Research article

Radiomics model based on computed tomography images for prediction of radiation-induced optic neuropathy following radiotherapy of brain and head and neck tumors

Elham Raiesi Nafchi [a], Pedram Fadavi [b], Sepideh Amiri [c], Susan Cheraghi [d,a,*], Maryam Garousi [b], Mansoureh Nabavi [e], Iman Daneshi [f], Marzieh Gomar [e], Malihe Molaie [g], Ali Nouraeinejad [h]

[a] Department of Radiation Sciences, Faculty of Allied Medicine, Iran University of Medical Sciences, Tehran, Iran
[b] Department of Radiation Oncology, School of Medicine, Iran University of Medical Science, Tehran, Iran
[c] Department of Information Technology, Faculty of Electrical and Computer Engineering, University of Tehran, Tehran, Iran
[d] Radiation Biology Research Center, Iran University of Medical Sciences, Tehran, Iran
[e] Radiation Oncology Research Center (RORC), Cancer Institute, Tehran University of Medical Sciences, Tehran, Iran
[f] Department of Clinical Oncology, Haft-e-Tir Hospital, Iran University of Medical Science, Tehran, Iran
[g] Control and Intelligent Processing Center of Excellence, School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran
[h] Department of Optometry and Vision Science, School of Rehabilitation, Tehran University of Medical Science, Tehran, Iran

ARTICLE INFO

ABSTRACT

Purpose: We aimed to build a machine learning-based model to predict radiation-induced optic neuropathy in patients who had treated head and neck cancers with radiotherapy.
Materials and methods: To measure radiation-induced optic neuropathy, the visual evoked potential values were obtained in both case and control groups and compared. Radiomics features were extracted from the area segmented which included the right and left optic nerves and chiasm. We integrated CT image features with dosimetric and clinical data subsequently, ranked 5 supervised ML models Bernoulli Naive Bayes, Decision Tree, Gradient Boosting Decision Trees, K-Nearest Neighbor, and Random Forest on 4 input datasets to predict radiation-induced visual complications classifiers by implementing 5-fold cross-validation. The F1 score, accuracy, sensitivity, specificity, and area under the ROC curve were compared to access prediction capability.
Results: radiation-induced optic neuropathy affected 31 % of the patients. 856 radiomic characteristics were extracted and selected from each segmented area. Decision Tree and Random Forest with the highest AUC (97 % and 95 % respectively) among the five classifiers were the most powerful algorithms to predict radiation-induced optic neuropathy. Chiasm with higher sensitivity and precision was able to predict radiation-induced optic neuropathy better than right or left optic nerve or combination of all radiomic features.
Conclusion: We found that combination of radiomic, dosimetric, and clinical factors can predict radiation-induced optic neuropathy after radiation treatment with high accuracy. To acquire more reliable results, it is recommended to conduct visual evoked potential tests before and after radiation therapy, with multiple follow-up courses and more patients.

---

* Corresponding author.Radiation Biology Research Center, Iran University of Medical Sciences, Tehran, Iran.
E-mail address: cheraghi.s@iums.ac.ir (S. Cheraghi).

1. Introduction

Half of all cancer patients in the world are thought to be candidates for radiation therapy (RT), either curative or palliative [1,2]. The goal of modern treatment tactics is always to provide higher dosages to the targeted tissue to improve tumor control, but before such approaches can be properly adopted, an accurate and reliable understanding of the toxic effects on surrounding tissues must be established [3]. Normal tissues around the target region referred to as organs at risk (OARs), might be damaged too, resulting in RT-induced toxicity. Acute or short-term toxicity occurs during treatment or within three months of RT completion, and complete recovery usually takes weeks to months [4,5]. On the other hand, late effects, are considered to be permanent and increase over time. As a result, the potential advantages of any RT treatment must be evaluated against the risk of harm to normal tissues, with the ultimate goal of increasing therapeutic response while reducing the risk of normal tissue damage. When RT is used to treat a disease, target coverage should not be compromised to save OARs [6].

Radiation-induced optic neuropathy (RION) is a type of delayed radionecrosis of the anterior visual pathways that occurs in one or both eyes, months to years following external cranial irradiation and results in permanent vision loss [3]. As a result of damage to the optic nerve, the visual field is restricted to a variable extent. RION usually appears between 10 and 20 months following treatment, with an average of 18 months [3]. RION is mainly caused by cumulative doses of radiation exceeding 50 Gy or single doses to the anterior visual pathway exceeding 10 Gy. Early symptoms of radiation damage to the visual system have also been detected using electrophysiological testing [7]. Increased latency in the presence of preserved waveform morphology has been thought to be an indication of the demyelinating process [8]. As the nerve sheath is damaged, the time it takes for electrical signals to be transmitted to the eyes is prolonged, resulting in abnormal visual evoked potential (VEP). The VEP test results may be abnormal in patients with anterior visual pathway radionecrosis several months before the vision loss [8]. Therefore, abnormal VEP test results will be very valuable in head and neck patients.

The radiation tolerance of the optic apparatus often limits the therapeutic dose levels for tumors of the central nervous system and head and neck area [3,9]. Due to injury to a part of the optic nerves/chiasm, patients frequently lose vision in one-half or a quadrant of the visual field [10]. The average time between RT and the onset of visual problems is three years (mode: one to 1.5 years; median: 2.5 years) [3,10]. Except when exposure happens very close to the optic chiasm, optic nerve injury usually results in monocular visual loss. Bilateral vision loss can result from injury to the whole chiasm [10]. Also, concurrent chemotherapy has been identified as a possible risk factor for RION. Chemotherapeutics such as vincristine, nitrosourea, and cisplatin have been linked to direct optic nerve injury and are thought to have a radio sensitizing effect, increasing the risk of RION in patients taking adjuvant chemotherapy [11–14]. For example, when cisplatin is used, it might cause vision loss due to retinopathy [15–19]. The cause of cisplatin-induced ocular damage is unknown, however, it could be due to drug accumulation in the central nervous system after repeated doses, especially in high-dose platinum-containing regimens [13].

According to prior studies, it seems radiation toxicity prediction and assessment techniques are critical for limiting RT-related side effects. In the RT scientific community, there is a rising interest in machine learning (ML) for the prediction of radiation-related damage, based on medical images utilizing radiomics and machine learning models [1,20–24]. Several studies in recent years have

Table 1

Patients' characteristics, clinic and dosimetric data, and result of visual evoked potential test.

Data	Type	Parameters	Range (Median percent)
Clinical patient data	Patient characteristics	Age	20–67 years (46 years)
		Male	28 (54 %)
		Female	24 (46 %)
		weight	48-85 Kg (68 Kg)
		height	154 Cm (168 Cm)
		eye color	41 brown (79 %) 10 black (19 %) other (2 %)
		surgery history	13 (12.5 %)
		alcohol consumption	1 (2 %)
		smoking	15 (29 %)
	Control group characteristics	Age	20–70 years (49 years)
		Male	26 (50 %)
		Female	26 (50 %)
	Tumor types	GBM (Glioblastoma multiform)	24 (46 %)
		Meningioma	4 (7.68 %)
		Astrocytoma	4 (7.68 %)
		ODM (Oligodendroglioma)	3 (5.76 %)
		NPC (Nasopharyngeal carcinoma)	17 (32.64 %)
Dosimetric data	Treatment protocol	Beam energy	6 Mev
		dose per fraction	2 Gy
		number of fractions	25–35
	Radiation dose	Total prescribed dose	50 -70 (58.9 Gy)
		Left optic nerves dose	(13.35 Gy)
		Right optic nerves dose	(15.76 Gy)
		Chiasm dose	(24.67 Gy)
Result of visual evoked potential test	Patient optic neuropathy	Positive	15 Male (15 %)/female (13 %)
		Negative	37 Male (38 %)/female (32 %)

focused on radiomics models and their potential to extract high-dimensional quantitative features that can be implemented in personalized treatment [20,25–27]. Radiomics is a technique for extracting quantitative information from medical images [21]. Some studies use ML approaches to determine side effects, factors associated with morbidity, adaptive RT, quantitative medical images, and tumor grading in head and neck cancers is useful [20,25,26,28–30]. In the present study, we aimed to build a machine learning-based model to predict the risk of RION after RT of brain and head and neck tumors.

