Radiomic-Assistant Response Prediction to Stereotactic Body Radiotherapy in Early Stage Lung Cancer

Melek YAKAR[1], Durmus ETIZ[1], Eyyup GULBANDILAR[2], Kerem DURUER[1], Ergin ERDEN[1]

[1] EskiSehir Osmangazi University, Faculty of Medicine, Department of Radiation Oncology

[2] EskiSehir Osmangazi University, Faculty of Engineering, Department of Computer Engineering

ABSTRACT

The aim of this study was to predict SBRT response in patients with early-stage lung cancer who underwent SBRT using 4-dimensional computed tomography (4DCT) radiomics. 44 cases diagnosed with early-stage lung cancer and treated with SBRT between 2020-2024 were included in the study. The radiomic features of the patients were obtained from the planning 4DCT with the Lifex program. The LASSO method was used to determine important variables. The SMOTE method was used to create a balanced data set. SBRT response estimate (complete response/partial response/stable response) was created using artificial intelligence methods using important variables. Median BED10 was 100 (min: 72, max: 132) Gy. SBRT scheme was applied as 8-12.5 Gy x 4-6 fr. Median PFS and OS after SBRT were 15 and 20 months at median 20-month follow-up. SBRT response assessment was performed using RECIST criteria. Complete, partial and stable response rates among patients were 36.4%, 36.4% and 27.3%, respectively. 7 of 55 radiomic features obtained with Lifex program were determined as significant variables with LASSO method. Prediction models were created with 5 different artificial intelligence algorithms using 7 significant variables. When the test groups are examined, SBRT response prediction was performed with 71%, 78%, 64%, 92% and 72% accuracy rates using MLPNN-1, MLPNN-2, ANFIS-1, ANFIS-2 and MLPC algorithms, respectively. Radiomics are easy to obtain, non-invasive and contains patient-specific information. However, the imaging method, segmentation differences between users, obtaining Radiomics and creating prediction algorithms are quite heterogeneous, and standardization should be provided with multi-center studies with more patients. Radiomics can be a potential biomarker in SBRT response prediction when these steps are standardized. In the current study, the highest accuracy rate was created with the ANFIS-2 algorithm and studies with more patients are needed.

Keywords: Early-stage lung cancer, Stereotactic body radiotherapy, Radiomics, treatment response prediction, Artificial intelligence

INTRODUCTION

Lung cancer is the most common cause of cancer-related deaths.[1] With the increased availability of screening methods and computed tomography, early-stage lung cancer is more frequently detected. The gold standard for treating early-stage lung cancer is surgery. However, in some cases, especially those with a long-term history of smoking, surgery cannot be performed due to insufficient lung and heart function. Sometimes patients do not accept surgery due to surgical complications and the risk of surgical mortality. For these reasons, the standard treatment for patients who cannot undergo surgery is stereotactic body radiotherapy (SBRT).[2] The 3-year local control rate with SBRT is 90%, comparable to surgery.[3] SBRT is a noninvasive radiotherapy technique that uses multiple small and precise beams of radiation to deliver high doses to tumors in extracranial areas in 1 to 5 fractions.[4]

Response rates are not always similar when the same dose of SBRT is applied to patients with the same tumor and patient characteristics. While some patients who receive similar treatment at a similar stage achieve a complete response, others may progress.

For this reason, personalized treatment is gaining importance in oncology. Methods that allow for more detailed evaluation of the tumor are needed for personalized treatments. The suffix "-omics" is widely used in biological disciplines to indicate the extraction of important information from a large dataset. Radiomics is the extraction of quantitative information from medical imaging data.[5] Radiomics is an emerging field of research that can extract features from medical images that the clinician cannot see or measure with the naked eye.These features can be used to create models for clinical outcomes, including information on diagnostic, prognostic or oncological outcomes and treatment toxicities.[6]

In radiological practice, imaging features other than size measurements are descriptive and qualitative. Quantitative features of imaging data can be better captured by radiomic features, which are higher-order measurements. Quantitative tumor information that is generally invisible to the human eye can be obtained with the help of radiomics.[7] Radiomic image features can increase the accuracy of tumor response prediction to SBRT by detecting quantitative features of the tumor that are not observed by a physician. In the current study, it was aimed to predict the treatment responses of 44 patients who underwent SBRT with early-stage lung cancer diagnosis in the Department of Radiation Oncology between 2020-2024 with radiomics.

PATIENTS AND METHODS

Patient Selection and Post-Treatment Follow-up

Between 2020 and 2024, 44 patients diagnosed with early-stage lung cancer with 4DCT simulations and who underwent SBRT were included in the study. Patients who were > 18 years old, had a KPS ≥ 60, completed their treatment as planned, and came to their follow-ups regularly for response evaluation and had the requested tests performed were included. In the early-stage lung cancer group, patients with T3 mediastinal region invasion, regional lymph node or distant metastasis detected, and previously treated with RT within the planned volume were not included in the study. Cases with hilar and/or mediastinal lymph nodes ≤ 1 cm and no involvement on Positron Emission Tomography-Computed Tomography (PET-CT) were considered N0, and cases with lymph node pathology negative for cases > 1 cm and/or abnormal involvement on PET-CT were also considered N0. Some cases underwent SBRT without tissue diagnosis due to complications secondary to biopsy/patient refusal of biopsy. In patients without tissue diagnosis, the tumor was considered malignant if it progressed on follow-up thorax CT and had radiological malignant characteristics and if there was involvement on FDG-PET-CT. Patients were evaluated at the Chest Diseases Oncology Council prior to SBRT. The council includes a radiation oncologist, a chest diseases specialist and a radiologist, and a thoracic surgeon's opinion is always sought for operability. The council evaluates imaging methods and, if available, biopsy and pathology-based staging, and a treatment decision is made. As recommended in the RTOG 0915 study, PET-CTs are requested within 8 weeks before SBRT.[8]

