

Machine Learning-Based Radiomic Signatures for Predicting Immunotherapy Response in Advanced Non-Small Cell Lung Cancer

E.M.J. Karthikeyan^{1,*}, Manoranjan Parhi², P. Grace Kanmani Prince³, Arvindkumar M. Chauhan⁴, Shitij Goyal⁵ and Prerak Sudan⁶

¹Department of General Surgery, Vinayaka Mission's Kirupananda Variyar Medical College & Hospitals; Vinayaka Mission's Research Foundation (DU), Salem, Tamil Nadu, India

²Department of Center for Data Science, IMS and SUM Hospital, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India

³Department of Biomedical Engineering, Sathyabama Institute of Science and Technology, Chennai, Tamil Nadu, India

⁴Faculty of Allied and Healthcare, Gokul Global University, Sidhpur, Gujarat, India

⁵Quantum University Research Center, Quantum University, Roorkee, Uttarakhand, 247667, India

⁶Center for Research Impact & Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India

Abstract: Immunotherapy has brought about many advances in treatment options for patients with advanced non-small cell lung cancer (NSCLC), although there is great variability among patients' responses to the treatment; thus, there is a need for accurate predictive methods. This study seeks to analyze the predictive power of machine learning models using computed tomography (CT)-based radiomic features to identify responders to immunotherapy in patients with metastatic NSCLC. The use of pretreatment CT images and responses to treatment was analyzed in a retrospective cohort of 150 patients with advanced non-small cell lung cancer. The responders were those who demonstrated full or partial effectiveness of treatment, while the non-responders were the patients who had cancer stabilisation or progressive illness. Tumor segmentation was performed, and radiomic features of intensity, shape, texture, and higher-order features were extracted. Feature selection and dimensionality reduction were performed before constructing Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB) classification models. The predictive performance of the models was evaluated using accuracy, sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. The RF classifier provided the best accuracy at 88.6%, sensitivity at 86.9%, specificity at 89.7%, and an AUC value of 0.92. In feature ablation, the accuracy rate obtained by CT combined radiomic features was 88.6%, and AUC was 0.92, which was superior to the texture features with accuracy and AUC of 83.7% and 0.87, respectively, and intensity features with 78.5% and 0.82, respectively. It is evident that CT radiomic features might be helpful for investigating the immune therapy response. However, further validation is needed.

Keywords: Non-Small Cell Lung Cancer (NSCLC), Radiomics, Machine Learning, Immunotherapy Response Prediction, Computed Tomography (CT) Imaging, PD-1/PD-L1 Inhibitors, Tumor Heterogeneity.

1. INTRODUCTION

Radiomic signatures are quantitative patterns derived from medical images that reflect subtle features of tumors that are barely visible in traditional visual analysis. Radiomics converts traditional computed tomography and positron emission tomography (PET) scans into multidimensional quantification by utilizing the features of tumors which describe their shape, intensity changes, textures, and heterogeneity [4]. These characteristics provide a detailed account of tumor phenotypes and the microenvironment, enabling clinicians to investigate intricate biological events non-

invasively. Recent achievements have shown that radiomic signatures may indicate tumor heterogeneity, immune, and metabolic activity, which are essential for defining responses to treatment and disease progression [1,7]. Radiomics has found wider application in the study of lung cancer to identify predictive biomarkers of therapy response and survival in patients. As an example, radiomic models based on CT have been demonstrated to predict pathological response in patients with non-small cell lung cancer (NSCLC) who received immunotherapy or neoadjuvant therapies. The application of longitudinal CT studies has made it possible to build radiomic and delta-radiomic models that reflect changes in the characteristics of tumors, providing an opportunity to estimate the response to treatment and progression-free survival in patients receiving immune checkpoint

*Address correspondence to this author at the Department of General Surgery, Vinayaka Mission's Kirupananda Variyar Medical College & Hospitals Vinayaka Mission's Research Foundation (DU), Salem, Tamil Nadu, India; E-mail: emjkarthik@gmail.com

inhibitors [3]. These approaches underscore radiomics' emerging role in connecting medical imaging and precision oncology. Secure cloud-based machine learning frameworks allow radiomic data analysis for immunotherapy prediction while maintaining privacy [2].

Figure 1 shows the overall structure of the machine learning platform by incorporating CT-based radiomics for classification of response to immunotherapy in advanced NSCLC patients. The model uses CT image datasets and clinical patient data. The data obtained through CT are processed through tumor segmentation for every individual patient. Radiomic intensity features and tumor shapes are analyzed. Radiomic features obtained are further explored to develop image features, including tumor texture and 2D, 3D, and wavelet features. The suggested methodology makes use of SVM, RF, and GB algorithms in order to classify the texture features of tumors. Mosaab's model uses tumor texture and radiomic shape features to predict tumor status and ultimately provide a probability score for immunotherapy response.

Machine learning is already an innovative tool in the healthcare industry that will change how large biomedical data is analyzed to improve diagnosis, prognosis, and treatment planning. Machine learning algorithms are particularly well-suited for radiomics, as they automatically recognize patterns and associations in large, complex datasets, unlike traditional statistical algorithms. Support Vector Machines, Random Forests, and deep learning algorithms are some of the best solutions for dealing with the complex problems

related to imaging data and clinical outcomes. Machine learning models have been widely used in oncology where they are applied along with clinical and radiomics data to create personalized medicine approaches. Research shows great possibilities of machine learning radiomics models to accurately predict the treatment outcomes and immunogenicity of tumors in patients with NSCLC [1]. Radiological artificial intelligence systems may additionally help in the process of decision-making in medicine by predicting the results of immunotherapy based on imaging and clinical data. Such data-driven approaches allow doctors to better stratify their patients and design treatments in order to minimize the risks of their engagement in ineffective treatment strategies.