2. Material and method

2.1. Patient information

The VEP test was performed on 52 adult patients (28 males and 24 female) who had completed RT of head and neck or brain cancers 1–3 years before that. The results of patients' VEP were compared with a reference or control group. The control group consisted of 52 patients who had head and neck or brain cancer but whose treatment sessions had not started. Patients' characteristics, clinic and dosimetric data, and result of visual evoked potential test information is presented in Table 1. We obtained IRB/Ethics approval no. IR. IUMS.REC.1398.1336 to conduct the study, also Informed consent was provided in both case and control groups of patients. Patients who had chemoradiation therapy, recurring or metastatic diseases, hypertension, diabetes, glaucoma, refractive errors, neurological diseases, and a history of eye inflammation were excluded from this study. The clinical and dosimetric features investigated in this study were as follows: the total dose of radiation, minimum, maximum, and mean dose of right and left optic nerves and chiasm, total prescribed dose, dose per fraction, number of dose fractions, sex, age, weight, height, pathology of tumor, surgery history, smoking, alcohol consumption, and eye color.

### 2.2. Radiation therapy or radiation chemotherapy protocols

All 52 patients were treated with 3-dimensional conformal radiation therapy (3D-CRT) at 50–70 Gy and five fractions per week in 1–3 phases of treatment planning. RT treatment planning was performed based on head and neck CT scan images of patients. Images were prepared by 0.5–2 mm slice thicknesses and 2 mm intervals.

2.3. VEP test

VEPs were performed to confirm the presence of visual pathology or to detect asymptomatic subclinical visual pathway dysfunction [31,32]. After proper skin preparation and application of a conducting electrode gel or paste, three electrodes were placed on the scalp such that the reference electrode (Fz) was placed at the forehead, the ground electrode (Cz) at the vertex, and the active electrode (Oz) at approximately 2 cm above the inion. The patient sat at a suitable distance from the visual stimulus in a dim room, and a stimulus generator was employed to run the required type of pattern stimulus while the non-tested eye was occluded. Responses were amplified and averaged by the machine, while the patient kept focus on the stimulus displayed on the monitor. Recorded responses were documented on graphs and analyzed based on their latencies and amplitudes in milliseconds and microvolts, respectively [31–33]. The VEP results were classified as "0" and "1". Zero represented patients whose optic nerve function was healthy, whereas "1" denotes patients who had functional visual impairment.

2.4. Proposed framework

The presented method involves five main steps: Preprocessing, Segmentation, Feature Extraction, Feature Selection, and
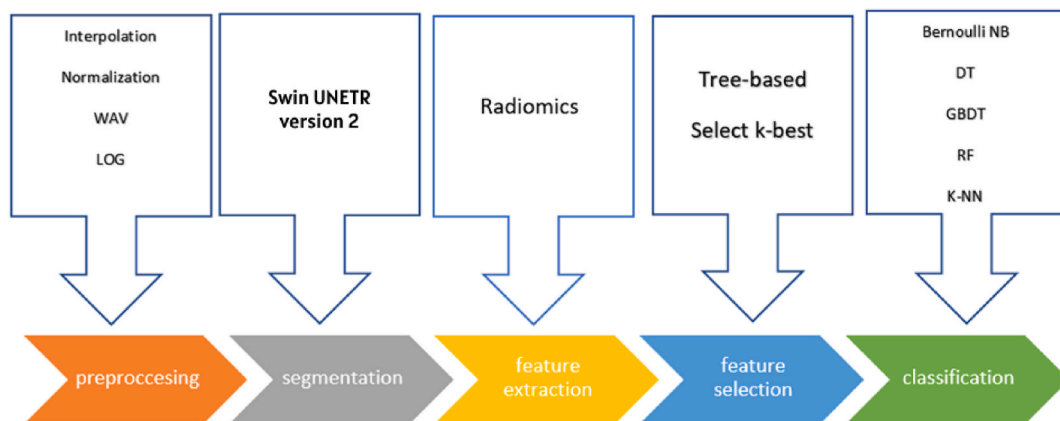
Fig. 1. Proposed framework for radiomics analysis on computed tomography images for predicting radiation-induced optic neuropathy by machine learning models.

Classification. The flowchart of the method is shown in Fig. 1 and explained in detail below.

Step 1. (Preprocessing):

The images were resampled using the voxel size resampling method to create an isotropic dataset. For all of the patients, interpolation was used to facilitate comparisons of image data from different samples and scanners. All CT scan images were preprocessed with preprocessing algorithms including interpolation and Wavelet Decomposition (WAV) [34] improve the quality of CT images.

Step 2. (Segmentation):

Since we aim to develop a fully automated deep learning pipeline, we require an expertly segmented dataset to guide our implementation. Therefore, a radiation oncologist manually segmented our dataset Fig. 2 shown segmented area of right and left optic nerves and chiasm. The Swin UNETR version 2 deep learning model was employed to pursue precise optic nerve and chiasm segmentation, the results of which can be seen in Table 2. During the training phase, a multitude of data augmentation techniques were harnessed, including random flips across spatial axes 1, 2, and 3, random rotations with a 10% flipping probability, and random shift intensity with a 50% flipping probability within an offset range of 10%. The computational framework featured an NVIDIA GeForce RTX 3080 GPU, complemented by Python 3.8.10 and Torch 1.9.0 + cu111. Patch-based Swin UNETR was employed, utilizing patch sizes of 96 by 96 by 96 over a span of 30,000 training iterations [35]. This method was implemented using the MONAI library [36], with batch sizes of 2 for training and 1 for testing. An Adam optimizer was chosen, accompanied by a weight decay of 1e-5, and the learning rate was set at 1e-4. In the initial stage, the dataset was partitioned into distinct segments, with 12% reserved for evaluation, 10% for testing, and the remaining 78% allocated for the training process. Swin UNETR is a novel architecture based on shifted windowing transformers (Swin) that limits self-attention computations to non-overlapping windows while allowing cross-window communication to improve performance. The model was implemented as a patch-based Swin UNETR, utilizing patch sizes of 96 by 96 by 96, and each patch is embedded into a high-dimensional space. This can be represented as $X = = \{x_p^1, x_p^2, ..., x_p^3\}$.

Here, X is the set of patch embeddings, and x_p^i represents the embedded vector of the ith patch. Swin transformer blocks are the core of the model and involve self-attention mechanisms. The self-attention calculation in transformers is generally given by *Attention* $(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right) V$. Here, Q, K, and V are the query, key, and value matrices, respectively, and d_K is the dimension of the key. The model architecture and design were detailed in the work by Hatamizadeh [35].