Follow-up was planned at 1st month after SBRT and then every 3 months for the first 3 years, every 6 months for the 4th-5th years and then annually. Thoracic CT was requested from the patients within 1-3 months after SBRT and PET CT was requested 3 months after RT, and their responses were evaluated multidisciplinary. PET-CTs were requested in the 1st year of follow-up as recommended in the RTOG 0915 study.[8]

Tumors were evaluated at each follow-up visit using the Response Evaluation Criteria in Solid Tumors, and the response was graded according to the criteria proposed in the Response Evaluation Criteria in Solid Tumors Guideline version 1.1.5.[9]

Treatment Planning

Patients were immobilized on a T-bar/Wingboard with their hands above their heads in the supine position. 1-2 mm thick images were obtained between the cricoid cartilage and the upper border of the L2 vertebra with GE Discovery RT CT® devices. An external respiratory tracking system (Real-time Position Management [RPM] System, Varian ® Medical Systems, Palo Alto, CA, USA) was used in 4D-CT. The RPM system uses an infrared tracking camera to determine the phases of the respiratory cycle. For all patients, 4DCT con-

Table 1. First and Second Order Radiomics

First Order	CONVENTIONAL_HUmin, CONVENTIONAL_HUmean, CONVENTIONAL_HUstd, CONVENTIONAL_Humax, CONVENTIONAL_HUSkewness, CONVENTIONAL_HUKurtosis, CONVENTIONAL_HUExcessKurtosis, DISCRETIZED_HUmin, DISCRETIZED_HUmean, DISCRETIZED_HUstd, DISCRETIZED_HUmax, DISCRETIZED_HUSkewness, DISCRETIZED_HUKurtosis, DIS CRETIZED_HUExcessKurtosis, DISCRETIZED_HISTO_Entropy_log10, DISCRETIZED_HISTO_Entropy_log2, DISCRETIZED_HISTO_Energy, DISCRETIZED_AUC_CSH, SHAPE_Volume (mL), SHAPE_Volume (vx), SHAPE_Sphericity, SHAPE_Surface (mm2), SHAPE_Compacity
Second Order	GLCM_Homogeneity, GLCM_Energy, GLCM_Contrast, GLCM_Correlation, GLCM_Entropy_log10, GLCM_Entropy_log2,GLCM_Dissimilarity, GLRLM_SRE, GLRLM_LRE, GLRLM_LGRE, GLRLM_HGRE, GLRLM_SRLGE, GLRLM_SRHGE, GLRLM_LRLGE, GLRLM_LRHGE, GLRLM_GLNU, GLRLM_RLNU, GLRLM_RP, NGLDM_Coarseness, NGLDM_Contrast, NGLDM_Busyness, GLZLM_SZE, GLZLM_LZE, GLZLM_LGZE, GLZLM_HGZE, GLZLM_SZLGE, GLZLM_SZHGE, GLZLM_LZLGE, GLZLM_LZHGE, GLZLM_GLNU, GLZLM_ZLNU, GLZLM_ZP

* GLCM: The grey level co-occurrence matrix

** GLRLM: The grey-level run length matrix

** GLRLM_SRE: Short-Run Emphasis

** GLRLM_LRE: Long-Run Emphasis

** GLRLM_LGRE: Low Gray-level Run Emphasis

** GLRLM_HGRE: High Gray-level Run Emphasis

** GLRLM_SRLGE: Short-Run Low Gray-level Emphasis

** GLRLM_SRHGE: Short-Run High Gray-level Emphasis

** GLRLM_LRLGE: Long-Run Low Gray-level Emphasis

** GLRLM_LRHGE: Long-Run High Gray-level Emphasis

** GLRLM_GLNUr: Gray-Level Non-Uniformity for run

** GLRLM_RLNU: Run Length Non-Uniformity

** GLRLM_RP: Run Percentage

***NGLDM: The neighborhood grey-level difference matrix

****GLZLM: The grey-level zone length matrix

**** GLZLM_SZE: Short-Zone Emphasis

**** GLZLM_LZE: Long-Zone Emphasis

**** GLZLM_LGZE: Low Gray-level Zone Emphasis

**** GLZLM_HGZE: High Gray-level Zone Emphasis

**** GLZLM_SZLGE: Short-Zone Low Gray-level Emphasis

****GLZLM_SZHGE: Short-Zone High Gray-level Emphasis

****GLZLM_LZLGE: Long-Zone Low Gray-level Emphasis

****GLZLM_LZHGE: Long-Zone High Gray-level Emphasis

****GLZLM_GLNUz: Gray-Level Non-Uniformity for zone

****GLZLM_ZLNU: Zone Length Non-Uniformity

****GLZLM_ZP: Zone Percentage

sisting of 10 respiratory phases was obtained and these 10 phases were determined as a percentage of the respiratory cycle. For the internal target volume (ITV), the clinician first determined in which phases the treatment would be performed and MIP was obtained over these phases. The gross tumor volume (GTV) was contoured over the 4DCT MIP and the ITV was obtained by merging these volumes.