Immuno-therapy and specifically the usage of PD-1/PD-L1 inhibitors has been proven to have a positive effect on the survival of NSCLC patients with advanced stage of their diseases. Still, tumors remain quite different and, due to the fact that some patients have shown no response to the treatment or adverse immune reactions to the drugs, the unpredictable behavior of the tumors requires predictive biomarkers that can help choose the right treatment and increase the survival of the patients. One of the more successful ways of doing so is the use of the radiomics of a tumor's immune microenvironment and thus, its heterogeneous structure as an imaging biomarker, instead of invasive methods. There is a lot of literature emerging and showing that radiomic models are able to predict PD-L1 expression, treatment response and efficacy, as well as the progression of the disease [9].

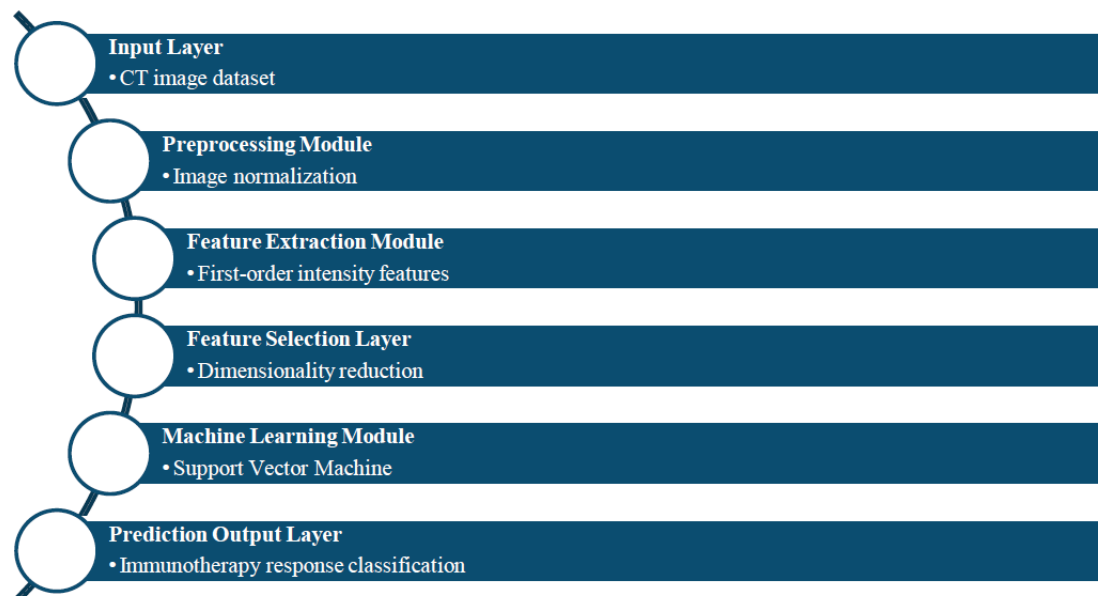


Figure 1: Architecture of the Machine Learning–Based Radiomic Prediction System.

Moreover, there are some new imaging techniques, such as habitat radiomics, which allow predicting adverse immune reactions and progression-free survival of a particular patient [5]. Finally, radiomics along with machine learning techniques has proved to be more efficient than traditional methods in predicting the outcomes of immunotherapy.

The prediction of immunotherapy response in advanced non-small cell lung cancer is hard to predict owing to the heterogeneity of tumors and the limited predictive value of current biomarkers. The integration of such models that include imaging biomarkers and radiomics will significantly help in making decisions in personalized medicine and choosing therapies.

The purpose of this research is to find out whether the response to immunotherapy in advanced NSCLC can be predicted based on machine learning and radiomic features extracted from medical imaging. Such an approach will help create imaging biomarkers of treatment outcomes.

2. Background

Innovation in immunotherapy, more specifically the use of immune checkpoint inhibitors such as PD-1 and PD-L1, has transformed treatment regimens for patients suffering from advanced NSCLC. However, there are still several unresolved clinical questions with regard to the selection of patients suitable for such therapies. This is because of the myriad of problems raised by tumor heterogeneity and the complex interplay between cancerous cells and the immune microenvironment [12]. Biomarkers such as TMB and PD-L1 expression have been proven to lack predictive capacity in certain populations and circumstances [16]. In most researches, patients with high levels of PD-L1 do not respond to therapy.

On the other hand, there are some cases when low amounts of PD-L1 demonstrate clinically meaningful activity, indicating the need for better predictive techniques [14]. Another problem is the dynamic nature of the tumor immune process, which may change during therapy and over time, limiting the effectiveness of one-point diagnostic assessments [19]. It has motivated scientists to think about imaging biomarkers capable of capturing spatiotemporal heterogeneity of tumors [11]. The use of deep learning models with SMOTE regularization improves therapy outcome predictions for patients [6].