Step 3. (Feature Extraction):

Radiomic features are extracted from the segmented area. We used a bin width of 25 Hounsfield unit for intensity discretization of voxels, and voxel size resampling $1 \times 1 \times 1$ mm^3 These settings have been proven to provide reliable results in similar radiomics-based machine learning studies, ensuring comparability and reproducibility of our findings [37–39]. The performance of decision support models can be improved by combining radiomic data with other patient characteristics [40,41]. We combined radiomic features with
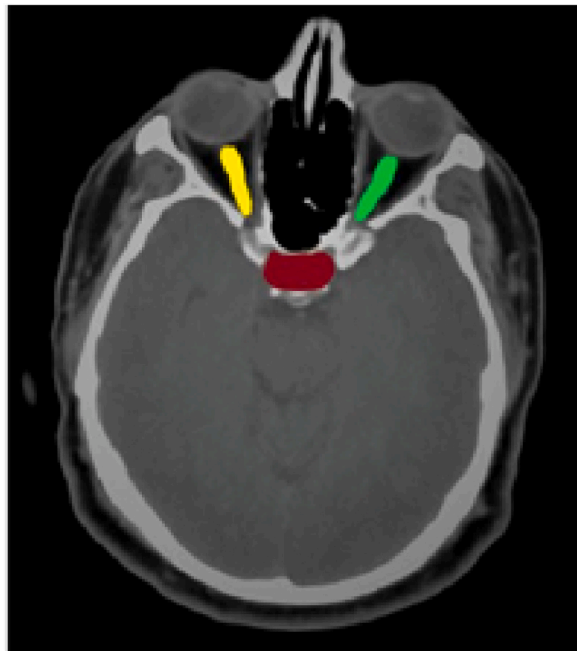

Fig. 2. Segmented area of right and left optic nerves and chiasm.

Table 2

Comparison of segmentation performance for right optic nerve, left optic nerve, and chiasm using Swin UNETR version 2: Mean $\pm$ standard deviation of dice similarity coefficient (DSC) Results.

Organ	Dice Score
Right optic nerve	0.89 ± 0.09
Left optic nerve	0.92 ± 0.13
Chiasm	0.91 ± 0.21

patient characteristics and dosimetric parameters in this study. Here, we presented a concise summary of several radiomic features. First-order statistics describe the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics. A homogenized execution size is measured for each gray surface using the Gray-Level Run Length Matrix (GLRLM). The second-order joint probability function of an image area bound by the mask is described by the Gray-Level Co-occurrence Matrix (GLCM). A Gray Level Dependence Matrix (GLDM) quantifies gray-level dependencies in an image. A gray level dependency is defined as the number of connected voxels within distance δ that is dependent on the center voxel. We applied Z-score normalization across all features to ensure that each feature had a mean of 0 and a standard deviation of 1. The normalization process was carried out after preprocessing and prior to the data being split. This approach was done to scale independent features and prevent any single feature from disproportionately influencing the model during training.

Step 4. (Feature Selection):

Selected features were used as input to each of the 5 ML classifiers. There are several methods for selecting features that lead to a high-performing classifier and the discovery of "effective" features. Univariate feature selection selects the best features based on univariate statistical tests. Select K-Best removes all but the k highest scoring features, and Select Percentile removes all but a user-specified highest scoring percentage of features. Also, tree-based estimators can be used to compute impurity-based feature importance, which in turn can be used to discard irrelevant features. Since the aim of this research was to determine the feature's importance and its relation to RION, we chose numerical extracted features and then implemented K-Best and tree-based feature selection strategies to select the effective features in RION. Because of a lot of features selected as output in every two methods of feature selection all of the features that were the same in the two techniques were selected as input of classifiers.

Step 5. (Hyperparameter optimization)

We performed hyperparameter optimization to maximize the performance of each model (Table 3). Utilizing a grid search strategy, we identified the optimal hyperparameter combinations for each machine learning model. This process involved testing various combinations and evaluating model performance based on accuracy, AUC, and F1 score through 5-fold cross-validation. By fine-tuning these hyperparameters, we successfully enhanced the performance of each model. The final selection of parameters was based on the model's performance metrics, specifically the Area Under the Curve (AUC) and the cross-validated F1 score.

Step 6. (Classification):

We implemented K-Nearest Neighbor (KNN), Bernoulli Naive Bayes (Bernoulli NB), Decision Tree (DT), Gradient Boosting Decision Trees, and Random Forest (RF) using Python 3.7 and Pycharm 2019.3.1. The KNN algorithm tools learn based on the nearest neighbors of each query point, where an integer value defined number of neighbors is five in this study. Bernoulli NB can compute how often each feature corresponding to each class isn't zero. Decision Trees (DTs) are an anon-parametric supervised learning approach used for classification and regression. The aim is to produce a model that predicts the value of a target variable by learning simple decision rules concluded from the data features. A tree can be seen as a piecewise constant approximation. Gradient boosting decision tree builds an additive model in a forward stage-wise fashion; it allows for the optimization of arbitrary differentiable loss functions. A random forest

Table 3

The ranges and the selected best hyperparameters for the algorithms.

Algorithms	Hyperparameters	Values	Best value/Best parameter
Random Forest (RF)	**Number of Estimators (trees)**	50, 100, 200	200
	Max Depth	10, 20, 30, None	20
	Min Samples Split	2, 5, 10	5
Gradient Boosting Decision Trees (GBDT)	**Learning Rate**	0.01, 0.1, 0.2	0.1
	Number of Estimators (trees)	50, 100, 200	200
	Max Depth	3, 5, 10	5
K-Nearest Neighbors (KNN)	**Number of Neighbors (K)**	3, 5, 7, 10	5
	Weights	Uniform, Distance	Distance
	Algorithm	Auto, Ball Tree, KD Tree	Auto
Decision Tree (DT)	**Max Depth**	10, 20, 30, None	10
	Min Samples Split	2, 5, 10	5
Bernoulli Naive Bayes (NB)	**Alpha (Smoothing parameter)**	0.01, 0.1, 1.0, 10.0	1.0

is a meta-estimator that fits several decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. We employed a data splitting strategy where 70 % was assigned for training 20 % for validation and 10 % of the dataset for testing. The algorithms were trained using 5-fold cross-validation in our study. Instead of separating the dataset into training and a test set, cross-validation splits the data repeatedly and trains numerous models, based on the summation of correct and incorrect estimation. When applied to new data, this strategy gives us an idea of how our model will perform in the best- and worst-case scenarios. Six evaluation criteria were used for comparison: accuracy, sensitivity, specificity, mean cross-validation, F1 score, and area under the receiver operating characteristic (ROC) and curve (AUC). AUC was used to evaluate the prognostic performance of different feature selection and classification methods. The confusion matrix highlights the balance between the model's ability to correctly identify patients with RION and its ability to avoid false cases. To further assess the robustness of our models and account for small data sizes, we used bootstrapping with 1000 iteration. The dataset was resampled with replacement in each iteration, and the model was retrained. Then, we calculated the mean, standard deviation, and 95 % confidence interval (accuracy, precision, recall, F1 score. This approach ensures that our evaluation does not depend on any data partitioning.

2.5. Datasets

We created four datasets as input for 5 ML algorithms. The first dataset was based on features extracted from the right optic nerve, the second dataset for the left optic nerve, the third dataset for chiasm, and finally, the last dataset for all features from both optic nerves and chiasm features together. Data sets included patient data (dosimetric, clinical, and VEP results) plus each selected feature listed above. In this study, patients were selected based on the availability of complete clinical and dosimetric data. As a result, there were no missing data points that required deletion or extrapolation. All included patients had comprehensive datasets, ensuring the integrity of our analysis. Annotation of the right and left optic nerves and chiasm using 3D-Slicer software is done by experienced radiotherapists to define the region of interest (ROI). This process involves meticulously iterating through an image's slices and drawing a contour at the boundary to accurately delineate the areas of interest.