A 3-5 mm isotropic margin was given around the ITV to create the planning target volume (PTV). The patients were treated with 6 MV-FFF energy and partial volumetric arc therapy (VMAT). All patients received SBRT with the Varian TrueBeam® linear accelerator. Based on the RTOG 0915 study as the planning criterion, planning was made in such a way that at least 95% of the PTV would receive the prescribed dose.[8] Varian Real-Time Position Management was used to monitor the patient's breathing.

Different fractionation schemes were used depending on the location of the tumor. In ultracentral tumors, regimens with lower BED10 values such as 8-9 Gy x 5-6 fractions were preferred, while in central and peripheral tumors, regimens with higher BED10 values of 10-12 Gy x 5 fractions were preferred. Control was performed with 3D-CBCT images before and after SBRT application in each fraction.

No concomitant chemotherapy/immunotherapy or targeted therapies were administered during SBRT. There was a minimum of 48 hours between two fractions. Patients were evaluated twice weekly in the outpatient clinic for possible toxicity, their symptoms were questioned, and physical examinations were performed.

Radiomic Feature Extraction

Radiomic features were obtained using computed tomography (CT) images taken for RT simulation purposes at Eskişehir Osmangazi University Faculty of Medicine, Department of Radiation Oncology. The GTV region of interest (ROI) information was obtained from the RT treatment planning system (Varian Eclipse model (Version: 15.6), Varian ® Medical Systems, Palo Alto, CA, USA). The LIFEx (Version: 7.0.0) program was used to extract radiomic features. No preprocessing was performed while extracting the radiomic features. Spatial resampling was normalized to 1 mm x 1 mm x 1 mm, with a gray level number of 400 and a bin width of 10. With these parameters set, a three-dimensional (3D) processing was performed, and a total of 55 radiomic features, including first-order and second-order features, were extracted (Table 1).

Statistical Analysis and Artificial Intelligence

Statistical Analysis and Radiomics Feature Selection

A Least Absolute Shrinkage and Selection Operator (LASSO) logistic regression model with a binomial family and 10-fold cross-validation was employed to identify the most predictive radiomic features. LASSO is particularly advantageous in high-dimensional datasets where multicollinearity and the presence of many irrelevant features can complicate traditional regression methods. By applying an L1 penalty, LASSO effectively shrinks the coefficients of less relevant features toward zero, leaving only those variables with the most predictive power. This helps mitigate the overfitting problem and improves the interpretability of the model by selecting a parsimonious set of features.

The LASSO model was implemented using the glmnet package in R. We performed 10-fold cross-validation to determine the optimal lambda value, which controls the strength of the L1 penalty. The lambda value that minimized the cross-validated deviance was selected to balance model complexity and prediction accuracy. Cross-validation splits were performed on the training data, ensuring that the model was not overfitted to any particular fold, thus enhancing the generalizability of the results.

For each feature, coefficients were calculated, and those with non-zero values were considered as significant predictors. The final model included significant radiomic features, which were robustly associated with the treatment response based on their respective non-zero coefficients.[10-12]

Artificial Intelligence-Based Prediction Algorithms

Data and Data Preparation

A data table was obtained with 55 columns consisting of the features and the number of 51 rows which is determined with the tumor of patients. Tumor response was also added as an output variable to this data table. The output variable was divided into three classes: Complete, partial and stable response.

Developing Models

Since the distribution of the numbers of these classes in the output variable is imbalanced, the number of the data was increased and the class distribution was balanced with the Synthetic Minority Oversampling Technique (SMOTE) algorithm by using Python.[13] Finally, the data was randomly divided into 80% training and 20% test groups, using the code prepared in Python programming language.

Developing models are planned under three main headings, namely artificial neural networks (Multiple-layer perceptron neural network-MLPNN), an adaptive network-based fuzzy inference system (ANFIS) and Multi-layer Perceptron (MLP) Classifier. MLPNN and ANFIS models were developed using MatLab Toolbox, while MLPClassifier networks were developed using Python programming language. A randomly allocated training data-

set was used to develop the models and then the trained models were tested with test data.

Multiple-layer Perceptron Neural Network (MLPNN)

MLPNN models were developed and trained in two different types. In the first model training process, there are three layers: input, hidden and output layer. There are 10 neurons in the hidden layer. The network type is "feed-forward backpropagation", the training function is "trainlm", the adaptation learning function is "learngdm", the performance function is Mean squared error (MSE) and the transfer function is "tansig".

Accuracy, recall, precision and F1 score were used to evaluate the performance results of the models. Recall is defined as the ratio of the total number of correctly classified positive classes divided by the total number of positive classes. In other words, it defines how many of the positive classes are correctly predicted. Also, precision is defined as the ratio of the total number of correctly classified positive classes divided by the total number of predicted positive classes. F1 score is a number between 0 and 1 and is the harmonic mean of precision and recall.[14]

In the second model training process, there is an input layer, two hidden layer and an output layer. There are 10 neurons in first the hidden layer and 5 neurons in the second hidden layer. Similarly, the network type is "feed-forward backpropagation", the training function is "trainlm", the adaptation learning function is "learngdm", the performance function is Mean squared error (MSE) and the transfer function.