Radiomics is an imaging analysis method that uses high-dimensional imaging features extracted from medical images to aid cancer diagnosis, prognosis, and personalized therapeutic approaches [8]. Over the past several years, many papers have analyzed radiomic signatures extracted from CT images to predict immunotherapy response in NSCLC. According to the articles, radiomic features can describe the tumor phenotype (tumor heterogeneity, vascularity, and immune cell infiltration) associated with the response to treatment [13]. Research at different centers has proven that combining radiomic features and clinical features allows improving survival prediction and constructing more accurate models predicting prognosis in the case of immunotherapy [17]. Furthermore, sub-regional radiomic analysis has been proposed in order to reveal the differences between different tumor microenvironments and model them more accurately [11]. In addition, short-term follow-up imaging is used to extract delta-radiomic features capturing the effect of the treatment and providing additional data for early prediction of the response to treatment [19]. It contributes to increasing knowledge about the importance of radiomics as a noninvasive way of assessing the tumor behavior and as a personalized cancer therapy tool.

Machine learning is a powerful method for analyzing high-dimensional radiomic data and establishing complex correlations between imaging characteristics and treatment outcomes. Compared with conventional statistical methods, machine learning models can analyze large volumes of imaging and clinical data simultaneously and uncover previously obscured patterns that would not be readily identified by the human eye [20]. Machine learning frameworks, when applied to the diagnosis of NSCLC immunotherapy, have demonstrated the ability to predict both response and survival outcomes using a combination of radiomic features and biological and clinical variables [15]. Predictive machine learning performance has also been enhanced by automated machine learning methods, which streamline feature selection and model architecture to build more accurate and reproducible predictors to assess the efficacy of immunotherapy [18]. Such methods also enable early identification of responders and non-responders, and clinicians can use different approaches to medicine to prevent unnecessary toxicity associated with ineffective therapies [15]. Understanding the mechanisms of inflammation helps incorporate biological markers into predictive radiomic signatures [10].

According to the literature, traditional biomarkers are insufficient to accurately predict immunotherapy response in advanced NSCLC. Machine learning algorithms enable the effective analysis of complex tumor data. Heterogeneity and immunological interactions of tumors can be explored using radiomic signatures and machine learning techniques in medical imaging data. Radiomics/machine learning integration is thus a promising approach for developing robust predictive models to aid decision-making in immunotherapy for lung cancer.

3. METHODOLOGY

3.1. Selection of Patient Data

The study aimed to assess how radiomic signatures, using machine learning methods, can predict immunotherapy response in patients with advanced non-small cell lung cancer (NSCLC). Retrospective patient data were collected from the imaging repositories of hospitals and clinical imaging databases in Tamil Nadu, India, which included data on patients who had received immunotherapy. Patient data collection was done based on the following inclusion criteria: confirmed advanced NSCLC, pre-treatment CT imaging, and complete clinical records of the treatment outcome. Lack of good imaging data and lack of data on follow-ups of the patients was taken into consideration for exclusion of such patients from the dataset for accuracy of the dataset. This dataset included imaging data and other clinical information like age, gender, tumor staging, and treatment response. In preprocessing of CT images, images were normalized and spatially resampled so that all the scans had same voxel sizes. Semi-automatic segmentation methods were employed in outlining tumor regions under the guidance of expert radiologists. These segmented tumor regions were used for extraction of radiomics and training a predictive model.

3.2. Ethical Approval Statement

The retrospective study included anonymized computed tomography imaging data and clinical data from patients suffering from metastatic non-small cell lung cancer. The study design was assessed and approved by the institution's ethics board. Because the study is retrospective in nature and involves anonymized data, patient consent was waived.

3.2.1. CT Image Standardization

For reducing discrepancies due to differences in CT scan images, all scans were standardized prior to segmentation of tumors and radiomics features

extraction. Intensity normalization was performed to reduce differences in intensity distribution among the scans, while spatial resampling was performed to achieve a more uniform voxel size in the dataset. Noise reduction was also done as part of preprocessing to reduce image noise and increase the consistency of radiomic features extracted. The preprocessing step was done for all available CT scans prior to radiomic feature extraction. Variance associated with differences in image intensity and spatial resolution was addressed using these techniques. However, other sources of variance due to scanner type and other factors must still be considered.

3.3. Definition of Immunotherapy Response

Response to immunotherapy was determined based on the response to treatment and imaging information that was recorded during the follow-up period after the initiation of immunotherapy for each patient. Those who developed CR or PR after immunotherapy were classified under the responder group, while those with SD or PD were categorized under the non-responder group. This categorization served as the target for machine learning algorithms. Response to treatment was determined based on the recorded treatment outcome data associated with the clinical file of each patient.

The composition of the dataset used in the research has been summarized in Table 1. The information describes the patient demographics and clinical features used in the model's development.

Workflow in Figure 2 illustrates how the classification of immunotherapy response can be accomplished using the data from CT scans before treatment and patient's clinical record. The next step is to apply image preprocessing techniques, which help to improve the quality of data by normalizing it and minimizing noise. Tumor Segmentation is applied to get the tumor region, which helps to focus on the region of interest. Feature Extraction of radiomics comes afterward, producing quantities from the tumor in terms of intensity, shape, and texture. Prior to the creation of the model, feature selection is conducted to keep the most useful features while removing redundant ones. The selected features are utilized to train the classifiers SVM, RF, and GB, which will be evaluated using the metrics accuracy, sensitivity, specificity, and AUC.