3. Results

A total of 856 features were extracted from the segmented area, including the right and left optic nerve and chiasm. Although optic nerves and chiasm were OARs that were not exposed to direct radiation, the mean radiation dose to the right optic nerve, left optic nerve, and chiasm was 14.95 Gy, 12.85 Gy, and 23.32 Gy, respectively. Visual complications occurred in 15 of 52 patients (29 %). It means 15 % of males and 13 % of females experienced RION 3 years after RT treatment. The result of the modeling with five implemented models is shown in Table 4. As shown, ML results of using a dataset that includes right optic nerve features show that the DT classifier has the best performance among the other classifiers (AUC = 0.97). 14 important features in this data set are identified in Fig. 3A. Table 4 shows that the KNN classifier outperforms the others when we used the dataset containing the left optic nerve for prediction of RION (AUC = 0.84). 11 essential features are shown in Fig. 3B. When we used the dataset containing visual chiasm, the RF classifier had more sensitive and accurate results than the other classifiers (AUC = 0.95), and 13 important features are shown in Fig. 3C. In addition, in the fourth scenario, we merged all features extracted from both optic nerves and chiasm in combination with

Table 4
Results of machine learning models performances in predicting radiation-induced optic neuropathy based on the right optic nerve-left optic nerve, chiasm and merged dataset.

Dataset	Algorithms	AUC	F1 Score	Accuracy	Sensitivity	Specificity	Mean cross validation
right optic nerve dataset	RF	0.85	0.95	0.89	0.92	0.63	0.778
	GBDT	0.77	0.86	0.83	0.90	0.67	0.84
	KNN	0.82	0.84	0.84	0.91	0.54	0.82
	DT	0.97	0.76	0.76	0.85	0.25	0.79
	NB	0.78	0.74	0.60	0.68	0.98	0.64
left optic nerve dataset	RF	0.82	0.92	0.91	0.88	0.25	0.655
	GBDT	075	0.96	0.83	0.91	0.69	0.81
	KNN	0.84	0.92	0.92	0.95	0.5	0.79
	DT	0.78	0.73	0.88	0.94	0.66	0.71
	NB	0.66	0.84	0.84	0.90	0.5	0.640
chiasm dataset	RF	0.95	0.95	0.92	0.95	0.75	0.678
	GBDT	0.73	0.93	0.88	0.95	0.50	0.74
	KNN	0.84	0.86	0.88	0.95	0.68	0.76
	DT	0.90	0.92	0.88	0.94	0.66	0.70
	NB	0.78	0.74	0.60	0.58	0.88	0.64
merged dataset	RF	0.83	0.92	0.88	0.95	0.50	0.768
	GBDT	0.79	0.87	0.80	0.80	0.75	0.77
	KNN	0.85	0.88	0.82	0.95	0.50	0.77
	DT	0.93	0.93	0.88	0.98	0.50	0.72
	NB	0.78	0.75	0.68	0.60	0.98	0.64

RF= Random Forest, GBDT = Gradient Boosting Decision Trees, RF =Random Forest, KNN = K-Nearest Neighbor, DT = Decision tree, NB= Bernoulli Naive Bayes.
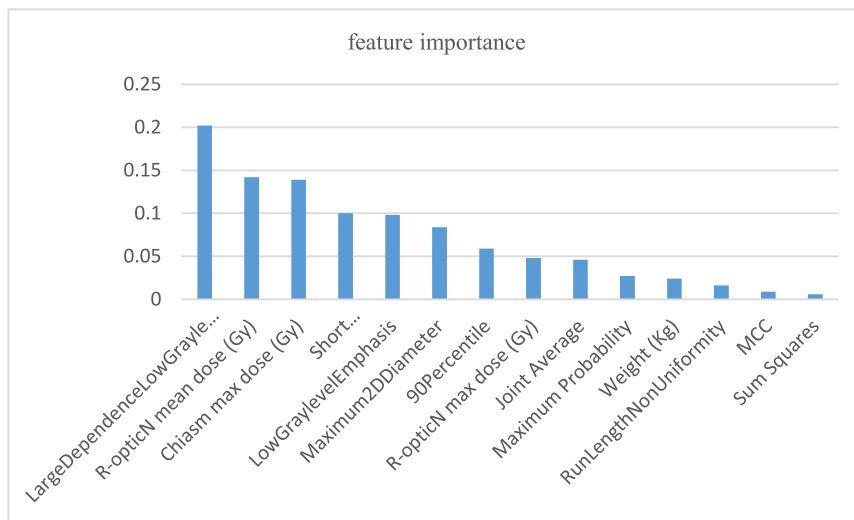
A)
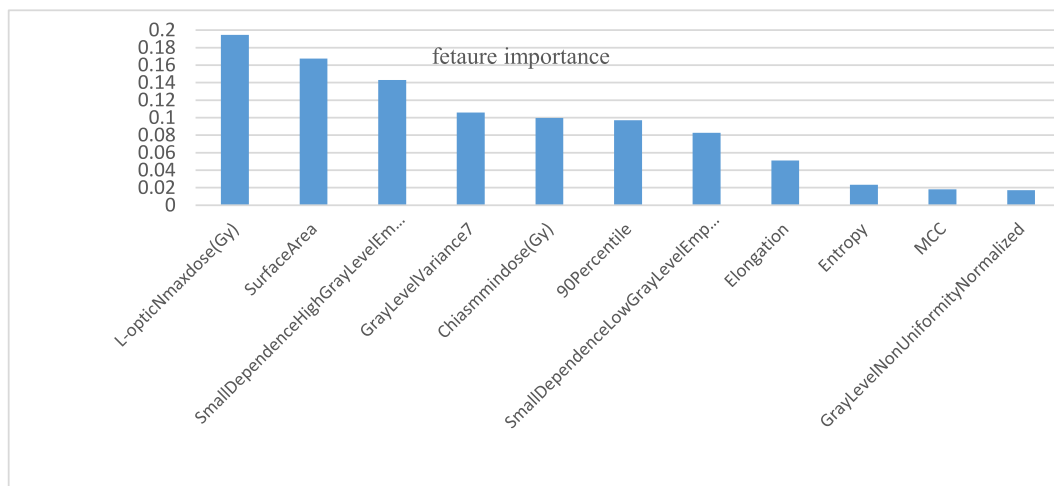
B)

Fig. 3. Random forest classifier feature importance in (A) right optic nerve dataset (B) left optic nerve dataset (C) chiasm dataset and (D) merged dataset (left nerve optic, right nerve optic and chiasm radiomics features).