Adaptive Network-based Fuzzy Inference System (ANFIS)

Neuro-fuzzy systems are the combination of the parallel computation and learning capabilities of artificial neural networks with the ability to derive conclusions using expert knowledge of fuzzy logic. As a result, artificial neural networks become more understandable thanks to neuro-fuzzy systems. ANFIS can assign all possible rules according to the structure created for the problem to be solved or allows the assignment of rules with the help of data. ANFIS's rule creation or rule creation allows it to benefit from the opinions it receives. Therefore, since it allows artificial neural networks to benefit from expert opinions in many prediction problems, it provides better results according to the mean square error criterion. ANFIS's learning algorithm is a hybrid learning algorithm consisting of the combination of the least squares method and the backpropagation learning algorithm.[15]

Similarly, we designed two different models in ANFIS. In developing first model (ANFIS-1), seven input and one output variables were used. Triangular membership function was preferred in the process of fuzzification the input variables. Other input variables are also defined with similar triangular membership functions and three membership functions are used to fuzzification the crisp values. The developing ANFIS-1 model continued training until it reached root mean square error (RMSE) = 0.00908 with 30 epochs.

In the ANFIS second model (ANFIS-) training process, there are the same seven input variables and one output variable. But in the AFIS-2 model, the trapezoidal membership function was preferred in the fuzzification process of the seven input variables. Similarly, for the other input variables, three trapezoidal membership functions are used to fuzzification the crisp values. The developing ANFIS-2 model continued training until it reached RMSE = 0.198964 with 30 epochs.

MLP Classifier

To create the fifth model using the same dataset, a deep classifier algorithm, MLPClassifier, was used. This model was designed using libraries in the Python programming language. Three hidden layers were preferred in MLPClassifier model. The number of neurons in the hidden layers were 4, 7 and 8. "Tanh" was preferred as the activation function. 'lbfgs' algorithm was preferred to optimize the MLPClassifier model. Solver is used for classification problems and uses the limited memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm for optimizing the weights. As another parameter, the maximum iteration number was selected as 500.

Table 2. Patient, tumor characteristics and oncological outcomes

Features	N (%) / Median (Minimum, Maksimum)
Age	69 (52-86)
KPS	70 (70-80)
Gender	
Female	
Male	36 (82%)
Tumor Location	
Right lower lobe	8 (18%)
Right middle lobe	1 (2%)
Right upper lobe	14 (32%)
Left lower lobe	10 (23%)
Left upper lobe	11 (25%)
Tumor size (cm)	2.35 (0.70-6.40)
GTV (cc)	10.65 (0.90-50.90)
PTV(cc)	28.85 (6-88.20)
Fraction dose (Gy)	10 (8-12.5)
Number of fractions	5 (4-6)
BED10	100 (72-132)
RECIST 1.1	
Complete response	16 (36%)
Partial response	16 (36%)
Stable response	12 (28%)
Current Status	
Alive	33 (75%)
Ex	11 (25%)
OS (ay)	20 (4-54)
PFS (ay)	15 (4-51)

KPS: Karnofsky performance status, GTV:Gross tumor volume, PTV: planning target volume, Gy:Gray, BED: Biologically effective dose, OS: Overall survival, PFS: Progression-free survival

Testing the Models

We applied our data set aside for testing to developed five different models whose training process was completed.

Ethic Committee Permission: Permission was obtained from Osmangazi University Non-Interventional Clinical Research Ethics Committee (Acceptance date: 26 September 2023, Decision Number: 07).

RESULTS

Patient, Tumor Characteristics, and Oncological Outcomes

The study included patients who were planned to receive 4DCT-guided SBRT, completed their treatments as planned, and attended regular follow-ups. The median age was 69 years, 8 (18.2%) of the patients were female and 36 (81.8%) were male. The median tumor size was 2.3 cm, and the median GTV and PTV volumes were 10.6 cc and 28.8 cc, respectively. The SBRT dose was 8-12.5 Gy x 4-6 fractions. BED10 was determined as $\alpha/\beta=10$ and was calculated with the formula $D(1 + d/[\alpha/\beta])$. The median BED10 value was 100 (min: 72, max: 132) Gy. During the median follow-up of 20 months, 33 (75%) patients were alive and 11 (25%) patients had died. Of the patients who died during follow-up, 4 died due to cancer progression, 6 died due to non-cancer reasons, and 1 died due to systemic treatment toxicity due to disease progression.

Median OS and PFS after SBRT were 20 and 15 months, respectively. RECIST assessment was performed with thorax CT 3 months after the end of SBRT, and there were 16 (36.4%) complete responses (CR), 16 (36.4%) partial responses(PR), and 12 (27.3%) stable responses(SR). Response assessment based on BED10 values showed that among the six patients with a BED10 value below 100 Gy, 2 (33.3%) achieved CR, 3 (50.0%) had PR, and 1 (16.6%) had SR. Among the 38 patients with a BED10 value of 100 Gy or higher, 14 (36.8%) achieved CR, 13 (34.2%) had PR, and 11 (28.9%) had SR.Patient, tumor characteristics, and oncologic outcomes are given in Table 2.

Important Radomic Features

A total of 7 out of 55 radiomic features were determined as important variables with the LASSO method. Four of these variables are first order and three are second order radiomic features. The first order radiomics determined as significant variables are 'CONVENTIONAL_HUmin', 'CONVENTIONAL_HUKurtosis', 'DISCRETIZED_HUKurtosis', 'DISCRETIZED_HUExcessKurtosis', while the second order radiomics are 'Gray Level

Table 3. The confusion matrix of training group models

MLPNN-1		**Predicted classes**			**Total**
		1 (Complete)	2 (Partial)	3 (Stable)	
True classes	1 (Complete)	**21**	4	0	25
	2 (Partial)	1	**17**	2	20
	3 (Stable)	0	1	**10**	11
Total		22	22	12	**56**