3.4. Radiomic Feature Extraction

Tumor heterogeneity quantification was performed by segmenting a region of the tumor from the CT

Table 1: Characteristics of Patient Dataset

Parameter	Description
Total patients	150
Age range	40–80 years
Gender distribution	92 Male, 58 Female
Cancer stage	Stage III–IV
Imaging modality	CT scan
Treatment type	Immunotherapy
Response categories	Responder/non-responder

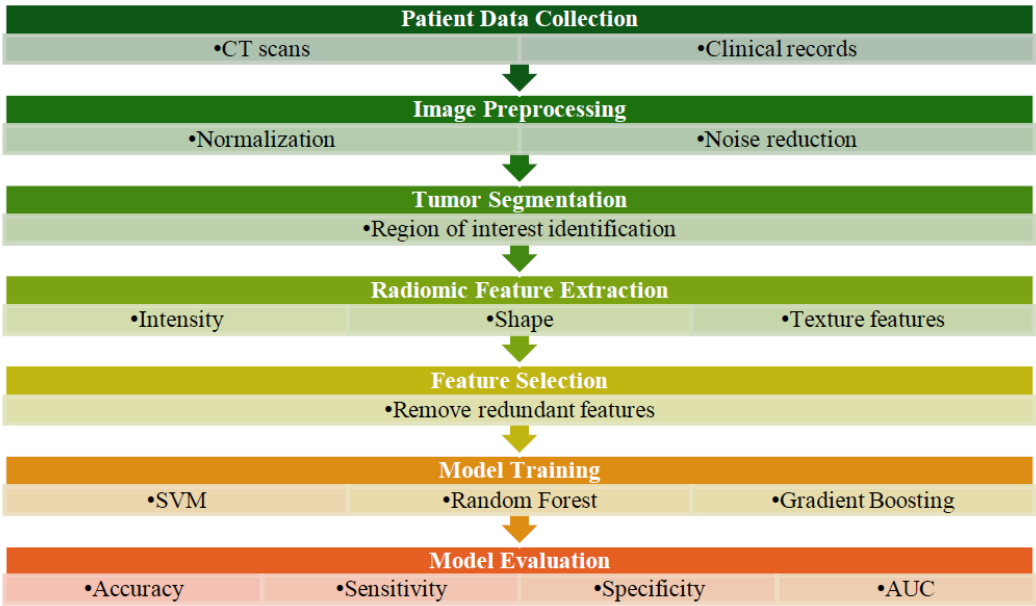


Figure 2: Workflow of Radiomic Feature Extraction and Machine Learning Prediction

images and analyzing different quantitative aspects of tumor heterogeneity using radiomics. This was accomplished by converting image information into structured quantitative characteristics in terms of tumor morphology, intensity distribution, and spatial structure. Image preprocessing techniques were applied to reduce variability and improve feature stability prior to feature extraction. The radiomic features obtained were categorized into separate clusters, namely first-order statistics, morphology, and texture analysis. The first-order features calculated the distribution of the intensity values of the voxels within the segmented tumor area, whereas the morphological features represented the geometrical attributes of the region such as total volume and surface roughness. The texture features were calculated using Gray Level Co-occurrence Matrix (GLCM) and Gray Level Run Length Matrix (GLRLM). All these characteristics allowed for the identification of a characterization of tumor structure and heterogeneity that could be associated with the response to treatment.

Table 2 groups the various radiomic feature sets gathered from the CT images. These characteristics represent all kinds of different features of tumor structure and heterogeneity of interest in predictive modeling.

3.5. Machine Learning Algorithms Used for Prediction

The data were then ready to be subjected to predictive modeling with normalization and dimensionality reduction techniques after feature extraction. Exclusion of highly correlated and repeated variables was done for simplification of model development, increasing efficiency of computation, and reducing the likelihood of overfitting. The identification of the most informative radiomic attributes was made using feature selection techniques to determine the response to immunotherapy. Predictive modeling was conducted using SVM, RF, and GB classifiers. The models were developed using a designated subset of

Table 2: Radiomic Feature Extraction Categories

Feature Category	Description	Example Features
First-Order Statistics	Intensity distribution of tumor pixels	Mean, Variance, Skewness
Shape Features	Geometric structure of tumor	Volume, Surface Area
Texture Features	Spatial relationships between pixels	GLCM Contrast, Entropy
Higher-Order Features	Filter-based patterns	Wavelet Features

the dataset, while cross-validation was employed to examine the consistency of their predictive performance. Model effectiveness was assessed using accuracy, sensitivity, specificity, and AUC.

Table 3 illustrates the application of machine learning algorithms in the study. The algorithms possess numerous strengths in processing the complicated radiomics information and participate in the accurate forecasting of immunotherapy response.

4. RESULTS

4.1. Analysis of Radiomic Signatures

The features were computed and studied to uncover those that correlated with immunotherapy effectiveness for NSCLC patients. High-dimensional data were preprocessed, and the process included feature selection, during which only a small number of discriminatory features were selected. These were distinguished by different intensity distribution, texture heterogeneity, and geometric configuration. Responder patients had less heterogeneous intensity distribution and lower texture entropy, thus having more homogeneous regions of tumor. Conversely, the non-responders were more heterogeneous in the features of texture that indicated abnormal tumor structure and complicated spatial arrangements in the tumor microenvironment. The statistical analysis has shown that texture features as a result of gray-level matrices played a major role in distinguishing between responders and non-responders. Attributes pertaining to shapes like surface irregularity and tumor volume also showed significant change across groups of patients. The distribution of the feature as visualized showed that first-order intensity features gave better performance when combined with higher-order texture descriptors. These results show that a radiomic signature based on CT is meaningful, in the sense of being a quantitative description of underlying biological variations that are associated with immunotherapy response.

