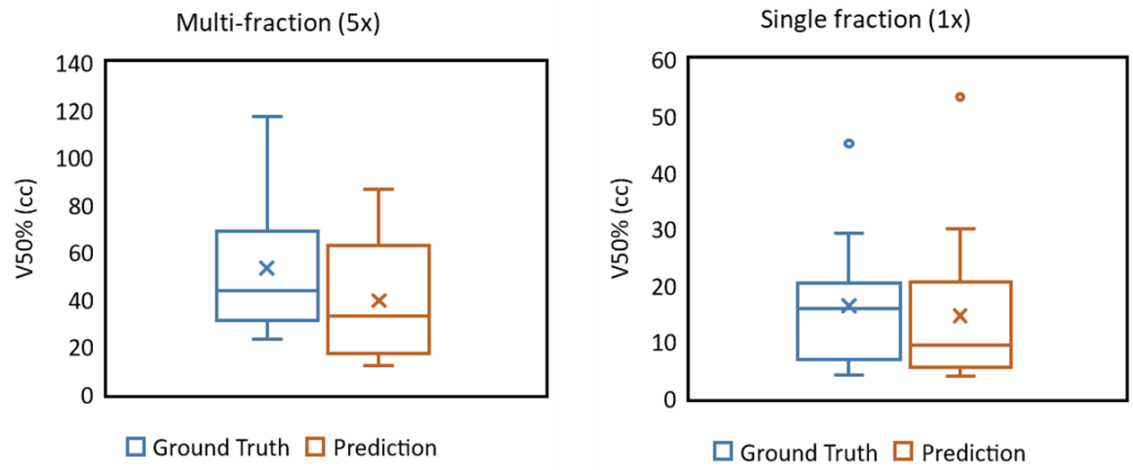


**Figure 11: single fraction versus multiple fraction predictions for V50% using standard train/test split model**

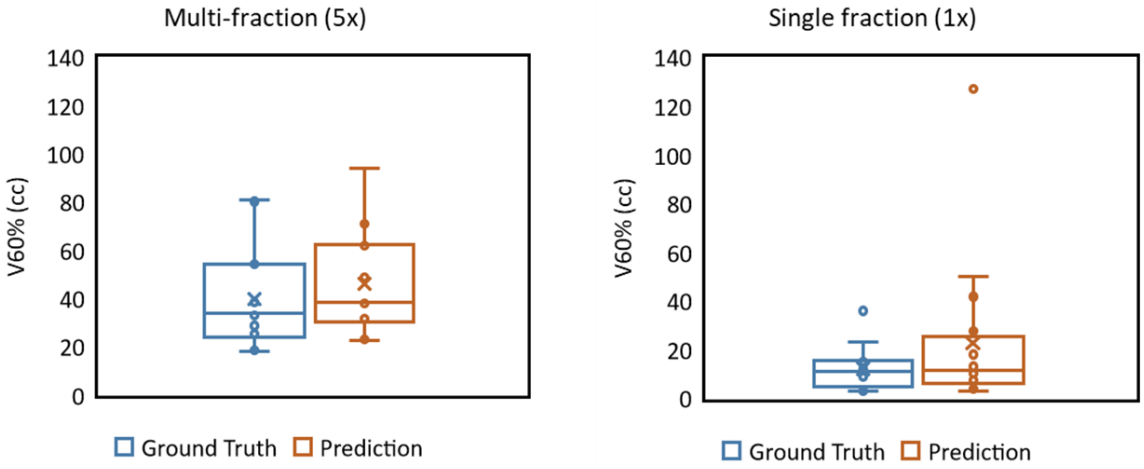


**Figure 12: single fraction versus multiple fraction predictions for V60% using standard train/test split model**

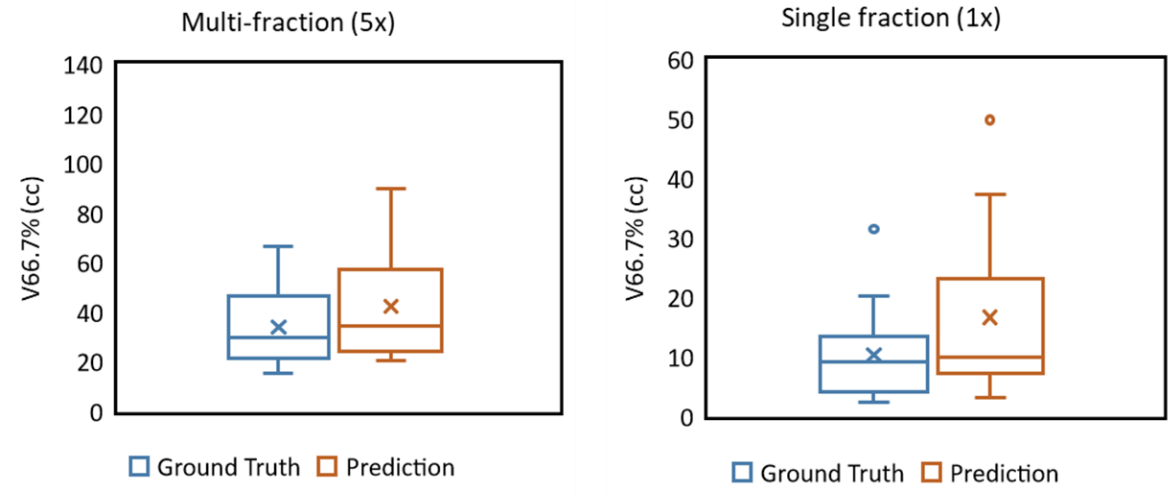


**Figure 13: single fraction versus multiple fraction predictions V66.7% using standard train/test split model**

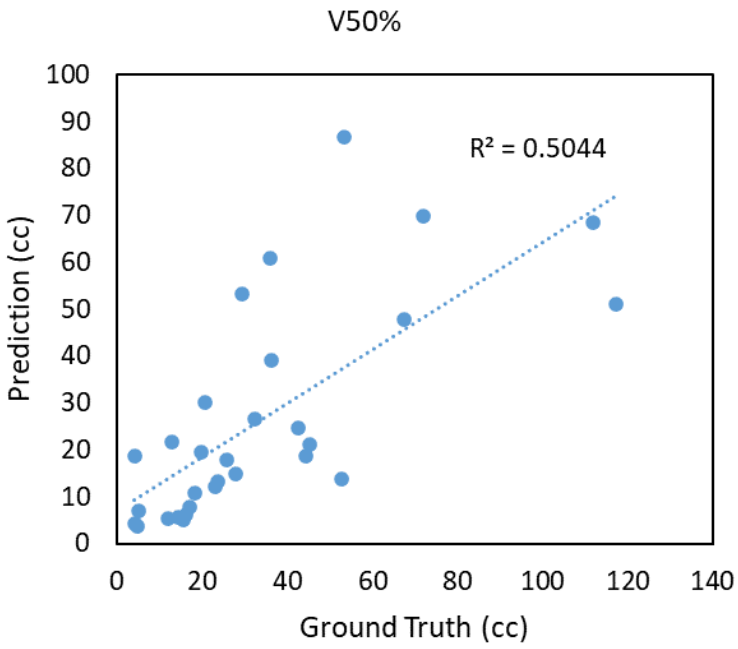