clinical and individual features of patients and implemented the models to identify the role of important features in RION complications. DT classifier was the most important algorithm (AUC = 0.93). Ten important features are presented in Fig. 3D Pie chart in Fig. 4A-D representing the distribution of various families of important radiomic features in the: right nerve dataset, left nerve dataset, chiasm dataset and combined dataset, respectively. Also, the area under the receiver operating characteristic (ROC) curve of Gradient Boosting Decision Trees in all datasets is shown in Fig. 5A-D. The RF model consistently demonstrated strong performance across all datasets, particularly with high AUC and F1 scores. Sensitivity was generally higher than specificity, indicating a stronger ability to detect positive cases. We generated a confusion matrix for the RF (chiasm dataset) and DT model (right optic nerve dataset) as the best-performing models in Table 5. Furthermore, to increase the robustness of the results, we applied bootstrapping with 1000 iterations to calculate the mean, standard deviation, and 95 % confidence interval for each metric. Confidence intervals further confirm the validity of the reported measures. Table 6 summarizes bootstrapping result, the performance metrics of RF ML algorithms applied to the right optic nerve, left optic nerve, chiasm, and merged datasets. These measures include AUC, F1 score, accuracy, sensitivity, specificity, and mean cross-scores. Finally, we analyzed feature combinations, to assess the effect of combining radionics, clinical, and dosimetric features on model performance. The results of implemented the RF models for all radiomics or top ranked radiomics features with all clinical or all dosimetric data and all clinical with all dosimetric features are comparable in Table 7.
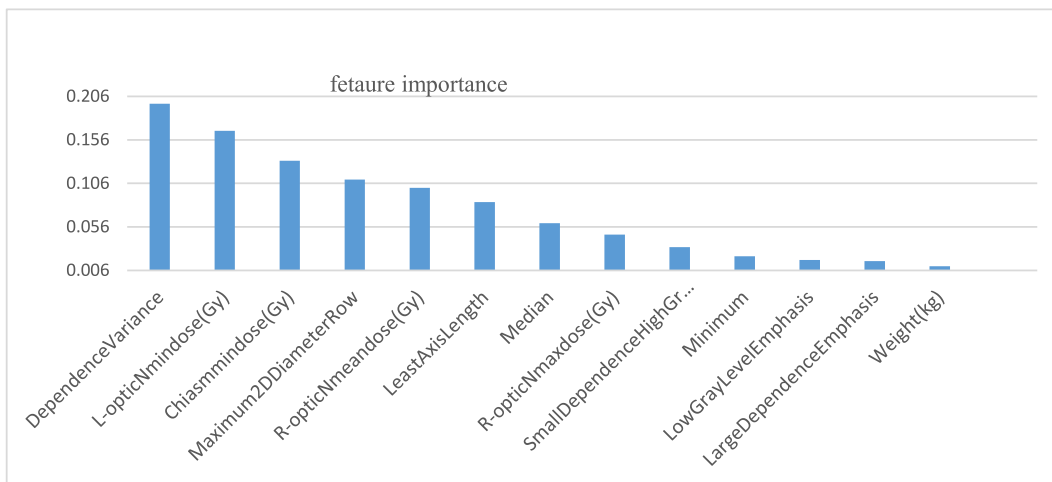
C)
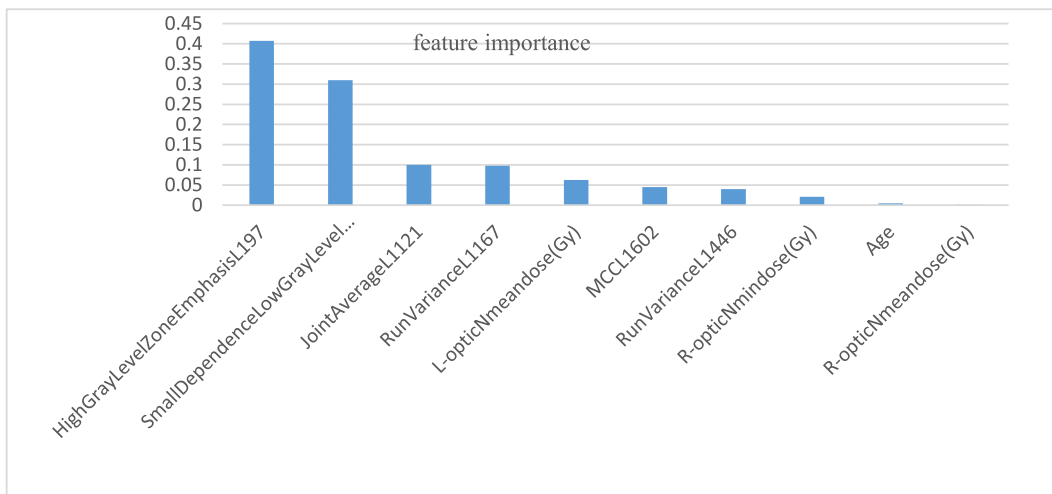
D)

Fig. 3. (*continued*).

4. Discussion and conclusion

The purpose of this work is to implement 5-fold cross-validation to evaluate five supervised ML models to predict visual complications in the patients who had treated head and neck cancers with RT. It is important to identify high risk patients prior to starting treatment to prevent potential vision problems that could result from the treatment. Our study findings identified an incidence of visual complications in 15 out of 52 patients (29 %), indicating a significant risk associated with RT. In traditional RT studies, it is estimated that 75 % of patients exposed to radiation will experience vision loss in both eyes [42]. However, modern IMRT techniques and proton therapy may reduce the incidence and severity of RION [43,44] as, incidence rate in proton therapy is reported to be 4.2 % [44]. In our study, the dose received by the optic nerve and chiasm dose were among the most critical dosimetric parameters influencing the development of RION, emerging as a top feature across all datasets. These findings align with the results of other studies in this domain [10,45,46]. The high importance of the left or right optic nerve dose in the chiasm dataset (Fig. 3C) could be attributed to the close anatomical proximity of the optic system. This result suggests that interactions between dosimetric factors of the left and right optic nerves play a significant role in predicting RION via chiasm dataset.

We observed a strong power of integrating clinical, and dosimetric with imaging features for predictive RION, with notable advantages. The RF model's robust performance with the chiasm dataset (AUC = 0.95) and the DT model's effectiveness with the right optic nerve dataset (AUC = 0.97) showcase the potential of targeted datasets to enhance predictive accuracy. Importantly, the chiasm
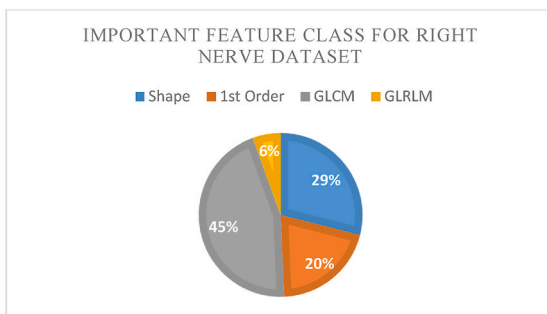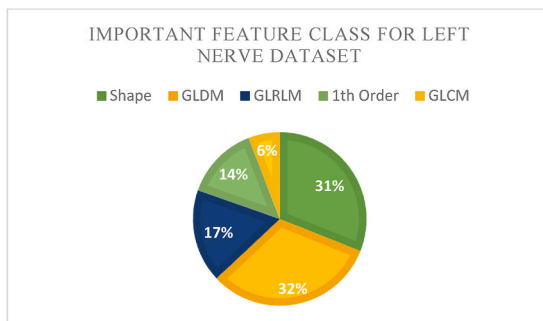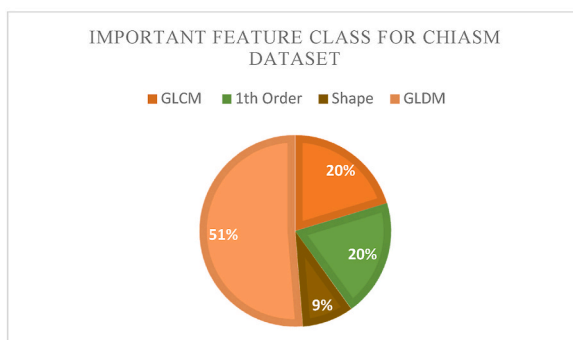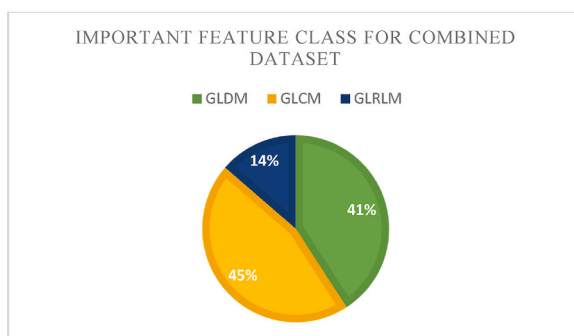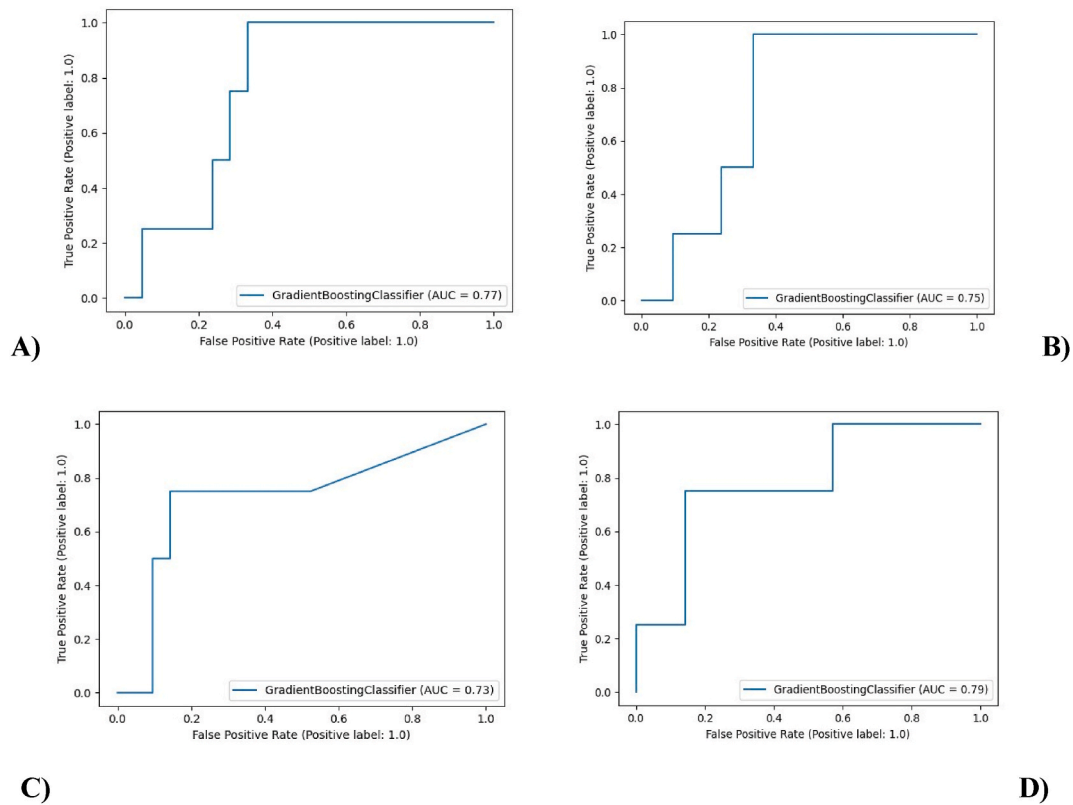
Fig. 4. Pie Chart representing the distribution of various families of important radiomic features in the: A) right nerve dataset B) left nerve dataset C) chiasm dataset D) combined dataset.

