Gonul Caliskan, Emre Alatas, Sevilay Koksal, Safak Yigit, Resit Akyel, Iffet Cavdar, Goksel Alcin, Esra Arslan, Tuba Ciftci-Kusbeci, Elcin Dereli, Elife Akgun and Handan Tanyildizi-Kokkulunk*

Classification of malignant/benign groups in lung cancer by machine learning and investigation of feature significance of parameters

https://doi.org/10.1515/bmt-2024-0628
Received January 2, 2025; accepted July 5, 2026;
published online July 28, 2026

Abstract

Objectives: This study aimed to apply machine learning (ML) models to enhance lung cancer (LC) classification, distinguishing malignant from benign tumors, using data from 73 patients.

Methods: The dataset included PET/CT biomarkers and demographic factors, with SMOTE applied to address class imbalance. Three models were evaluated using 10-fold cross-validation to compare Random Forest (RF), Decision Tree (DT), Extra Trees Classifier (ETC), and XGBoost algorithms based on accuracy and AUC.

Results: ETC performed best in Model 1 (86 % accuracy, AUC 0.95), RF in Model 2 (94 % accuracy, AUC 0.97), and XGBoost in Model 3 (94 % accuracy, AUC 0.98). XGBoost consistently outperformed others, particularly in Model 3, which included age and smoking. Feature importance analysis highlighted SUVmax as the most predictive variable, with smoking having a moderate influence and gender being minimal.

Conclusions: Integrating clinical and lifestyle data with PET/CT parameters significantly improved LC classification. XGBoost emerged as the most effective model, demonstrating that comprehensive models enhance diagnostic accuracy beyond traditional metrics.

Keywords: lung cancer; malignant; benign; machine learning; XGBoost; smote

***Corresponding author: Handan Tanyildizi-Kokkulunk**, Vocational School of Health Sciences, Radiotherapy Program, Altınbaş University, Kartaltepe Dist., No.11, Bakirkoy, Istanbul, 34147, Türkiye; and Department of Arts & Sciences, Maine Maritime Academy, Castine, 04420, ME, USA, E-mail: handan.kokkulunk@mma.edu. https://orcid.org/0000-0001-5231-2768

Gonul Caliskan, Physiotherapy and Rehabilitation Master Program, Istanbul Bilgi University, Institute of Graduate Programs, Istanbul, Türkiye, E-mail: gonul.caliskan@bilgiedu.net. https://orcid.org/0009-0008-4778-1920

Emre Alatas, Faculty of Management Sciences, Department of Management Information Systems, Bogazici University, Istanbul, Türkiye, E-mail: emre.alatas@bogazici.edu.tr. https://orcid.org/0000-0003-4579-5057

Sevilay Koksal, Nuclear Physics Master Program, Istanbul University, Institute of Science and Technology, Istanbul, Türkiye, E-mail: sevilaykksl@gmail.com. https://orcid.org/0009-0005-0556-2808

Safak Yigit, Vocational School, Physiotherapy Program, Istanbul Galata University, Istanbul, Türkiye, E-mail: safak.yigit@galata.edu.tr. https://orcid.org/0000-0003-4815-3674

Resit Akyel, Department of Nuclear Medicine, Yedikule Chest Diseases Hospital, Istanbul, Türkiye, E-mail: akyelresit@hotmail.com. https://orcid.org/0000-0001-9373-8693

Iffet Cavdar, Faculty of Science, Department of Nuclear Physics, Istanbul University, Istanbul, Türkiye, E-mail: icavdar@istanbul.edu.tr. https://orcid.org/0000-0003-0089-0499

Goksel Alcin, Esra Arslan and Elife Akgun, Department of Nuclear Medicine, Istanbul Training and Research Hospital, Istanbul, Türkiye, E-mail: gokselalcin@hotmail.com.tr (G. Alcin), dresraarslan@gmail.com (E. Arslan), elifekaymak@hotmail.com (E. Akgun). https://orcid.org/0000-0003-2268-9606 (G. Alcin). https://orcid.org/0000-0002-9222-8883 (E. Arslan). https://orcid.org/0000-0001-5625-9749 (E. Akgun)