**Figure 14: Correlation of single split model predictions in 30 test cases to ground truth V50% (cc).**

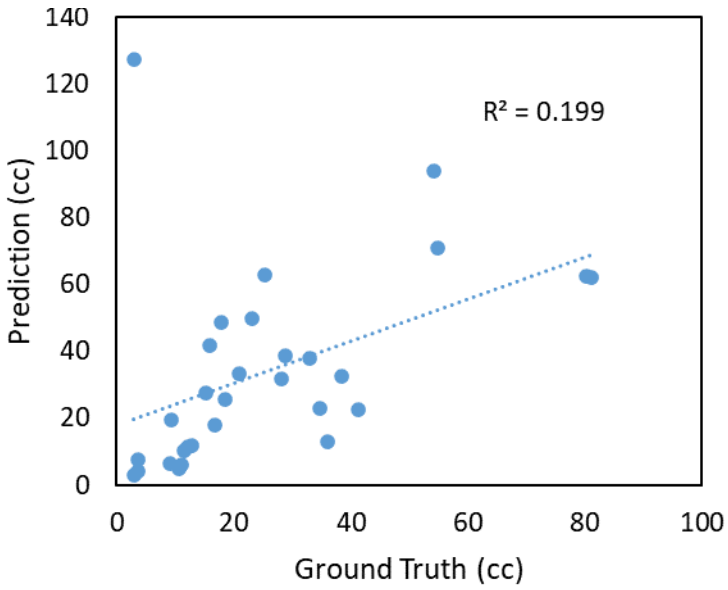


**Figure 15: Correlation of single split model predictions in 30 test cases to ground truth V60% (cc).**

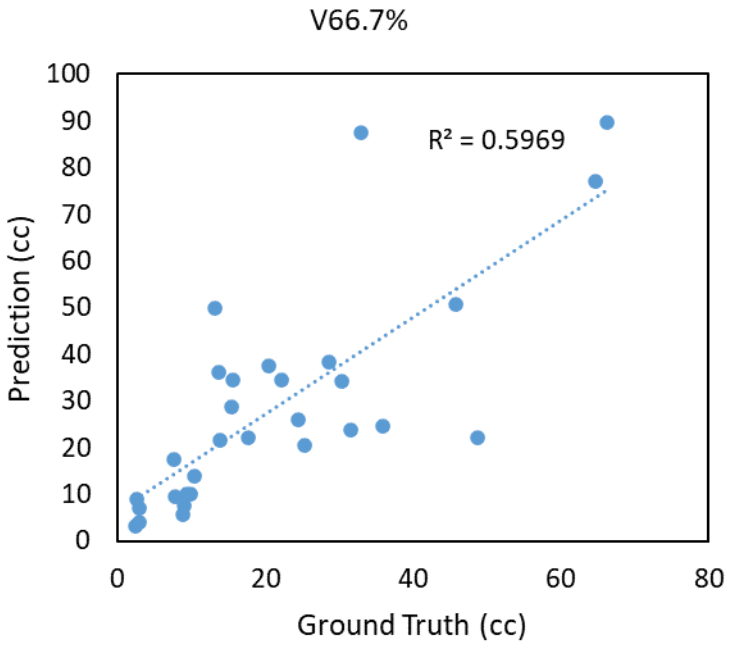


**Figure 16: Correlation of single split model predictions in 30 test cases to ground truth V66.7% (cc).**

## *3.2 10-fold cross validation*

Training loss and $R^2$ coefficients during the 10-fold cross validation for V50%, V60%, and V66.7% models is shown in Fig 17, 18, and 19, respectively.