MLPNN-2		**Predicted classes**			**Total**
		1 (Complete)	2 (Partial)	3 (Stable)	
True classes	1 (Complete)	**24**	1	0	25
	2 (Partial)	6	**12**	2	20
	3 (Stable)	0	0	**11**	11
Total		30	13	13	**56**

ANFIS-1		**Predicted classes**			**Total**
	1 (Complete)	2 (Partial)	3 (Stable)		
True classes	1 (Complete)	**25**	0	0	25
	2 (Partial)	0	**20**	0	20
	3 (Stable)	0	0	**11**	11
Total		25	20	11	**56**

ANFIS-2		**Predicted classes**			**Total**
	1 (Complete)	2 (Partial)	3 (Stable)		
True classes	1 (Complete)	**20**	5	0	25
	2 (Partial)	1	**19**	0	20
	3 (Stable)	0	3	**8**	11
Total		21	27	8	56

MLP Classifier		**Predicted classes**			**Total**
	1 (Complete)	2 (Partial)	3 (Stable)		
True classes	1 (Complete)	**24**	1	0	25
	2 (Partial)	0	**20**	0	20
	3 (Stable)	0	0	**11**	11
Total		24	21	11	**56**

Co-occurrence Matrix (GLCM)_Correlation', 'Neighborhood Gray-Level Dependence Matrix (NGLDM)_Coarseness' and 'NGLDM_Busyness'.

Results of AI-Based Prediction Algorithms

The confusion matrix showing the training results of the first model developed (MLPNN-1) is given in Table 3. The diagonal cells in this table show correct predictions, while the other cells show incorrect predictions. In general, as a result of the training of the developed MLPNN-1 model, 48 out of 56 data were predicted correctly (85.7%). The training performance results of our MLPNN-1 model are shown in Table 4.

When the training performance of the MLPNN-1 model is examined in terms of its success in correctly classifying classes, that is, the recall value reveals high accuracy values such as 0.840 for 1 (complete response), 0.850 for 2 (partial response) and 0.909 for 3 (stable response). The mean recall value is 0.866.

Table 4. The training performance results of developed models

MLPNN-1						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Complete)	0.840	0.866	0.955	0.854	0.894	0.858
2 (Partial)	0.850		0.772		0.810	
3 (Stable)	0.909		0.833		0.870	
Accuracy	0.857					
MLPNN-2						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Complete)	0.960	0.853	0.800	0.856	0.872	0.839
2 (Partial)	0.600		0.923		0.727	
3 (Stable)	1.000		0.846		0.917	
Accuracy	0.839					
ANFIS-1						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Complete)	1.000	1.000	1.000	1.000	1.000	1.000
2 (Partial)	1.000		1.000		1.000	
3 (Stable)	1.000		1.000		1.000	
Accuracy	1.000					
ANFIS-2						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Complete)	0.800	0.826	0.953	0.885	0.870	0.840
2 (Partial)	0.950		0.703		0.806	
3 (Stable)	0.727		1.000		0.842	
Accuracy	0.839					
MLP Classifier						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Complete)	0.960	0.987	1.000	0.984	0.980	0.985
2 (Partial)	1.000		0.952		0.976	
3 (Stable)	1.000		1.000		1.000	
Accuracy	1.000					

The precision values for 1 (complete response), 2 (partial response) and 3 (stable response) which is all the predictive positive classes are 0.955, 0.772 and 0.833, respectively. This reveals a very satisfactory result for the developed model.

When the F1-score, which is the success measurement criterion of the developed model, is examined, a similar high success is seen (values range from 0.810 to 0.894).

In order to determine the training performance results of the developed MLPNN-2 model for each class in detail, the confusion matrix shown in Table 3. The F1-score value, which is a generally accepted metric, has a good performance value of 0.839 for MLPNN-2. When Table 4 is examined in detail, similar success results to the success of the MLPNN-1 model in class prediction are also seen in MLPNN-2.The training results of the second model developed (MLPNN-2) are shown in Table 4.

The confusion matrix shown in Table 3 was created in order to determine the training performance results of the developed ANFIS-1 model for each class in detail.The training results of the developed ANFIS-1 model are shown in Table 4. In this table, the diagonal cells show the correct predictions, which shows that 56 out of 56 data were predicted correctly (100%) for the training result of the developed ANFIS-1 model. The accuracy value, which is a generally accepted metric, has a good performance value of 1.000 for ANFIS-1. It is seen that the accuracy, precision and F1 score values of the ANFIS-1 model are above 1.000 for each class.

The training results of the developed ANFIS-2

Table 5. The confusion matrix of test group models

MLPNN-1		**Predicted classes**			**Total**
		1 (Complete)	2 (Partial)	3 (Stable)	
True classes	1 (Complete)	1	0	0	1
	2 (Partial)	2	7	0	9
	3 (Stable)	0	2	2	4
Total		3	9	2	14
MLPNN-2		**Predicted classes**			**Total**
		1 (Complete)	2 (Partial)	3 (Stable)	
True classes	1 (Complete)	1	0	0	1
	2 (Partial)	0	8	1	9
	3 (Stable)	0	2	2	4
Total		1	10	3	14
ANFIS-1		**Predicted classes**			**Total**
	1 (Complete)	2 (Partial)	3 (Stable)		
True classes	1 (Complete)	1	0	0	1
	2 (Partial)	3	5	1	9
	3 (Stable)	1	0	3	4
Total		5	5	4	14
ANFIS-2		**Predicted classes**			**Total**
	1 (Complete)	2 (Partial)	3 (Stable)		
True classes	1 (Complete)	1	0	0	1
	2 (Partial)	0	9	0	9
	3 (Stable)	0	1	3	4
Total		1	10	3	14
MLP Classifier		**Predicted classes**			**Total**
	1 (Complete)	2 (Partial)	3 (Stable)		
True classes	1 (Complete)	1	0	0	1
	2 (Partial)	1	8	0	9
	3 (Stable)	0	2	2	4
Total		2	10	2	14

model are summarized in Table 3. When the ratio of the sum of the diagonal cells in Table 4. It is determined that the accuracy, precision and F1 score values of the ANFIS-2 model for each class are above 0.8.