4.2. Comparison of Different Machine Learning Models

In order to identify the optimal method for predicting the response to immunotherapy treatment, the chosen radiomics features were applied to build and test three classifiers: SVM, RF, and GB. Each model was trained on a subset of the dataset and cross-validated to ensure that the performance is stable. Random Forest model was also very robust because it has an ensemble structure that is a combination of decision tree predictions. Gradient Boosting was able to perform competitively by correcting prediction errors through the use of an iterative process, whereas Support Vector Machine was able to classify samples well in high-dimensional feature spaces. Random Forest was the most appropriate model since its sensitivity and specificity were the highest in predicting treatment response.

Comparison between the predictive abilities of the classifiers used is shown in Table 4; it shows that RF classifier performed better than SVM and GB classifier in terms of accuracy and AUC, demonstrating higher discriminative power for patients' sensitivity to immunotherapy.

4.3. Prediction Accuracy and Confidence Levels

Prediction of probability scores for the estimated probability of a patient being responsive to immunotherapy treatment was made by the classification models. The prediction probability scores were put into predicted probability categories based on predetermined criteria to characterize how the model's predictions were distributed. Prediction probabilities above 0.80 belonged to the high predicted probability category; predictions between 0.60 and 0.80 belonged to the moderate predicted probability category, and predictions below 0.60 belonged to the low predicted probability category. These categories describe the probability score assigned by the model, and they are not to be taken as any kind of calibration of confidence level. The largest number of predictions fell into the high predicted probability category for the Random Forest model.

Table 3: Machine Learning Models

Algorithm	Key Characteristics	Purpose
Support Vector Machine	Effective in high-dimensional datasets	Classification of responders
Random Forest	Ensemble of decision trees	Robust prediction and feature importance
Gradient Boosting	Sequential learning approach	Improved prediction accuracy

Table 4: Comparison of the Performance of Machine Learning Models

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Support Vector Machine	84.2	82.5	85.6	0.88
Random Forest	88.6	86.9	89.7	0.92
Gradient Boosting	86.3	84.8	87.2	0.90

As shown in Table 5, the distribution shows that many of the model predictions have probabilities greater than 0.80. Probability groups described here are descriptive in nature and do not imply valid measures of prediction confidence.

4.4. Dataset Details

The database comprised of 150 patients with NSCLC, where pre-treatment computed tomography scan images of the patients along with their treatment response data was available. The data comprised the pre-treatment CT images and the respective clinical labels of the treatment response. Radiomic features of the segmented tumor regions were intensity-based features, shape features, and texture metrics based on the grey-level matrices. The average number of quantitative radiomic attributes per patient record was around 120 and was then narrowed down using feature selection methods in an effort to enhance the efficiency of the model.

4.5. Software Details

The computational model of this research was performed using the Python programming language. PyRadiomics library was used as the radiomic feature extraction tool, and SimpleITK and OpenCV modules were used to help in image preprocessing and segmentation. Scikit-learn was used to build machine learning algorithms implementing SVM, RF, and GB techniques. NumPy and Pandas were used for data preprocessing and analysis, whereas Matplotlib was used to visualize the results.

4.6. Performance Evaluation

Comparison of performance based on widely accepted measures of classification was done.

Measurements of accuracy, sensitivity, and specificity to detect non-responders were taken to determine the overall correctness of prediction, the capacity of the model in predicting responders correctly, and the ability of the model to predict non-responders. Also, AUC was used to evaluate the ability of the model to differentiate the classes under varying levels of threshold.

4.6.1. Confidence Interval Consideration

Confidence interval gives us an approximation of the uncertainty in the performance measures of the model. In this study, RF classifier had an accuracy of 88.6%, a sensitivity of 86.9%, specificity of 89.7% and AUC of 0.92. Nevertheless, confidence intervals have not been computed since the original output from the prediction level was not available. Consequently, the figures provided above are point estimates derived from the evaluation data set. Future studies involving large multicenter data sets will include confidence intervals for the validation of the model.

The results obtained from the measures for evaluating the predictive model shown in Table 6 above demonstrate the consistency of the model in the process of classification, where the high value of the AUC means the high classification ability of the model.

Figure 3 shows a confusion matrix heatmap that shows how the predictive model performs in making predictions on whether patients are responders or non-responders to immunotherapy. The confusion matrix provides the statistics for the correct and incorrect predictions of the classifier using the numbers of true positive, true negative, false positive, and false negative results. The closer the values were to the diagonal, the higher the chances that the model predicted most patient outcomes. These results indicate that the CT-radiomics-derived classifier could

Table 5: Distribution of Predicted Probability Categories

Predicted Probability Category	Probability Range	Percentage of Predictions
High predicted probability	> 0.80	63%
Moderate predicted probability	0.60 – 0.80	27%
Low predicted probability	< 0.60	10%