Fig. 5. Area under the receiver operating characteristic (ROC) curve of Gradient Boosting Decision Trees in (A) Right optic nerve dataset (B) Left optic nerve dataset (C) Chiasm dataset and (D) Merged dataset (left optic nerve, right optic nerve and chiasm).

Table 5
Confusion matrix for best classifiers.

Algorithms		Predicted Positive	Predicted Negative
Decision Tree classifiers (right optic nerve dataset)	**Actual Positive**	12	3
	Actual Negative	5	32
Random Forest classifiers (chiasm dataset)	**Actual Positive**	13	2
	Actual Negative	1	36

Table 6
Aggregated results of the bootstrapping process for RF machine learning algorithms, showing performance metrics for right optic nerve, left optic nerve, chiasm, and merged datasets. Measures include mean values and standard deviations for AUC, F1 score, accuracy, sensitivity, specificity, and mean cross-validation scores, along with 95 % confidence intervals calculated through bootstrapping.

Metrics	Mean	Standard Deviation	95 % Confidence Interval
AUC	**0.821**	**0.075735**	**(0.7855549312512157, 0.8564450687487842)**
F1 Score	**0.8635**	**0.079225**	**(0.826421622680746, 0.9005783773192538)**
Accuracy	**0.826**	**0.096048**	**(0.7810480430122196, 0.8709519569877807)**
Sensitivity	**0.8745**	**0.118298**	**(0.8191347192337798, 0.9298652807662203)**
Specificity	**0.6185**	**0.197545**	**(0.5260461884431542, 0.7109538115568459)**
Mean Cross Validation	**0.73295**	**0.066373**	**(0.7018863271709395, 0.7640136728290605)**

dataset's superior sensitivity and precision illustrate its capability to identify fine changes in adjacent structures affected by radiation, likely due to its unique position within the optic apparatus and its sensitivity to radiation effects on surrounding tissues. Based on confusion matrix, RF algorithm in chiasm dataset performed very well, with only 1 false positive and 2 false negatives while, DT in right optic nerve dataset had slightly more false positives (5) and false negatives (3). Radiation damage to the chiasm is important because it can cause loss of bilateral vision, while damage to the optic nerve usually results in loss of monocular vision [10]. Also, optical chiasm has a key role in the processing of visual information, spatial vision, and convergence of binocular information, which

Table 7
Results of running the models for the following combinations: Radiomics or Top-ranked Radiomics combined with Clinical features, Radiomics or Top-ranked Radiomics combined with Dosimetric features, and Clinical combined with Dosimetric features.

Feature Combination	Algorithm	AUC	F1 Score	Accuracy	Sensitivity	Specificity	Mean Cross-validation
Radiomics + Clinical data	Random Forest (RF)	0.90	0.88	0.85	0.90	0.70	0.82
Top-ranked Radiomics + Clinical data		0.91	0.89	0.87	0.91	0.75	–
Radiomics + Dosimetric data	Random Forest (RF)	0.93	0.92	0.89	0.93	0.75	0.85
Top-ranked Radiomics + Dosimetric data		0.95	0.93	0.91	0.94	0.80	–
Clinical + Dosimetric	Random Forest (RF)	0.77	0.74	0.62	0.78	0.65	0.70

enhances three-dimensional acuity and improves motion detection. Our results align with previous research, showed that tree based models can easily be combined with various data such as clinical, imaging and genetic data to provide more comprehensive analysis [47–49]. These models due to attention to strong interactions between variables, can well identify nonlinear relationships between characteristics and outputs, consequently, they are more suitable for use with decision-making algorithms that follow the same structure which is very important in the field of medicine [48].

By removing common features in 4 dataset results, 19 effective radiomic features were identified as the most important in RION prediction. These features are classified into 5 classes, including GLRLM (5 features), GLCM (4 features), shape features (4 features), GLDM (3 features), and first-order features (3 features). By identifying GLRLM and GLCM-based features as key predictors, reinforces the role of advanced radiomic features in detecting texture variations that correlate with physiological damage, offering clinicians a new layer of diagnostic insight. GLRLM is a texture analysis technique utilized in radiomics, and its contribution to prediction merit further exploration; as understanding their biological foundations could offer insights for reducing complications and enhancing patient outcomes. A Gray-level run can be described as a line of pixels in a certain direction with the same intensity value. Other studies have also highlighted the importance of the features of the GLRLM in treatment response and complication predicting related RT, including, Handong Li et al. reported that the GLRLM can enhance the prediction of survival in patients with cervical cancer [50]. Similarly, Liting Shi's study confirmed that GLRLM influences treatment outcome prediction for non-small cell lung cancer [51]. In another study by Muditha S Bandara, GLRLM-based Normalized Run Length non-uniformity was identified as the most significant feature in differentiating between normal and diseased kidneys (accuracy = 0.85, sensitivity = 1.0) in ultrasound images [52]. GLCM features are known as spatial gray level dependency matrices that are the most widely used second-order statistical tools for extracting texture information from images that was important for Rion prediction in our study. Similar to our result, Khaled Bousabarah et al. utilizing radiomic features of the primary tumor, predict radiation-induced lung injury and outcome in non-small cell lung cancer patients treated with robotic stereotactic body radiation treatment [53]. They found that GLCM features had a considerable impact on all endpoints, which was maintained in a multifactorial Cox regression analysis that included other clinical and dosimetric factors.