Tuba Ciftci-Kusbeci, Faculty of Medicine, Department of Chest Diseases, Altınbaş University, Istanbul, Türkiye, E-mail: ciftcituba@gmail.com. https://orcid.org/0000-0002-6359-5908

Elcin Dereli, Faculty of Health Sciences, Physiotherapy and Rehabilitation Program, Istanbul Bilgi University, Istanbul, Türkiye, E-mail: elcin.dereli@bilgi.edu.tr. https://orcid.org/0000-0002-1572-0620

Introduction

Lung cancer continues to be the primary cause of cancer-related death globally, with notable regional differences in incidence rates. Lung cancer is the second most prevalent cancer diagnosed worldwide, accounting for 11.4 % of all new cases in 2020, according to the Global Cancer Observatory [1]. Genetic predispositions, environmental exposures, and smoking prevalence all have an impact on the incidence of lung cancer. Malignant and benign are the two main classifications of lung cancer. The more common malignant lung tumors are small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), with NSCLC making up about 85 % of cases [2]. Despite being less frequent, benign lung tumors can nevertheless be extremely dangerous to patient health because they can mimic malignant tumors and result in

consequences, including airway obstruction [3]. Accurate diagnosis and treatment planning depend on the distinction between benign and malignant lung tumors, underscoring the significance of sophisticated imaging methods and histological examination in clinical practice.

Positron Emission Tomography/Computed Tomography (PET/CT) is an invaluable tool, especially when assessing tumor anatomical aspects and metabolic activity in diagnosing and treating cancer. Important PET/CT-based metrics for oncological evaluations are the metabolic tumor volume (MTV), total lesion glycolysis (TLG), and maximum standardized uptake value (SUVmax). SUVmax is a measurement of a tumor's maximum concentration of radiotracer uptake that offers information about its metabolic activity; larger values are frequently associated with more aggressive disease [4]. TLG provides a thorough assessment of the tumor's overall metabolic load. It is especially helpful in determining treatment response and prognosis because it integrates metabolic activity (SUV) and the tumor volume [5]. MTV is a critical criterion for treatment planning and staging because it represents the metabolically active tumor volume [6]. Regarding lung cancer, several characteristics have demonstrated noteworthy predictive significance. Elevated SUVmax, TLG, and MTV are frequently linked to unfavorable results and can help distinguish benign from malignant lesions, direct biopsy choices, and customize treatment strategies.

The categorization and diagnosis of many diseases, including lung cancer, have greatly improved using machine learning (ML) in healthcare. Personalized medicine could benefit greatly from machine learning's capacity to automate and enhance categorization. This could lead to more focused treatment plans and better patient outcomes through early and accurate diagnosis.

In light of this information, this study's primary goal is to create ML models that use PET/CT-based anatomical-metabolic tumor parameters, such as SUV_{max}, TLG, and MTV, to classify lung cancer patients that are malignant and benign. Furthermore, the research aims to examine the importance of these characteristics in distinguishing between benign and malignant lung lesions.

Literature review

Recent developments in ML have improved lung cancer categorization significantly, allowing for a more accurate distinction between benign and malignant lesions. Research has indicated that ML algorithms have the capacity to handle and evaluate intricate imaging data, specifically from PET/CT scans, to enhance prognostic evaluations and diagnosis precision. By using deep learning models to analyze of low-dose chest CT scans, Ardila et al. [7] were able to diagnose lung cancer with a high degree of specificity and sensitivity. Convolutional neural networks (CNNs) were also used by Liu et al. [8], showing notable gains over conventional techniques in the categorization of lung nodules as benign or malignant.