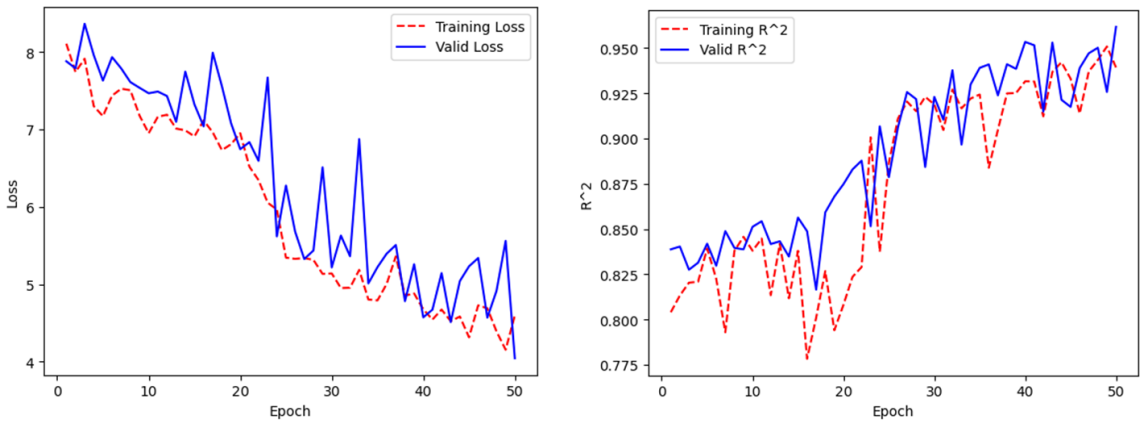


**Figure 17: 10-fold cross validation training loss and $R^2$ for V50% using 10-fold cross validation**

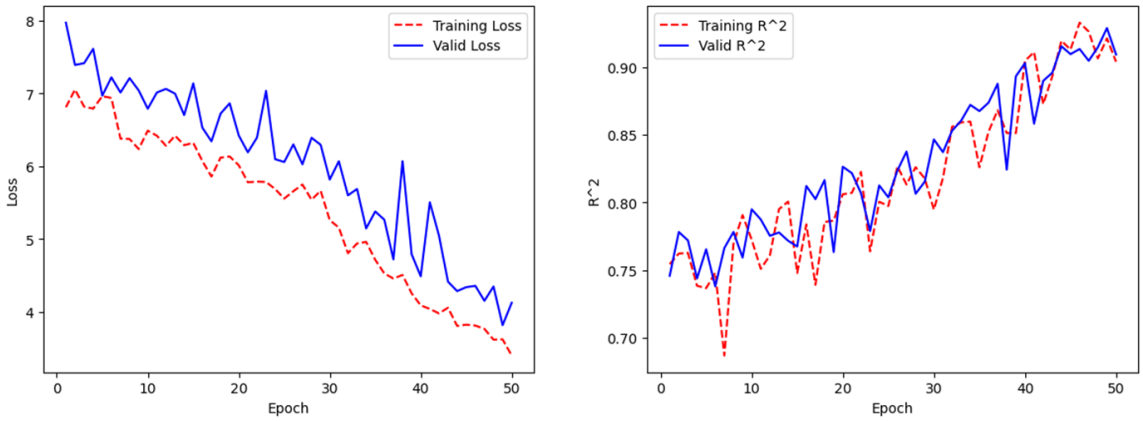


**Figure 18: 10-fold cross validation training loss and $R^2$ for V60% using 10-fold cross validation**

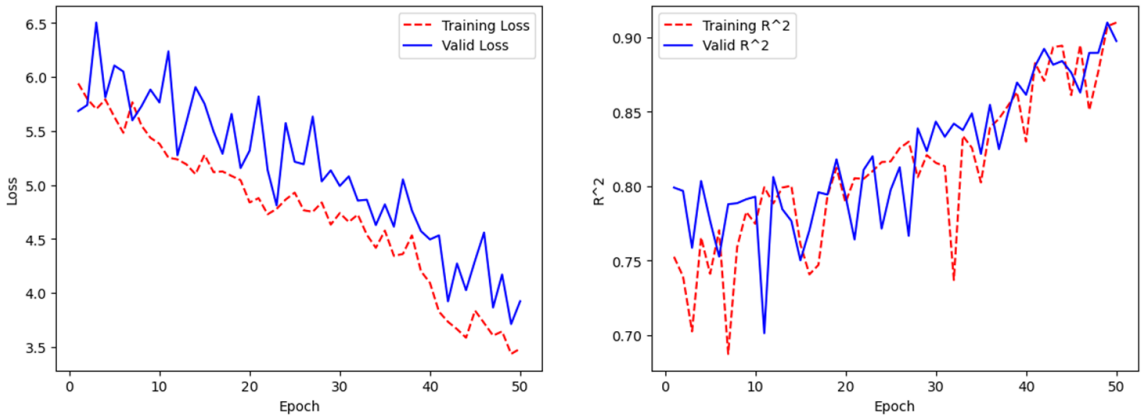


**Figure 19: 10-fold cross validation training loss for $R^2$ for V66.7% using 10-fold cross validation**

A box-and-whisker plot comparison between single and multi-fraction case predictions for V50%, V60%, and V66.7% models based on the 10-fold cross validation is shown in Fig. 20, 21, and 22, respectively.