The training results of the developed MLP Classifier model are shown in Table 3. In this table, the diagonal cells show the correct predictions, and it is seen that 55 out of 56 data were predicted correctly (98.21%) for the training result of the developed MLP Classifier model. In Table 4 was created to determine the training performance results of the developed MLP Classifier model for each class in detail. The accuracy value, which is a generally accepted metric, has a good performance value of 0.9821 for the MLP Classifier. The average values of accuracy, recall, precision and F1 score of the MLP Classifier model are 0.982, 0.987, 0.984 and 0.985, respectively.

The confusion matrices of the algorithms created with post-training test data are given in Table-5 and the performance scores of the algorithms are given in Table 6.

Table 6. The testing performance results of developed models

MLPNN-1						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Progressive)	1.000	0.760	0.333	0.704	0.500	0.648
2 (Regressive)	0.778		0.777		0.778	
3 (Stable)	0.500		1.000		0.667	
Accuracy	0.714					
MLPNN-2						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Progressive)	1.000	0.796	1.000	0.822	1.000	0.805
2 (Regressive)	0.889		0.800		0.842	
3 (Stable)	0.500		0.667		0.571	
Accuracy	0.786					
ANFIS-1						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Progressive)	1.000	0.769	0.200	0.650	0.333	0.599
2 (Regressive)	0.556		1.000		0.714	
3 (Stable)	0.750		0.750		0.750	
Accuracy	0.643					
ANFIS-2						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Progressive)	1.000	0.917	1.000	0.967	1.000	0.935
2 (Regressive)	1.000		0.900		0.947	
3 (Stable)	0.750		1.000		0.857	
Accuracy	0.929					
MLP Classifier						
Classes	**Recalls**	**Mean recall**	**Precisions**	**Mean precision**	**F1-scores**	**Mean F1-score**
1 (Progressive)	1	0.796	0.500	0.767	0.667	0.725
2 (Regressive)	0.889		0.800		0.842	
3 (Stable)	0.500		1.000		0.667	
Accuracy	0.725					

In the test group, the F1 scores of MLPNN-1, MLPNN-2, ANFIS-1, ANFIS-2 and MLP Classifier algorithms are 0.714, 0.786, 0.643, 0.929 and 0.725, respectively. The highest accuracy rate was achieved with the ANFIS-2 model. 14 cases were used in the test group of these models, and ANFIS-2 correctly predicted one case with complete response, all nine cases with partial response and three of four cases with stable response, and the mean recall was 0.917, precision was 0.967 and mean F1 score was 0.935, while the accuracy rate was 0.929, and it was determined as the best prediction algorithm in this patient group.

DISCUSSION

Imaging methods are an integral part of cancer treatment. They guide the clinician in both staging, treatment planning and post-treatment disease follow-up. However, the fact that similar results cannot be obtained in every patient with the same treatment at the same stage directs the clinician to personalized treatments. The personalized treatment decision is quite complex. It is known that many other factors that we do not yet know, from environmental and genetic factors to nutritional habits, can also change tumor behavior. Radiomics are promising as a guiding biomarker in this regard. If radiosensitive or radioresistant tumors can be determined before treatment, treatment strate-

gies can be changed accordingly. Radiomics is still in the study phase and is not used in routine clinical practice. However, if standard radiomics that show radioresistance and the most accurate algorithms can be found, we can be one step closer to personalized treatment. The -omic suffix emerged from molecular biology to describe the characterization of biological molecules (such as DNA and genomics, proteins and proteomics). Today, it is also used for research areas that produce complex and high-dimensional information from some data.[16] Radiomics can be thought of as signatures or fingerprints of tumors.

In the current pilot study, 7 important radiomic features were found in RT planning 4DCT for predicting SBRT response. These radiomics include both first order and second order features. First order features show the mean, median, maximum, minimum values of voxel intensities independent of spatial relationship and features such as asymmetry, kurtosis, entropy, and uniformity. Second order features show the relationships between neighboring voxels.[17] The first order features found to be important in the current study are 'CONVENTIONAL_HUmin', 'CONVENTIONAL_HUKurtosis', 'DISCRETIZED_HUKurtosis', 'DISCRETIZED_HUExcessKurtosis', while second order radiomics are 'GLCM_Correlation', 'NGLDM_Coarseness' and 'NGLDM_Busyness'.

Luo et al. created 3 models as radiomic model, clinical model and combined model in patients diagnosed with lung cancer who underwent SBRT and evaluated all 3 models. They extracted 1502 radiomic features from CT images taken before SBRT and 4 important radiomic features were found with LASSO method. These 4 important variables were found as wavelet-LLL_glszm_SmallAreaEmphasis, wavelet-LHH_glcm_JointAverage, wavelet-LHH_ngtdm_Complexity, and squareroot_glcm_DifferenceEntropy. As a result of multivariate analysis of clinical parameters, clinical stage, platelet count and minimum dose received by GTV were determined as important variables. Logistic regression method was used to create the model. AUC (area under curve) values of radiomic model, clinical model and combined model were 0.811, 0.845 and 0.911, respectively and the combined model with the highest performance was accepted. According to this study, a combined model based on radiological features, clinical and dosimetric parameters can improve the prediction of 1-year local control in lung cancer patients undergoing SBRT.[18] Although only lung cancer cases were included in our study, there were both early-stage, locally advanced and metastatic cases, and it is a heterogeneous group, and combined models were used. In our study, only early-stage lung cancer cases were included, but a single model (radiomic model) was used. Although different patient groups and different models were used, the accuracy rates were high in both the study of Luo et al. and our study.