Numerous research have examined the relative importance of SUV_{max}, MTV, and TLG in cancer diagnosis and prognosis within the context of PET/CT parameters. SUVmax is a commonly used metric to measure tumor metabolic activity. Higher SUVmax values were reported to be substantially linked to malignancy and a worse prognosis in lung cancer patients by Roengvoraphoj et al. [9]. On the other hand, there are studies reporting that SUV_{max} is not sufficient to represent the metabolic burden of the tumor entirely and that, in some cases, MTV and TLG are more successful in diagnosis [10–12].

Despite these findings, the relative importance of SUV_{max}, MTV, and TLG is still being researched. Combining these characteristics in ML models can enhance diagnostic performance and yield a more comprehensive understanding of tumor biology, according to studies by Ohri et al. [13] and Hyun et al. [14].

Materials and methods

The dataset

The study included 73 patients (F: 23, M: 50) scanned with suspected lung cancer or followed up in the Nuclear Medicine Department of Yedikule Chest Diseases Hospital. Inclusion criteria for the study were as follows: being 18 years of age or older, having no prior cancer diagnosis, signing the informed consent form, and having no active infection. All patients included had undergone FDG PET/CT imaging due to suspected lung cancer and had received a definitive diagnosis confirmed through histopathological examination and/or multidisciplinary clinical evaluation. Exclusion criteria included a history of other malignancies, uncontrolled severe systemic diseases (such as uncontrolled diabetes that could affect FDG uptake), and lack of cooperation that would impair the acquisition of high-quality PET/CT scans. FDG PET/CT was performed using a PET/CT scanner (Discovery 600, GE Medical Systems, Milwaukee, Wis, USA). The FDG PET/CT scan was performed after at least 6 h of fasting for all individuals. At the time of the FDG injection, their blood glucose levels were less than 150 mg/dL. FDG activity was 3.5 mCi/kg for each patient.

In the dataset, some demographic characteristics of the patients and some features based on PET/CT scans considered biomarkers for lung cancer were used to reveal the effect of ML applications for lung cancer diagnosis. In this dataset, the descriptive information of the attributes of 73 patients and the class numbers used in malignant/benign estimation are shown in Table 1 in detail.

In order to equalize the sample numbers, the data set was smoothed using virtual data replication. For this, the SMOTE technique was applied using the SMOTE function from the imblearn.over_sampling module in Python, with default parameters (random_state=42, sampling_strategy='auto'). Crucially, to prevent data leakage and artificially inflated performance metrics, the dataset was first split into training and testing sets. The SMOTE algorithm was then applied strictly to the training data alone; the test set remained completely unaltered and unseen during model training.

In the study, machine learning was performed by developing three different datasets in the form of PET/CT-based diagnostic parameters and the combination of parameters such as some demographic characteristics and smoking. Detailed information about the datasets is shown in Table 2.

Cross validation

The accuracy of the outcomes produced by systems designed for specific purposes can be used to assess their success. The k-fold cross-validation approach is the most widely used technique for this evaluation [15]. In this method, the dataset is divided into k subsets, each containing an equal number of data points. One cluster is utilized for testing, and k-1 clusters are used for training using this data set. Our hypothesis is supported by the average error value derived from k experiments.

Table 1: Descriptive information of attributes in the dataset.

Variable	Range	Mean
Smoking, year	Non-smoker (14), 1–10 (4), 11–20 (6), 21–30 (12), 30 and above (37)	–
SUV_{max}	0.4–39.4	11.7 ± 7.5
MTV	2.2–317.8	38.2 ± 52
TLG	3.7–4,292.4	307.6 ± 557.7
Age, year	41–89	64 ± 10
Class	**Category**	**Frequencies**
0	Malignant	59
1	Benign	14

Table 2: The features for three different datasets used in machine learning.

Model numbers	Used features					
Model 1	SUV_{max}	MTV	TLG			
Model 2	SUV_{max}	MTV	TLG	Age	Sex	Smoking
Model 3	SUV_{max}	Age	Sex	Smoking		

Due to the volume of data being collected and the amount of computing power needed, the k value for this experiment was set at 10. This technique was used to fill in the gaps the testtrain split strategy left. In addition to the cross-validation process, the data set was divided into training and test (7:3) halves, and the accuracy values were calculated and compared.