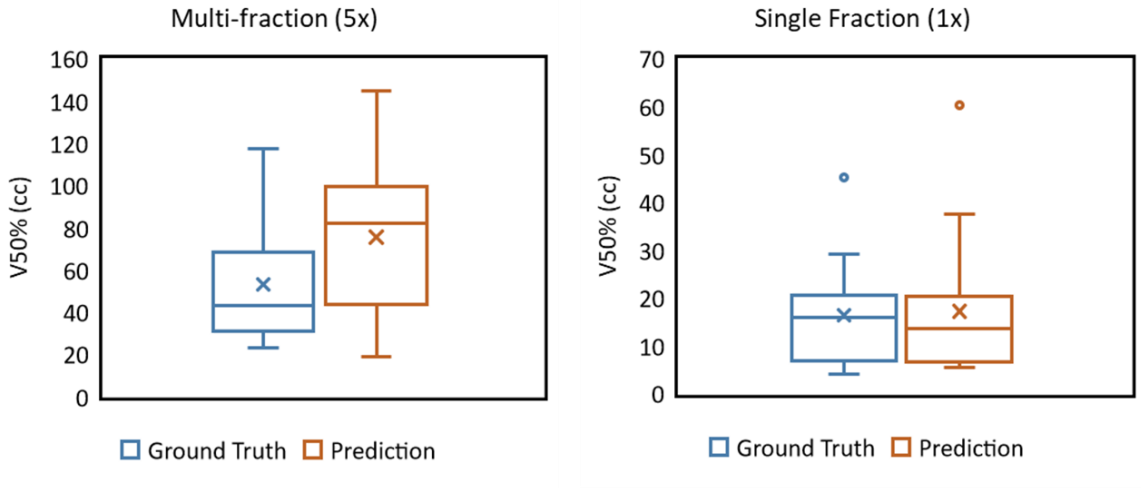


**Figure 20: single fraction vs multiple fraction predictions for V50% using 10-fold cross validation**

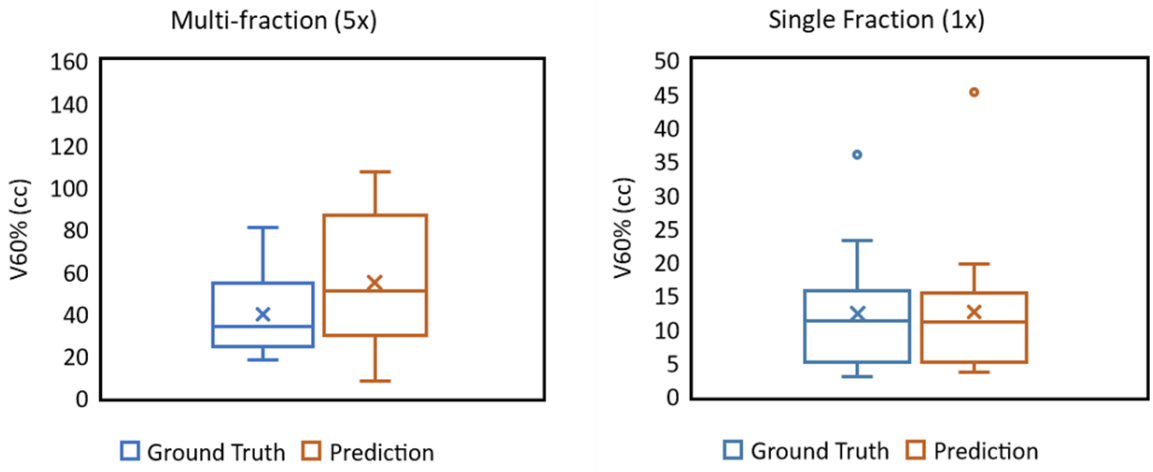


**Figure 21: single fraction vs multiple fraction predictions for V60% using 10-fold cross validation**

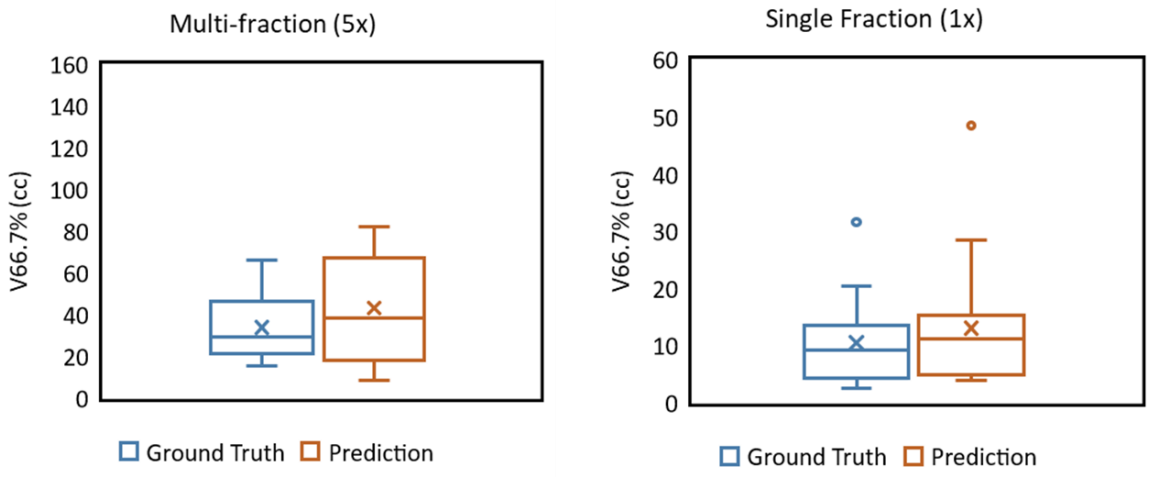


**Figure 22: single fraction vs multiple fraction predictions for V66.7% using 10-fold cross validation**

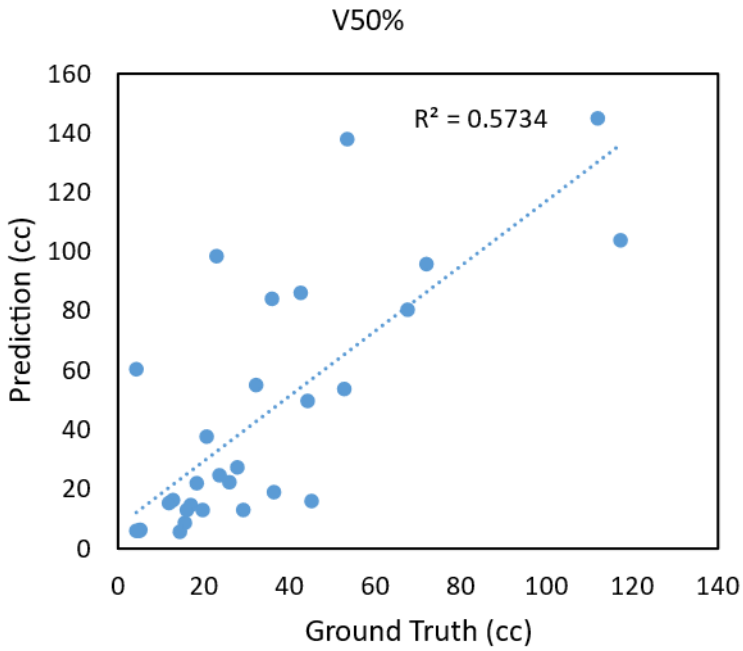


**Figure 23: Correlation of 10-fold cross validation model predictions in 30 test cases to ground truth V50% (cc).**

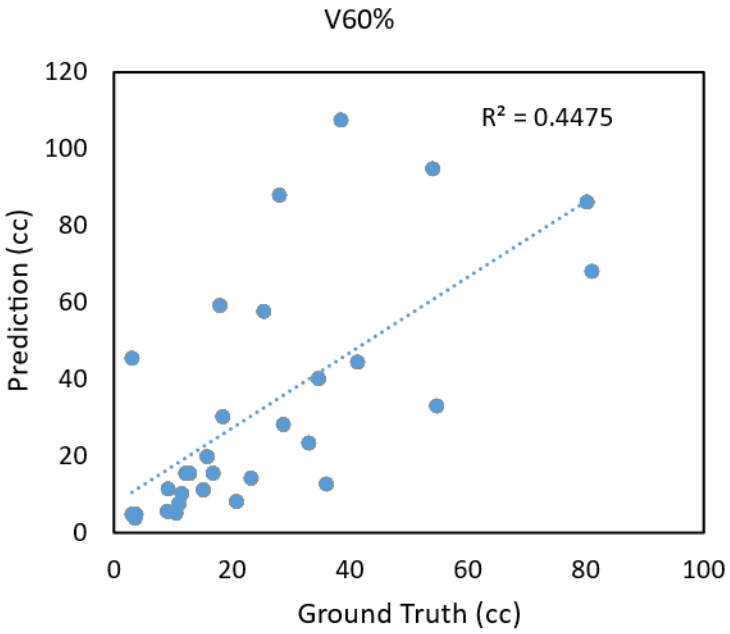


**Figure 24: Correlation of 10-fold cross validation model predictions in 30 test cases to ground truth V60% (cc).**

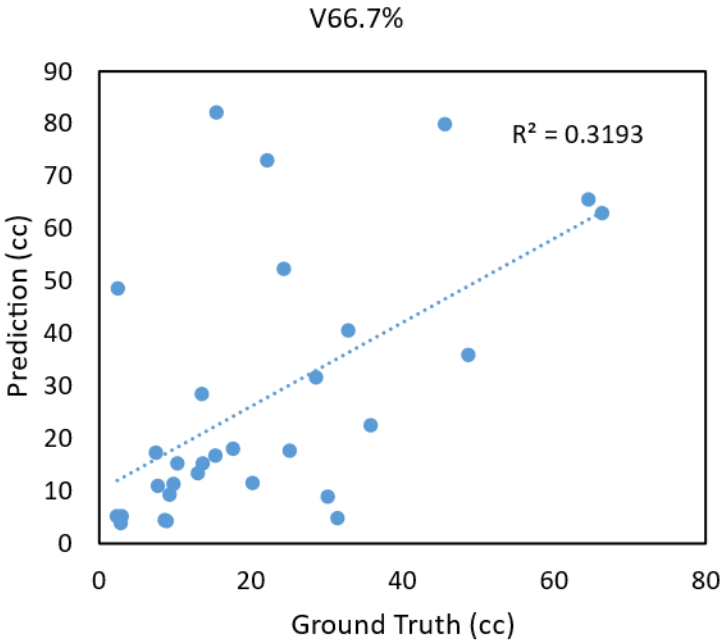


**Figure 25: Correlation of 10-fold cross validation model predictions in 30 test cases to ground truth V66.7% (cc).**

# 4. Discussion

For all three single train/test split model, we observe a decrease in train and validation loss as well as an increase in $R^2$ during each training epoch. When the train and validation loss decrease while $R^2$ increases during each training epoch, it indicates that the model is learning and improving its performance. A decrease in loss signifies that the model is becoming better at predicting the output based on the input. A high $R^2$ value indicates that the model is fitting the data well and explaining most of the variance in the output. Therefore, the combination of decreasing loss and increasing $R^2$ during each training epoch is an indication of the model's good performance. It suggests that the model is learning the underlying patterns in the data and making accurate predictions on both the training and validation sets. However, it is important to note that a high $R^2$ value does not necessarily mean that the model is the best fit for the data, and other factors such as overfitting and bias must also be considered.