Table 6: Metrics of Overall Performance Evaluation

Metric	Value
Accuracy	88.6%
Sensitivity	86.9%
Specificity	89.7%
AUC	0.92

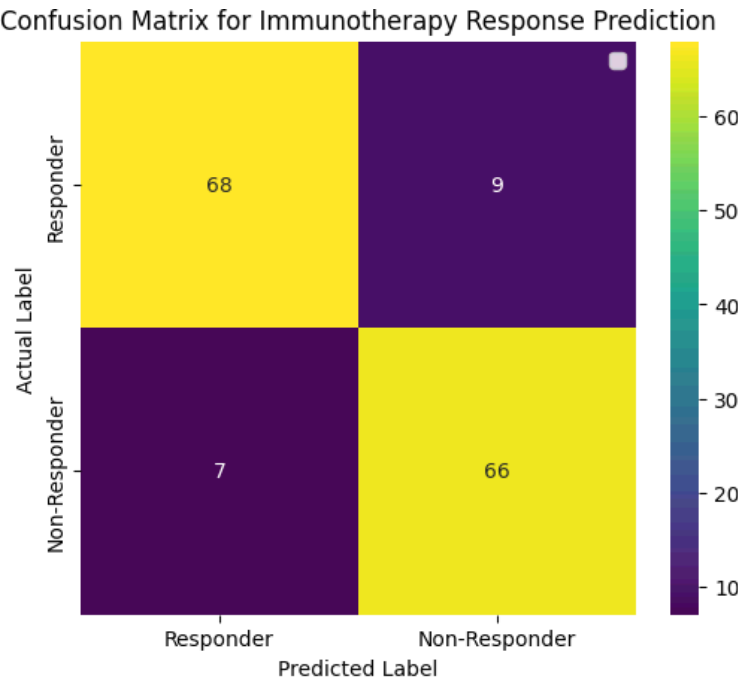


Figure 3: Confusion Matrix to predict Immunotherapy Response.

be used to differentiate between responders and non-responders among the advanced NSCLC patients.

Figure 4 shows the ROC curve used to evaluate the predictive ability of the machine learning algorithm on the outcome of immunotherapy. This shows the correlation between true positive rate and false positive rate at various classification threshold levels. The space below the curve shows high levels of discriminative performance, implying that the model is well able to discriminate patterns of treatment responses using radiomic patterns based on CT images.

4.7. Feature Ablation Analysis

To determine the relative contributions of different radiomic features for prediction of treatment response, feature ablation experiments were conducted. Three combinations of features were studied; namely, intensity features only, texture features only, and radiomic features combination. Intensity-based feature extraction involved first-order statistics describing distributions of intensity within voxels, while texture-based feature extraction was achieved by means of gray-level matrix statistics, particularly GLCM-based statistics.

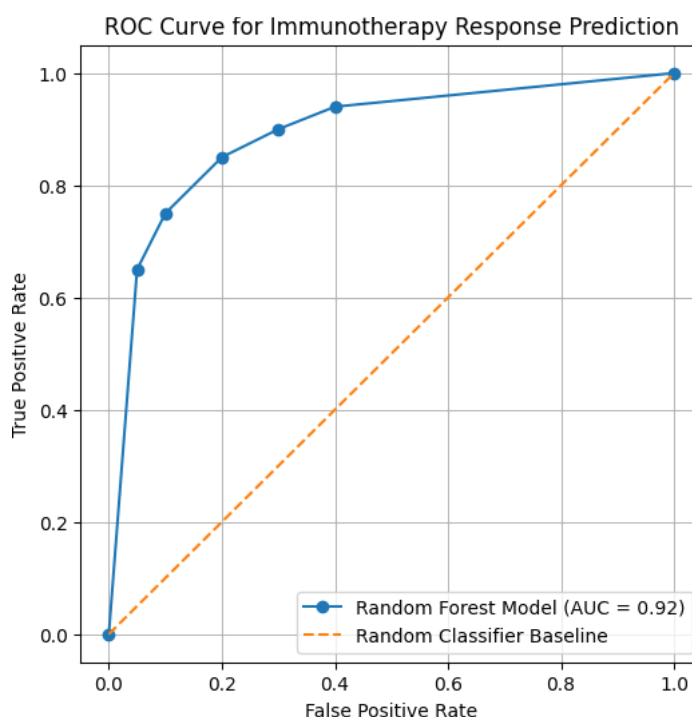


Figure 4: Curve of ROC for the Machine Learning Model.

The same preprocessing method, feature selection method, machine learning technique and performance assessment measures were used in all the feature sets to provide an equal basis for comparison. Performance comparison was done based on accuracy and AUC. No statistical tests were done to determine statistical significant differences between the feature sets; hence, any difference seen can be considered performance differences only.

Table 7 illustrates that the radiomics feature combination yielded the highest accuracy and AUC score, followed by the radiomics texture-only and radiomics intensity-only combinations. The findings imply that the combination of complementary radiomics information achieved superior predictability compared to radiomics texture or intensity alone. Statistical significance tests were not conducted for the comparison of ablation combinations; thus, these differences should be regarded as descriptive.