Machine learning algorithms

It has been preferred to utilize algorithms that can function with minimal data, are less affected by outlier observations, and show simplicity of application to any data format (particularly discrete, continuous, or mixed data) given the dataset used in the study. Furthermore, the Python software utilized scikit-learn, a free programming language library, for the machine learning application.

Four distinct ML techniques were assessed for binary classification, namely: Random Forest (RF), Decision Tree (DT), Extra Trees Classifier (ETC), and XGBoost. The accuracy, 10-fold cross validation based accuracy, and the receiver operating characteristic (ROC) curves were used to compare the discriminative performances of the four ML classifiers according to the definitions of malignant/benign. The performance evaluation measure was based on the area under the ROC curve (AUC). In addition, the importance of the attributes used on the malignant/benign classification was investigated.

To enhance the reliability of performance evaluation, 10-fold cross-validation was conducted during the modeling process. However, a 70:30 train-test split was primarily utilized for illustrative purposes, including performance comparison and the visualization of confusion matrices and ROC curves. Final performance metrics such as accuracy and AUC were calculated based on the results obtained from the 70:30 split.

The testing subset (30 %) was strictly held out and not used during model training. This allowed us to assess generalization on previously unseen data and to present confusion matrices and ROC curves based on truly internal validation using the held-out subset.

Results and discussion

In this study, ML was performed using RF, DT, ETC, and XGBoost algorithms for malignant/benign lung cancer classification as the dependent variable. Three different dataset models were created as combinations of potential independent variables used in lung cancer diagnosis. While the main dataset had six variables and 73 observations, the number of observations was increased to 118 by equalizing the number of patients belonging to the malignant/benign class with the SMOTE algorithm.

Accuracy (%), 10-fold cross validation based accuracy (%), and AUC values (%) were used to compare ML algorithms. Binary classification was used to predict malignant/benign groups, which were considered as two separate outputs. Figure 1 shows the confusion matrices produced by the four different ML techniques used for Models 1, 2, and 3. The values that were correctly predicted are shown in the first and fourth cells, whereas the values that were incorrectly predicted are shown in the other cells.

RF, DT, and ETC were the most accurate algorithms for Model 1. The numbers correctly classified for malignant/benign groups in all three models were 17/14. False negative rates for malignant and benign classification were 15 and 12.5 % for RF, DT, ETC algorithms, respectively, while 20 and 12.5 % for the XGBoost algorithm. For Model 2, the RF algorithm was the most accurate with 94 % prediction accuracy. In the malignant/benign classification, false negative rates were calculated as 5–6.2 %, 35–0 %, 10–12.5 %, and 15–6.2 % for RF, DT, ETC, and XGBoost algorithms, respectively. For Model 3, the most accurate algorithms were RF, DT, and XGBoost, with an accuracy rate of 94 %. The numbers correctly classified for malignant/benign groups in all three models were 19/15. False negative rates for malignant and benign classification were 5 and 6.2 % for RF, DT, and XGBoost algorithms, respectively, while 10 and 25 % for ETC algorithms.

ROC curves are a method that shows an algorithm's sensitivity and specificity capabilities. In addition to accuracy assessment, measuring how well the algorithm predicts classes is essential. The area under the ROC curve AUC value close to 1 is associated with algorithm's success. The AUC values, accuracy, and 10-fold cross validation based accuracy values of the algorithms for Models 1, 2 and 3 are shown in detail in Table 3.