To control for overfitting and bias, we ran the 10-fold cross validation technique. A decrease in the average train and validation loss, and an increase in the average $R^2$ over the folds, indicates that the deep learning model is performing well and is generalizing well to new data. The decrease in the average train and validation loss of our models suggests that the models are able to minimize the error between the predicted and actual output during training and testing. This implies that the model is

capturing the patterns and features in the data that are relevant to the output. From the quantitative results in Tables 1 and 2, we see an improvement in prediction accuracy following a 10-fold cross validation.

The median absolute difference (cc) was lowest for V66.7% and highest for V50%. This can be explained by the generally smaller values of V66.7% compared to the other parameter values. It is interesting to note that the median percent difference (%) was highest for V66.7%. Generally, the median percent difference (%) was high for all three parameters. A large median percent difference with a small median absolute difference may indicate that the data has some outliers or extreme values that are affecting the overall results, but the majority of the values are relatively consistent. A small median absolute difference (cc) indicates that most of the paired values are relatively close to each other, with only a few extreme values that are driving the large percent difference. We decided to use the median as a metric since it is a robust measure of error that is less affected by outliers than the mean. Median absolute error is useful when there are extreme values in the dataset that may skew the results and make it difficult to interpret the average error, such as in our results.

The box-and-whisker plots indicate that the models are better at predicting single fractionated cases than multi-fractionated cases. In fact, we see a closer prediction for single fractionated cases after 10-fold cross validation. Single fractionated have fewer targets, smaller targets, and less variability in the target distribution. Clinically, it makes

sense that the models would be better at predicting plan quality evaluators in single fractionated cases.

From the correlation plots, we see an improved correlation between predictions and ground truth values following the 10-fold cross validation for V50% and V60%. The correlation was worse for V66.7% following a 10-fold cross validation. The correlation plot is a graphical representation of how closely the predicted values of a model match the actual values. In this case, an improved correlation means that the predictions were more accurate and closely matched the ground truth values.

This regression problem was especially difficult to achieve due to the limited number of patients and the model complexity. Future work is directed at improving prediction accuracy before potential clinical application. Although deep learning models have great potential for clinical applications, their lack of explainability poses a challenge. Currently, these models remain a 'black box' in clinical settings, without proper transparency. To ensure that deep learning algorithms are driven by clinically appropriate features and make clinically defensible decisions, recent studies are directed at determining the internal mechanisms of these networks (Chaddad et al., 2023). Through explainable AI, we can increase accountability and confidence in the use of deep learning models for clinical applications.