Our results indicated that incorporating top-ranked radiomics features with all clinical data showed slight improvement across all metrics. Furthermore, adding top-ranked radiomics features to the dosimetric data further enhanced the model's performance, improving the AUC from 0.93 to 0.95. This combination can better differentiate between patients with and without RION. This approach confirms that radiomics features complement dosimetric data in predicting RION, with better sensitivity and specificity than either dataset alone. These findings further validate the robustness of radiomics in predicting radiation-induced complications as the other studies [24,54,55]. Shane P. Krafft et al. used pretreatment CT radiomics features extracted from the normal lung volume to see if there were any increases in prediction model performance for radiation pneumonitis. By using only, the clinical and dosimetric data, the average cross-validated AUC was 0.51. When entire lung CT radiomics features were included, the cross validated AUC was 0.68 [56].

In conducted RT studies, dose-response data modeling such as tumor control probability and normal tissue complication probability (NTCP) prediction, are limited to just considering organ radiation sensitivity and dosimetric parameters [57]. In contrast, ML algorithms consider a broad range of patient-specific information, dose-outcome connections, and imaging features to predict complications probability of normal tissue [58]. In a study, Vitali Moiseenko et al. investigated four NTCP models to fit data on radiation-induced dose-response neuropathy finding a correlation between higher D50 and increased RION [59]. Consistent with our study, they found radiation dose received by the right and left optic nerves and chiasm are related to RION. Of course, studies beyond our study are required to reach a common conclusion for choosing the predominant strategy, but what is certain is that in recent years many works have been done by radiomics and ML models to predict radiation-induced damage in normal tissue based on CT images [15,19,26]. Although our study supported our hypotheses statistically, there were some limitations. These included follow-up time because of the late effect of neuropathy and merging data manners. It seems it will be valuable to implement feature extraction in a multi-modality image approach which could expand the scope of radiomics in cancer treatment outcomes. Our results suggest that radiomic analysis could serve as a non-invasive adjunct to traditional dosimetric approaches, providing early indicators of RION risk and enabling proactive treatment adjustments. This work lays the foundation for developing personalized RT planning protocols that consider individual variability in tissue response, positioning ML-driven radiomics as a transformative tool in enhancing patient outcomes in radiation oncology. Furthermore, merging data related to radiomics with genomics, transcriptomics, and proteomics could guide researchers to in-depth consideration of various factors before irreversible consequences take place in radiotherapy.

In conclusion, we suggest clinical, dosimetric, and radiomics data provide better results than any dataset alone. our comparative analyses using ML models across various combinations of radiomic features and clinical/dosimetric data emphasize the interdependencies between these variables. Establishing these relationships could lead to the development of predictive protocols that

incorporate both clinical and radiomic data, allowing for a more patient-specific and therefore potentially more effective treatment planning process and ultimately improving patient quality of life and outcomes. Future studies could further investigate these promising areas, including the exploration of additional demographic, genetic, and biological risk markers. This exploration would enhance our understanding of RION and aid in the development of optimized patient care protocols.

CRediT authorship contribution statement

Elham Raiesi Nafchi: Writing – original draft, Resources, Investigation, Data curation. **Pedram Fadavi:** Supervision. **Sepideh Amiri:** Writing – review & editing, Validation, Methodology, Formal analysis. **Susan Cheraghi:** Supervision, Project administration. **Maryam Garousi:** Supervision. **Mansoureh Nabavi:** Resources. **Iman Daneshi:** Methodology. **Marzieh Gomar:** Supervision. **Malihe Molaie:** Formal analysis. **Ali Nouraeinejad:** Methodology.

Availability of data and material

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] L.J. Isaksson, et al., Machine learning-based models for prediction of toxicity outcomes in radiotherapy, Front. Oncol. 10 (2020) 790.
[2] E.B. Elkin, P.B. Bach, Cancer's next frontier: addressing high and increasing costs, JAMA 303 (11) (2010) 1086–1087.
[3] H.V. Danesh-Meyer, Radiation-induced optic neuropathy, J. Clin. Neurosci. 15 (2) (2008) 95–100.
[4] S.R. Mahdavi, et al., Comparison of radiation and chemoradiation-induced sensorineural hearing loss in head and neck cancer patients, J. Cancer Res. Therapeut. 16 (3) (2020) 539–545.
[5] A. Rezaeyan, et al., Audiometric findings in patients with head and neck chemoradiotherapy and radiotherapy: short-term outcomes, Int J Radiat Res 17 (2019) 633–641.
[6] L. Henke, et al., Phase I trial of stereotactic MR-guided online adaptive radiation therapy (SMART) for the treatment of oligometastatic or unresectable primary malignancies of the abdomen, Radiother. Oncol. 126 (3) (2018) 519–526.
[7] M. Lecornu, et al., Prospective assessment of early proton therapy-induced optic neuropathy in patients with intracranial, orbital or sinonasal tumors: impact of A standardized ophthalmological follow up, Front. Oncol. 11 (2021) 2271.
[8] K.A. Leber, J. Berglöff, G. Pendl, Dose—response tolerance of the visual pathways and cranial nerves of the cavernous sinus to stereotactic radiosurgery, J. Neurosurg. 88 (1) (1998) 43–50.
[9] S. Lessell, Friendly fire: neurogenic visual loss from radiation therapy, J. Neuro Ophthalmol. 24 (3) (2004) 243–250.
[10] C. Mayo, et al., Radiation dose–volume effects of optic nerves and chiasm, Int. J. Radiat. Oncol. Biol. Phys. 76 (3) (2010) S28–S35.
[11] J.D. Griffin, M.B. Garnick, Eye toxicity of cancer chemotherapy: a review of the literature, Cancer 48 (7) (1981) 1539–1549.
[12] P. Lambin, et al., Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[13] L. Leocani, S. Medaglini, G. Comi, Evoked potentials in monitoring multiple sclerosis, Neurol. Sci. 21 (2) (2000) S889–S891.
[14] P.A. Sanderson, T. Kuwabara, D.G. Cogan, Optic neuropathy presumably caused by vincristine therapy, Am. J. Ophthalmol. 81 (2) (1976) 146–150.
[15] H. Abdollahi, et al., Cochlea CT radiomics predicts chemoradiotherapy induced sensorineural hearing loss in head and neck cancer patients: a machine learning and multi-variable modelling study, Phys. Med. 45 (2018) 192–197.
[16] H.H. Chao, et al., Exploratory analysis using machine learning to predict for chest wall pain in patients with stage I non-small-cell lung cancer treated with stereotactic body radiation therapy, J. Appl. Clin. Med. Phys. 19 (5) (2018) 539–546.
[17] E.D. Gennatas, et al., Preoperative and postoperative prediction of long-term meningioma outcomes, PLoS One 13 (9) (2018) e0204161.
[18] V. Moiseenko, W.Y. Song, L.K. Mell, N. Bhandare, A comparison of dose-response characteristics of four NTCP models using outcomes of radiation-induced optic neuropathy and retinopathy, Radiat. Oncol. 6 (1) (2011) 1–10.
[19] S. Mostafaei, et al., CT imaging markers to improve radiation toxicity prediction in prostate cancer radiotherapy by stacking regression algorithm, La radiologia medica 125 (1) (2020) 87–97.
[20] W., Machine Learning to Predict Toxicity in Head and Neck Cancer Patients Treated with Definitive Chemoradiation, vol. 105, 2019.
[21] V. Kumar, et al., Radiomics: the process and the challenges, Magn. Reson. Imag. 30 (9) (2012) 1234–1248.
[22] P. Lambin, R.T.H.M. Larue, A.J.G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F.M. Mottaghy, J. E. Wildberger, S. Walsh, Radiomics: the bridge between medical imaging and personalized medicine, Nat. Rev. Clin. Oncol. 14 (12) (2017) 749–762.
[23] G. Valdes, et al., Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy, Phys. Med. Biol. 61 (16) (2016) 6105.
[24] S. Amiri, et al., A machine learning approach for prediction of auditory brain stem response in patients after head-and-neck radiation therapy, J. Cancer Res. Therapeut. 19 (5) (2023) 1219–1225.
[25] S. Amiri, et al., Radiomics analysis on CT images for prediction of radiation-induced kidney damage by machine learning models, Comput. Biol. Med. 133 (2021) 104409.
[26] C. Parmar, et al., Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer, Front. Oncol. 5 (2015) 272.
[27] S. Amiri, et al., Designing an intelligent lesion detection system using deep architecture neural networks in the lower limb X-ray images, Frontiers in Biomedical Technologies 10 (2) (2023) 169–176.
[28] J. Reddy, et al., Applying a machine learning approach to predict acute radiation toxicities for head and neck cancer patients, Int. J. Radiat. Oncol. Biol. Phys. 105 (1) (2019) S69.