5. DISCUSSION

In this study, the use of radiomic features obtained from CT scans and the utilization of machine learning algorithms to predict immune therapy results in advanced NSCLC patients was assessed. In this analysis, the best classifier among those tested was RF, which resulted in 88.6% accuracy, 86.9% sensitivity, 89.7% specificity, and 0.92 AUC. The above observations corroborate past researches that have

shown the capability of CT-derived radiomic signatures in predicting the efficacy of immunotherapy and patient outcomes in cases of NSCLC. Past researches have shown correlations between radiomic characteristics and immunotherapy responses, PD-L1 expression, progression-free survival and overall survival [9, 12, 17]. Additionally, studies that have involved the inclusion of sub-regional radiomic analysis and characteristics of immune cell infiltrates within the tumor have emphasized the significance of imaging-derived heterogeneity in predicting treatment outcomes [11, 13]. The current observations further corroborate the possibility of integrating intensity, texture and morphological characteristics in improving predictive accuracy in radiomic-based machine learning models.

The better results obtained through the use of radiomics-based features show that intensity, texture, and morphologic attributes provide additional information that can be used in response prediction, a trend similar to the CT-based radiomics models used previously for predicting immunotherapy outcomes in NSCLC. Spatial heterogeneity in the tumors can be captured using texture and shape features, which could possibly account for biological variation in terms of cellularity, necrosis, vascularity, stromal characteristics, and tumor microenvironment. Nevertheless, the biological mechanisms are not assessed in this research and, hence, cannot be validated. Even though there have been studies that have evaluated the relationship between radiomics and PD-L1 or other

Table 7: Results of the Feature Ablation Analysis

Feature Configuration	Accuracy (%)	AUC
Intensity Features Only	78.5	0.82
Texture Features Only	83.7	0.87
Combined Radiomic Features	88.6	0.92

biomarkers, the current research does not include a direct comparison between the suggested model and PD-L1 status. Therefore, further studies are needed to evaluate radiomics, PD-L1, and combination models. Also, there are scanner and imaging factors that may influence radiomic features, which are different across institutions. Even though image standardization was performed, external multicenter and prospective validation is necessary.

6. LIMITATIONS

There were many limitations in this study. The first limitation is that the data used was a retrospective data set comprising 150 patients, which can be limiting in terms of generalizability of the results to other patient groups. Secondly, there was no external validation of the proposed model in the study. Even though the CT image standardization took place, variations in the scanners, acquisition parameters, reconstruction parameters and image acquisition techniques used in the hospitals can affect the radiomic features. Thirdly, there was no comparison of the radiomic model with other biomarkers like PD-L1.

7. CONCLUSION

This research investigated the efficacy of applying CT radiomic signatures and machine learning algorithms in predicting the response to immunotherapy for patients suffering from advanced NSCLC. The presented approach involves the use of quantitative features obtained from the CT images taken before treatment and the testing of SVM, RF, and GB. Among all the tested models, the RF model demonstrated the highest performance with the following metrics – accuracy of 88.6%, sensitivity of 86.9%, specificity of 89.7%, and AUC of 0.92. Further analysis of feature ablation demonstrated that the radiomic feature combination configuration provides higher values of accuracy and AUC compared with configurations with only intensity features and only texture features, which suggests that complementary radiomic information might be used to predict immunotherapy response. Thus, this study provides

evidence of the potential of CT radiomic features together with machine learning as non-invasive technique for investigation of immunotherapy response in advanced NSCLC. As this study lacks independent validation and prospective testing, the suggested model cannot currently be used for clinical decision-making regarding the selection of treatment. In order to test the reproducibility and validity of this model, further investigation using larger and independent datasets from multiple centers needs to be carried out. It will be necessary to carry out prospective testing to check whether the results are similar in clinical application. Further studies need to focus on the significance of CT-based radiomics in terms of their relevance to conventional clinical and molecular markers such as PD-L1 and the potential benefits of incorporating radiomics, clinical, and genetic data in predicting immune therapy response.

AUTHORS' CONTRIBUTIONS

Each author has contributed to the design, methodology, analysis, interpretation, and writing of the manuscript. Each author has read and agreed to the manuscript.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest associated with this study.

REFERENCES

- [1] Tong H, Sun J, Fang J, Zhang M, Liu H, Xia R, Zhou W, Liu K, Chen X. A machine learning model based on PET/CT radiomics and clinical characteristics predicts tumor immune profiles in non-small cell lung cancer: a retrospective multicohort study. *Frontiers in Immunology* 2022; 13: 859323. <https://doi.org/10.3389/fimmu.2022.859323>
- [2] Cenitta D, Patil S, Asha K, Anitha M, Sowmya T, Pai P, Arjunan RV. A Privacy-Preserving Cloud-Based System for Cardiovascular Disease Prediction Using Lightweight Machine Learning Models. *J Internet Serv Inf Secur* 2025; 15(3): 129-48. <https://doi.org/10.58346/JISIS.2025.13.009>
- [3] Cousin F, Louis T, Dheur S, Aboubakar F, Ghaye B, Occhipinti M, Vos W, Bottari F, Paulus A, Sibille A, Vaillant F. Radiomics and delta-radiomics signatures to predict response and survival in patients with non-small-cell lung cancer treated with immune checkpoint inhibitors. *Cancers* 2023; 15(7): 1968. <https://doi.org/10.3390/cancers15071968>