# 5. Conclusion

A high plan quality is able to treat brain metastases while minimizing radiation's impact on the quality of life and neurocognition. Creating a high-quality treatment plan for stereotactic radiosurgery (SRS), particularly for a single isocenter multiple target (SIMT) plan, typically demands the planners' extensive experience in planning, multiple iterations of trial-and-error, and frequent communication among radiation oncology team members and physicians. In clinical settings, where resources may be limited, SIMT SRS planning could be time-consuming and exhibit significant variations in plan dosimetric quality. To improve planning efficiency and reduce plan quality variations, an estimation of achievable dosimetric outcomes could be useful.

The goal of this project was to utilize deep learning networks, specifically CNNs, to predict V50%, V60%, and V66.7% dosimetric evaluators based on each patient's unique target contours. A key technical innovation of this work is the spherical projection design: by projecting the target contours on a spherical surface, the target distribution in 3D is collapsed to a polar-azimuthal angular distribution map that is view-invariant, reduces data size, and preserves volumetric information. This design enables a dimensional reduction for deep learning input without losing the key volumetric information. While our results indicate promising potential, there is a need for further work to improve the accuracy of our predictions.

# References


Achrol, A. S., Rennert, R. C., Anders, C., Soffietti, R., Ahluwalia, M. S., Nayak, L., Peters, S., Arvold, N. D., Harsh, G. R., Steeg, P. S., & Chang, S. D. (2019). Brain metastases. *Nature Reviews Disease Primers*, *5*(1). https://doi.org/10.1038/s41572-018-0055-y

Blonigen, B. J., Steinmetz Rd Fau - Levin, L., Levin L Fau - Lamba, M. A., Lamba Ma Fau - Warnick, R. E., Warnick Re Fau - Breneman, J. C., & Breneman, J. C. Irradiated volume as a predictor of brain radionecrosis after linear accelerator stereotactic radiosurgery. (1879-355X (Electronic)).

Brown, P. D., Ballman, K. V., Cerhan, J. H., Anderson, S. K., Carrero, X. W., Whitton, A. C., Greenspoon, J., Parney, I. F., Laack, N. N. I., Ashman, J. B., Bahary, J.-P., Hadjipanayis, C. G., Urbanic, J. J., Barker, F. G., Farace, E., Khuntia, D., Giannini, C., Buckner, J. C., Galanis, E., & Roberge, D. (2017). Postoperative stereotactic radiosurgery compared with whole brain radiotherapy for resected metastatic brain disease (NCCTG N107C/CEC·3): a multicentre, randomised, controlled, phase 3 trial. *The Lancet Oncology*, *18*(8), 1049-1060. https://doi.org/10.1016/s1470-2045(17)30441-2

Chaddad, A., Peng, J., Xu, J., & Bouridane, A. (2023). Survey of Explainable AI Techniques in Healthcare. *Sensors*, *23*(2), 634. https://doi.org/10.3390/s23020634

Chang, J., Wernicke, A. G., & Pannullo, S. C. Restricted single isocenter for multiple targets dynamic conformal arc (RSIMT DCA) technique for brain stereotactic radiosurgery (SRS) planning. (2156-4647 (Electronic)).

Clark, G. M., Popple Ra Fau - Young, P. E., Young Pe Fau - Fiveash, J. B., & Fiveash, J. B. Feasibility of single-isocenter volumetric modulated arc radiosurgery for treatment of multiple brain metastases. (1879-355X (Electronic)).

Desai, R., & Rich, K. M. Therapeutic Role of Gamma Knife Stereotactic Radiosurgery in Neuro-Oncology. (0026-6620 (Print)).

Eaton, D. J., Lee, J., & Paddick, I. Stereotactic radiosurgery for multiple brain metastases: Results of multicenter benchmark planning studies. (1879-8519 (Electronic)).

Gállego Pérez-Larraya, J., & Hildebrand, J. Brain metastases. (0072-9752 (Print)).

Hartgerink, D., Swinnen, A., Roberge, D., Nichol, A., Zygmanski, P., Yin, F.-F., Deblois, F., Hurkmans, C., Ong, C. L., Bruynzeel, A., Aizer, A., Fiveash, J., Kirckpatrick, J., Guckenberger, M., Andratschke, N., De Ruysscher, D., Popple, R., & Zindler, J. (2019). LINAC based stereotactic radiosurgery for multiple brain metastases: guidance for clinical implementation. *Acta Oncologica*, *58*(9), 1275-1282. https://doi.org/10.1080/0284186x.2019.1633016

Huang, Y., Chin, K., Robbins, J. R., Kim, J., Li, H., Amro, H., Chetty, I. J., Gordon, J., & Ryu, S. Radiosurgery of multiple brain metastases with single-isocenter dynamic conformal arcs (SIDCA). (1879-0887 (Electronic)).

Kotecha, R., Gondi, V., Ahluwalia, M. S., Brastianos, P. K., & Mehta, M. P. (2018). Recent advances in managing brain metastasis. *F1000Research*, *7*, 1772. https://doi.org/10.12688/f1000research.15903.1

Lau, S. K., Zhao, X., Carmona, R., Knipprath, E., Simpson, D. R., Nath, S. K., Kim, G. Y., Hattangadi, J. A., Chen, C. C., & Murphy, K. T. Frameless single-isocenter intensity modulated stereotactic radiosurgery for simultaneous treatment of multiple intracranial metastases. (2218-676X (Print)).