Cilla et al. evaluated 80 lesions of 56 patients with lung oligometastasis who underwent SBRT. GTV was contoured on the CT scan taken before SBRT and 107 radiomic features were extracted from these GTVs. Four of the 107 radiomics were found to be significant radiomics. These radiomics were 'surface to volume ratio', 'the skewness', 'the correlation' and 'the grey normalized level uniformity'. No statistically significant relationship was found between the response in clinical parameters. Logistic regression (LR) and the classification and regression tree analysis (CART) models were used. Area under the curve for CR prediction was 0.707 and 0.753 for the LR and CART models, respectively.[19]

In another study conducted with both early stage and oligometastasis, 85 tumors and 69 patients were evaluated. GTV was contoured from CT images taken before SBRT and the aim was to predict the response to SBRT. 110 radiomic features were extracted. Support vector machine (SVM) was used to create the model. AUC was used in model performance evaluation. Skewness and root mean squared were identified as radiomic predictors of response to SBRT. The accuracy rate of the SVM machine learning model in prediction was 74.8%. AUCs in the prediction of complete response, partial response and non-response cases were 0.86, 0.94 and 0.85, respectively.[20] This study was conducted with a heterogeneous group and only the radiomic model was evaluated.

Kurtosis is a measure of the 'peakedness' of the distribution of values in the ROI. GLCM_Correla-

tion is Linear dependency of gray level values to their respective voxels in the GLCM.[21]

In a study evaluating SBRT-related lung toxicity and oncological outcomes with radiomics, GLCM-correlation radiomic was found to be associated with local recurrence and death. Tumors with high kurtosis radiomic were associated with worse overall survival.[22] In our study, kurtosis and GLCM_Correlation are among the important radiomics affecting the SBRT response. There are also other studies showing the relationship between radiomics including kurtosis features and disease-free survival.[23] In the study conducted by Yu et al. with patients diagnosed with stage 1 NSCLC who underwent SBRT, it was planned to evaluate oncological outcomes. 147 patients were included in the training cohort and 295 cases were included in the validation cohort, and 12 radiomic features were extracted from the CT images obtained before SBRT. Two of these 12 radiomics were determined as important variables, namely kurtosis and the GLCM_ homogeneity2. These radiomics were found to be associated with overall survival, distant metastasis and borderline regional recurrence (24). In the study conducted by Mattonen et al., they made early prediction of local recurrence in CT scans taken after SBRT in 45 early-stage lung cancer cases. Of the 44 radiomic features, 5 radiomics were determined as significant variables, namely GLCM_homogeneity, GLCM_correlation, GLCM_energy and grey_level uniformity.[25]

NGLDM is based on the gray level relationship between a pixel or voxel considered as the center and its neighbors. NGLDM features reflect the similarity in gray levels and gray level dependencies across a ROI, with a large dependency emphasis and a small dependency emphasis reflecting heterogeneity and homogeneity, respectively.[26] In another study conducted with 102 early stage NSCLC cases treated with SBRT, nodal relapse and disease-free survival were estimated. GTV was contoured using RT simulation 4DCT and 45 radiomic features were obtained. Six of the 45 radiomic features were determined as significant variables, namely Shape_Sphericity, Shape_Compacity, Histo_Energy, GLRLM_RLNU, GLRLM_LRE, and NGLDM_Coarseness. NGLDM_Coarseness, which was considered an important variable in progression-free survival in this study, was also found to be a significant variable in SBRT response estimation in our study.[27] NGLDM_Coarseness measures the spatial rate of change in intensity, so it can be thought that the response rate and PFS increase as a result of the change in tumor intensity.

Radiomics are quite promising in terms of personalized treatments in cancer treatment. However, standardization is required for their use in routine clinical practice. In the studies conducted, different CTs (3DCT, 4DCT, contrast +/-, different phases of 4DCT, etc.) are used to extract radiomic features, this heterogeneous approach needs to be standardized. There may be differences between clinicians in the segmentation phase of GTV, which may prevent finding the right radiomic. With a highly accurate artificial intelligence-based application accompanied by guides, contours without individual differences may be obtained in the future. In the radiomic extraction phase, there are many different commercial software or artificial intelligence methods, and important variables are evaluated with different artificial intelligence algorithms. All of these create heterogeneity in predicting oncological outcomes. In order to be able to move into clinical routine, these stages must first be standardized.

In our current study, GTV was determined by two different radiation oncologists with at least 10 years of experience via 4DCT. The limitations of this study are the small number of patients and the cohort was created with patients from a single center. 55 radiomics were extracted with the Lifex program and only pre-treatment radiomics were evaluated. Only radiomics belonging to GTV were evaluated and peritumoral radiomics were not evaluated. Considering that peritumoral immune response also affects oncological results, peritumoral radiomics may also be important. In addition, a model was created only with radiomics without including clinical parameters.