[29] G. Guidi, et al., A machine learning tool for re-planning and adaptive RT: a multicenter cohort investigation, Phys. Med. 32 (12) (2016) 1659–1666.
[30] T. Wang, et al., Machine learning in quantitative PET: a review of attenuation correction and low-count image reconstruction methods, Phys. Med. 76 (2020) 294–306.
[31] M.I. Roldán, et al., Assessing the Visual Afferent Pathway with the Multifocal Visual Evoked Potentials in the Radiologically Isolated Syndrome, 2024.
[32] A. Calcagni, M.M. Neveu, N. Jurkute, A.G. Robson, Electrodiagnostic tests of the visual pathway and applications in neuro-ophthalmology, Eye (2024) 1–14.
[33] R. Hamilton, et al., VEP estimation of visual acuity: a systematic review, Doc. Ophthalmol. 142 (2021) 25–74.
[34] Y. Zhang, W. Ding, Z. Pan, J. Qin, Improved wavelet threshold for image de-noising, Front. Neurosci. 13 (2019) 39.
[35] A. Hatamizadeh, et al., Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images, in: International MICCAI Brainlesion Workshop, Springer, 2021.
[36] M.J. Cardoso, et al., Monai: an open-source framework for deep learning in healthcare, arXiv preprint arXiv:2211.02701 (2022).
[37] Y.-m. Zheng, et al., A computed tomography–based radiomics signature for predicting expression of programmed death ligand 1 in head and neck squamous cell carcinoma, Eur. Radiol. 32 (8) (2022) 5362–5370.
[38] E. Kapetanou, et al., Developing a radiomics atlas dataset of normal abdominal and pelvic computed tomography (RADAPT), Journal of Imaging Informatics in Medicine (2024) 1–9.
[39] W. Sun, et al., A CT-based radiomics nomogram for distinguishing between benign and malignant bone tumours, Cancer Imag. 21 (2021) 1–10.
[40] V. Moiseenko, W.Y. Song, L.K. Mell, N. Bhandare, A comparison of dose-response characteristics of four NTCP models using outcomes of radiation-induced optic neuropathy and retinopathy, Radiat. Oncol. 6 (2011) 1–10.
[41] X. Jin, et al., Prediction of response after chemoradiation for esophageal cancer using a combination of dosimetry and CT radiomics, Eur. Radiol. 29 (2019) 6080–6088.
[42] F.-X. Borruat, et al., Visual recovery from radiation-induced optic neuropathy. The role of hyperbaric oxygen therapy, J. Neuro Ophthalmol. 13 (2) (1993) 98–101.
[43] Y.-L. Wu, et al., Long-term evaluation and normal tissue complication probability (NTCP) models for predicting radiation-induced optic neuropathy after intensity-modulated radiation therapy (IMRT) for nasopharyngeal carcinoma: a large retrospective study in China, Journal of Oncology 2022 (1) (2022) 3647462.
[44] A. Köthe, et al., Assessment of radiation-induced optic neuropathy in a multi-institutional cohort of chordoma and chondrosarcoma patients treated with proton therapy, Cancers 13 (21) (2021) 5327.
[45] A.R. Carey, B.R. Page, N. Miller, Radiation-induced optic neuropathy: a review, Br. J. Ophthalmol. 107 (6) (2023) 743–749.
[46] S. Brecht, et al., Radiation-induced optic neuropathy after stereotactic and image guided intensity-modulated radiation therapy (IMRT), Radiother. Oncol. 134 (2019) 166–177.
[47] M.M. Ghiasi, S. Zendehboudi, Application of decision tree-based ensemble learning in the classification of breast cancer, Comput. Biol. Med. 128 (2021) 104089.
[48] W.-T. Wu, et al., Data mining in clinical big data: the frequently used databases, steps, and methodological models, Military Medical Research 8 (2021) 1–12.
[49] H. Li, et al., Decision tree model for predicting in-hospital cardiac arrest among patients admitted with acute coronary syndrome, Clin. Cardiol. 42 (11) (2019) 1087–1093.
[50] H. Li, et al., Radiomic score as a potential imaging biomarker for predicting survival in patients with cervical cancer, Front. Oncol. 11 (2021) 706043.
[51] L. Shi, et al., Radiomics for response and outcome assessment for non-small cell lung cancer, Technol. Cancer Res. Treat. 17 (2018) 1533033818782788.
[52] M.S. Bandara, et al., Ultrasound based radiomics features of chronic kidney disease, Acad. Radiol. 29 (2) (2022) 229–235.
[53] K. Bousabarah, et al., Radiomic analysis of planning computed tomograms for predicting radiation-induced lung injury and outcome in lung cancer patients treated with robotic stereotactic body radiation therapy, Strahlenther. Onkol. 195 (9) (2019) 830–842.
[54] R. Agheli, et al., Predicting severe radiation-induced oral mucositis in head and neck cancer patients using integrated baseline CT radiomic, dosimetry, and clinical features: a machine learning approach, Heliyon 10 (3) (2024) e24866.
[55] Z. Yang, et al., Machine-learning with region-level radiomic and dosimetric features for predicting radiotherapy-induced rectal toxicities in prostate cancer patients, Radiother. Oncol. 183 (2023) 109593.
[56] S.P. Krafft, et al., The utility of quantitative CT radiomics features for improved prediction of radiation pneumonitis, Med. Phys. 45 (11) (2018) 5317–5324.
[57] A. Haghbin, et al., Prediction of chronic kidney disease in abdominal cancers radiation therapy using the functional assays of normal tissue complication probability models, J. Cancer Res. Therapeut. 18 (3) (2022) 718–724.
[58] J.D. Shur, et al., Radiomics in oncology: a practical guide, Radiographics 41 (6) (2021) 1717–1732.
[59] V. Moiseenko, W. Song, L. Mell, N. Bhandare, MO-EE-A2-03: comparison of four NTCP models to describe dose-response for radiation-induced optic neuropathy and retinopathy, Med. Phys. 37 (6Part26) (2010), 3348-3348.