Limon, D., McSherry, F., Herndon, J., Sampson, J., Fecci, P., Adamson, J., Wang, Z., Yin, F.-F., Floyd, S., Kirkpatrick, J., & Kim, G. J. (2017). Single fraction stereotactic radiosurgery for multiple brain metastases. *Advances in Radiation Oncology*, *2*(4), 555-563. https://doi.org/10.1016/j.adro.2017.09.002

Lippitz, B., Lindquist C Fau - Paddick, I., Paddick I Fau - Peterson, D., Peterson D Fau - O'Neill, K., O'Neill K Fau - Beaney, R., & Beaney, R. Stereotactic radiosurgery in the treatment of brain metastases: the current evidence. (1532-1967 (Electronic)).

Mahajan, A., Ahmed, S., McAleer, M. F., Weinberg, J. S., Li, J., Brown, P., Settle, S., Prabhu, S. S., Lang, F. F., Levine, N., McGovern, S., Sulman, E., McCutcheon, I. E., Azeem, S., Cahill, D., Tatsui, C., Heimberger, A. B., Ferguson, S., Ghia, A., Demonte, F., Raza, S., Guha-Thakurta, N., Yang, J., Sawaya, R., Hess, K. R., & Rao, G. (2017). Post-operative stereotactic radiosurgery versus observation for completely resected brain metastases: a single-centre, randomised, controlled, phase 3 trial. *The Lancet Oncology*, *18*(8), 1040-1048. https://doi.org/10.1016/s1470-2045(17)30414-x

Mehta, M. P., Tsao Mn Fau - Whelan, T. J., Whelan Tj Fau - Morris, D. E., Morris De Fau - Hayman, J. A., Hayman Ja Fau - Flickinger, J. C., Flickinger Jc Fau - Mills, M., Mills M Fau - Rogers, C. L., Rogers Cl Fau - Souhami, L., & Souhami, L. The

American Society for Therapeutic Radiology and Oncology (ASTRO) evidence-based review of the role of radiosurgery for brain metastases. (0360-3016 (Print)).

Minniti, G., Clarke, E., Lanzetta, G., Osti, M. F., Trasimeni, G., Bozzao, A., Romano, A., & Enrici, R. M. (2011). Stereotactic radiosurgery for brain metastases: analysis of outcome and risk of brain radionecrosis. *Radiation Oncology*, *6*(1), 48. https://doi.org/10.1186/1748-717x-6-48

Nelms, B. E., Robinson, G., Markham, J., Velasco, K., Boyd, S., Narayan, S., Wheeler, J., & Sobczak, M. L. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. (1879-8519 (Electronic)).

Noh, T., & Walbert, T. Brain metastasis: clinical manifestations, symptom management, and palliative care. (0072-9752 (Print)).

Otto, K. (2007). Volumetric modulated arc therapy: IMRT in a single gantry arc. *Medical Physics*, *35*(1), 310-317. https://doi.org/10.1118/1.2818738

Palmer, J. D., Sebastian, N. T., Chu, J., Dicostanzo, D., Bell, E. H., Grecula, J., Arnett, A., Blakaj, D. M., McGregor, J., Elder, J. B., Lu, L., Zoller, W., Addington, M., Lonser, R., Chakravarti, A., Brown, P. D., & Raval, R. (2020). Single-Isocenter Multitarget Stereotactic Radiosurgery Is Safe and Effective in the Treatment of Multiple Brain Metastases. *Advances in Radiation Oncology*, *5*(1), 70-76. https://doi.org/10.1016/j.adro.2019.08.013

Schouten, L. J., Rutten, J., Huveneers, H. A. M., & Twijnstra, A. (2002). Incidence of brain metastases in a cohort of patients with carcinoma of the breast, colon, kidney, and lung and melanoma. *Cancer*, *94*(10), 2698-2705. https://doi.org/10.1002/cncr.10541

Slagowski, J. M., & Wen, Z. (2020). Selection of single-isocenter for multiple-target stereotactic brain radiosurgery to minimize total margin volume. *Physics in Medicine & Biology*, *65*(18), 185012. https://doi.org/10.1088/1361-6560/ab9703

Soffietti, R., Rudā, R., & Mutani, R. (2002). Management of brain metastases. *Journal of Neurology*, *249*(10), 1357-1369. https://doi.org/10.1007/s00415-002-0870-6

Soliman, H., Das, S., Larson, D. A., & Sahgal, A. (2016). Stereotactic radiosurgery (SRS) in the modern management of patients with brain metastases. *Oncotarget*, *7*(11), 12318-12330. https://doi.org/10.18632/oncotarget.7131

Wowra, B., Muacevic A Fau - Tonn, J.-C., & Tonn, J. C. CyberKnife radiosurgery for brain metastases. (1662-3924 (Electronic)).