Conclusion

In oncological treatments, finding the most effective treatment with the least side effects and achieving personalized treatments with a good therapeutic index are very important in terms of

oncological results and toxicity. When radiomics are considered as a signature of the tumor, it promises hope for correct and effective treatment decisions. However, there are many steps that need to be standardized and more patient numbers and multicenter studies are needed.

Acknowledgments:

We would like to thank Professor Doctor Ertuğrul Çolak from Eskişehir Osmangazi University Faculty of Medicine, Department of Biostatistics, for his statistical evaluation support.

REFERENCES

1. Bray F, Laversanne M, Weiderpass E, Soerjomataram I. The ever-increasing importance of cancer as a leading cause of premature death worldwide. Cancer 127: 3029-3030, 2021.
2. Schneider BJ, Daly ME, Kennedy EB, et al. Stereotactic body radiotherapy for aerly-stage non-small-cell lung cancer: American Society of Clinical Oncology Endorsement of the American Society for Radiation Oncology Evidence-Based Guideline. J Clin Oncol 36: 710-719, 2018.
3. Chang JY, Senan S, Paul MA, et al. Stereotactic ablative radiotherapy versus lobectomy for operable stage I non-small-cell lung cancer: a pooled analysis of two randomised trials. Lancet Oncol 16(6): 630-637, 2015. Epub 2015 May 13. Erratum in: Lancet Oncol 16: e427, 2015.
4. Timmerman RD, Kavanagh BD, Cho LC, et al. Stereotactic body radiation therapy in multiple organ sites. J Clin Oncol 25: 947-952, 2007.
5. Chen M, Copley SJ, Viola P, et al. Radiomics and artificial intelligence for precision medicine in lung cancer treatment. Semin Cancer Biol 93: 97-113, 2023.
6. Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 5: 4006, 2014.
7. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14: 749-762, 2017.
8. Videtic MMG, Hu C, Singh AK, et al. NRG Oncology RTOG 0915 (NCCTG N0927): A randomized phase II study comparing 2 stereotactic body radiation therapy (SBRT) schedules for medically inoperable patients with stage I peripheral non-small cell lung cancer. Int J Radiat Oncol Biol Phys 93: 757-764, 2015.
9. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version1.1). Eur J Cancer 45: 228-247, 2009.
10. R Core Team (2022) R. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. URL https://www.R-project.org/.
11. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw 33: 1-22, 2010.
12. Lambin P, Leijenaar RTH, Deist TM, et al. "Radiomics: The bridge between medical imaging and personalized medicine." Nat Rev Clin Oncol 14: 749-762, 2017.
13. Dablain D, Krawczyk B, Chawla NV. DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data. IEEE Trans Neural Netw Learn Syst 34: 6390-6404, 2023.
14. Powers D. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. J Mach Learn Technol 2: 2229-3981, 2011.
15. JSR Jang. ANFIS: Adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern B Cybern 23: 665-685, 1993.
16. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 278: 563-77, 2016.
17. Rizzo S, Botta F, Raimondi S, et al. Radiomics: the facts and the challenges of image analysis. Eur Radiol Exp 2: 36, 2018.
18. Luo LM, Huang BT, Chen CZ, et al. A combined model to Improve the prediction of local control for lung cancer patients undergoing stereotactic body radiotherapy based on radiomic signature plus clinical and dosimetric parameters. Front Oncol 11: 819047, 2022.
19. Cilla S, Pistilli D, Romano C, et al. CT-based radiomics prediction of complete response after stereotactic body radiation therapy for patients with lung metastases. Strahlenther Onkol 199: 676-685, 2023.
20. Cheung BMF, Lau KS, Lee VHF, et al. Computed tomography-based radiomic model predicts radiological response following stereotactic body radiation therapy in early-stage non-small-cell lung cancer and pulmonary oligo-metastases. Radiat Oncol J. 2021 Dec;39(4):254-264.
21. Scapicchio C, Gabelloni M, Barucci A, et al. A deep look into radiomics. Radiol Med 126: 1296-1311, 2021.
22. Bousabarah K, Temming S, Hoevels M, 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: 830-842, 2019.
23. Huang Y, Liu Z, He L, et al. Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. Radiology 281: 947-957, 2016.
24. Yu W, Tang C, Hobbs BP, et al. Development and validation of a predictive radiomics model for clinical outcomes in stage I non-small cell lung cancer. Int J Radiat Oncol Biol Phys 102: 1090-1097, 2018.

25. Mattonen SA, Palma DA, Johnson C, et al. Detection of local cancer recurrence after stereotactic ablative radiation therapy for lung cancer: Physician performance versus radiomic assessment. Int J Radiat Oncol Biol Phys 94: 1121-1128, 2016.

26. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J Nucl Med 61: 488-495, 2020.

27. Franceschini D, Cozzi L, De Rose F, et al. A radiomic approach to predicting nodal relapse and disease-specific survival in patients treated with stereotactic body radiation therapy for early-stage non-small cell lung cancer. Strahlenther Onkol 196: 922-931, 2020.

Correspondence:

Dr. Melek YAKAR

Eskisehir Osmangazi Universitesi Tip Fakultesi

Radyasyon Onkolojisi Anabilim Dali

Odunpazari

ESKISEHIR / TURKIYE

Tel : (+90-538) 391 61 90

e-mail: myakar@ogu.edu.tr

ORCIDs:

Melek Yakar	0000-0002-9042-9489
Durmus Etiz	0000-0002-9042-9489
Eyyüp Gülbandilar	0000-0001-5559-5281
Kerem Duruer	0000-0003-2303-4070
Ergin Erden	0009-0003-4661-9626