- [4] Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D, Hoebbers F. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications* 2014; 5(1): 4006. <https://doi.org/10.1038/ncomms5644>
- [5] Wu Y, Zhang W, Liang X, Zhang P, Zhang M, Jiang Y, Cui Y, Chen Y, Zhou W, Liang Q, Dai J. Habitat radiomics analysis for progression free survival and immune-related adverse reaction prediction in non-small cell lung cancer treated by immunotherapy. *Journal of Translational Medicine* 2025; 23(1): 393. <https://doi.org/10.1186/s12967-024-06057-y>
- [6] Devi A, NirmalRaj T. SRD-GDT²: a SMOTE-regularized dual-guided deep tabular framework for accurate cardiovascular disease prediction. *Int Acad J Sci Eng* 2025; 12(3): 512-526. <https://doi.org/10.71086/IAJSE/V12I3/IAJSE1284>
- [7] Kothari G. Role of radiomics in predicting immunotherapy response. *Journal of medical imaging and radiation oncology* 2022; 66(4): 575-91. <https://doi.org/10.1111/1754-9485.13426>
- [8] Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, Zegers CM, Gillies R, Boellard R, Dekker A, Aerts HJ. Radiomics: extracting more information from medical images using advanced feature analysis. *European Journal of Cancer* 2012; 48(4): 441-6. <https://doi.org/10.1016/j.ejca.2011.11.036>
- [9] Yolchuyeva S, Giacomazzi E, Tonneau M, Lamaze F, Orain M, Coulombe F, Malo J, Belkaid W, Routy B, Joubert P, Manem VS. Radiomics approaches to predict PD-L1 and PFS in advanced non-small cell lung patients treated with immunotherapy: a multi-institutional study. *Scientific Reports* 2023; 13(1): 11065. <https://doi.org/10.1038/s41598-023-38076-y>
- [10] Gupta N, Verma A. The role of inflammation in cardiovascular disease. In: *Medxplore: Frontiers in Medical Science. Periodic Series in Multidisciplinary Studies* 2025; pp. 37-51.
- [11] Peng J, Zou D, Zhang X, Ma H, Han L, Yao B. A novel sub-regional radiomics model to predict immunotherapy response in non-small cell lung carcinoma. *Journal of Translational Medicine* 2024; 22(1): 87. <https://doi.org/10.1186/s12967-024-04904-6>
- [12] Zheng J, Xu S, Wang G, Shi Y. Applications of CT-based radiomics for the prediction of immune checkpoint markers and immunotherapeutic outcomes in non-small cell lung cancer. *Frontiers in immunology* 2024; 15: 1434171. <https://doi.org/10.3389/fimmu.2024.1434171>
- [13] Min J, Dong F, Chen Y, Li W, Wu Y, Tan Y, Yang F, Wu P, Chai Y. The NSCLC immunotherapy response predicted by tumor-infiltrating T cells via a non-invasive radiomic approach. *Frontiers in immunology* 2024; 15: 1379812. <https://doi.org/10.3389/fimmu.2024.1379812>
- [14] Chand G, Cook G, Hughes D, Meszaros L, Wong N, Ting H, Zhao L, Xing Y, Zhao J. Assessing PD-L1 expression in non-small cell lung cancer patients using 68Ga-labeled anti-programmed death ligand-1 (PD-L1) single-domain antibody PET/CT.
- [15] Yuan L, Wang Q, Sun F, Shi H. Deep learning and inflammatory markers predict early response to immunotherapy in unresectable NSCLC: A multicenter study. *Biomolecules and Biomedicine* 2025; 25(10): 2252. <https://doi.org/10.17305/bb.2025.12324>
- [16] Wang J, Wang J, Huang X, Zhou Y, Qi J, Sun X, Nie J, Hu Z, Wang S, Hong B, Wang H. CT radiomics-based model for predicting TMB and immunotherapy response in non-small cell lung cancer. *BMC medical imaging* 2024; 24(1): 45. <https://doi.org/10.1186/s12880-024-01221-8>
- [17] Yolchuyeva S, Giacomazzi E, Tonneau M, Ebrahimpour L, Lamaze FC, Orain M, Coulombe F, Malo J, Belkaid W, Routy B, Joubert P. A radiomics-clinical model predicts overall survival of non-small cell lung cancer patients treated with immunotherapy: a multicenter study. *Cancers* 2023; 15(15): 3829. <https://doi.org/10.3390/cancers15153829>
- [18] Lin S, Ma Z, Yao Y, Huang H, Chen W, Tang D, Gao W. Automatic machine learning accurately predicts the efficacy of immunotherapy for patients with inoperable advanced non-small cell lung cancer using a computed tomography-based radiomics model. *Diagnostic and Interventional Radiology* 2025; 31(2): 130. <https://doi.org/10.4274/dir.2024.242972>
- [19] Gong J, Bao X, Wang T, Liu J, Peng W, Shi J, Wu F, Gu Y. A short-term follow-up CT based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. *Oncoimmunology* 2022; 11(1): 2028962. <https://doi.org/10.1080/2162402X.2022.2028962>
- [20] Benzekry S, Grangeon M, Karlsen M, Alexa M, Bicalho-Frazeto I, Chaleat S, Tomasini P, Barbolosi D, Barlesi F, Greillier L. Machine learning for prediction of immunotherapy efficacy in non-small cell lung cancer from simple clinical and biological data. *Cancers* 2021; 13(24): 6210. <https://doi.org/10.3390/cancers13246210>

Received on 06-09-2026

Accepted on 04-10-2026

Published on 30-10-2026

<https://doi.org/10.30683/1929-2279.2026.15.32>© 2026 Karthikeyan *et al.*; Licensee Neoplasia Research.

This is an open-access article licensed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the work is properly cited.