Except for Model 1, other models' accuracy rates based on 10-fold cross-validation were low. For the Model 1 dataset, the ETC algorithm was the most successful method for malignant/benign classification, with an accuracy of 86 % and

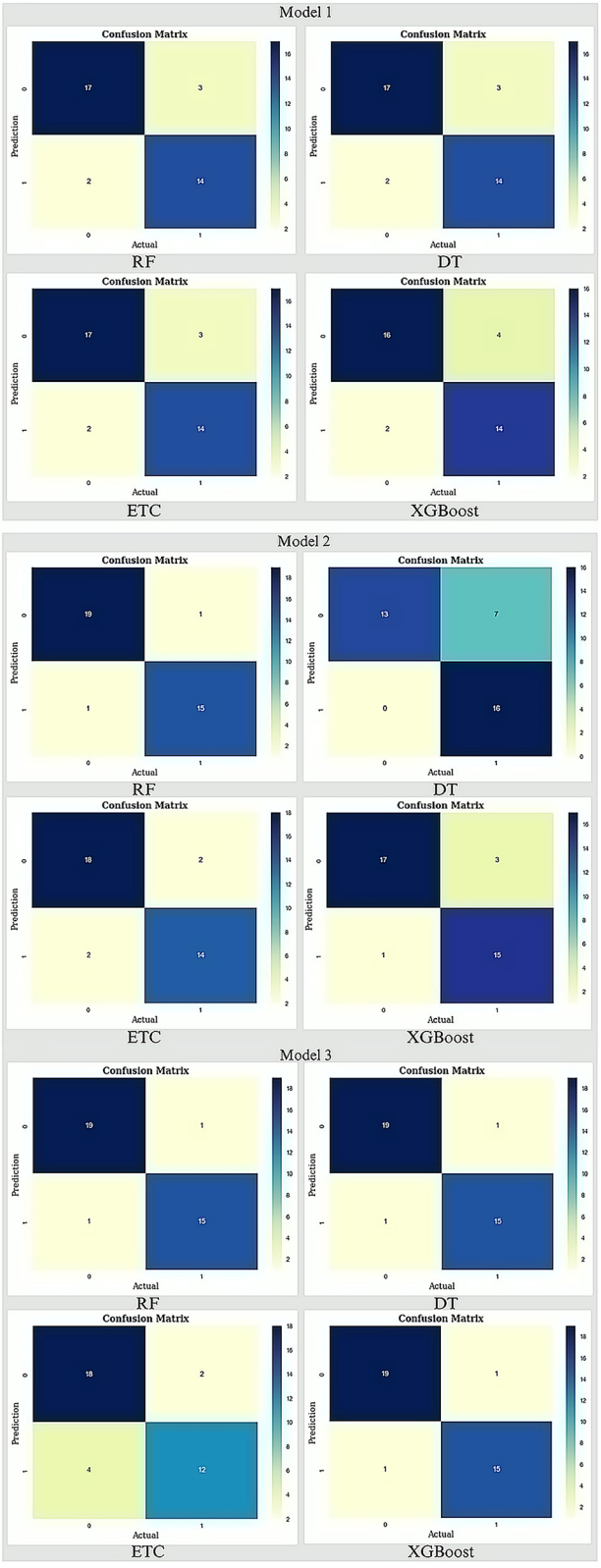

Figure 1: The confusion matrices produced by RF, DT, ETC, and XGBoost were used for Models 1, 2, and 3.

Table 3: The features of three different datasets used in ML.

		Accuracy	10-fold cross validation based accuracy	AUC based ROC curves
Model 1	RF	**0.86**	0.88	0.93
	DT	**0.86**	0.88	0.86
	ETC[a]	**0.86**	**0.91**	**0.95**
	XGBoost	0.83	0.87	0.91
Model 2	RF[a]	**0.94**	**0.91**	**0.97**
	DT	0.81	0.84	0.82
	ETC	0.89	0.90	0.96
	XGBoost	0.89	0.88	**0.97**
Model 3	RF	**0.94**	**0.90**	0.97
	DT	**0.94**	0.86	0.94
	ETC	0.83	0.88	0.95
	XGBoost[a]	**0.94**	0.87	**0.98**

[a]Indicates the optimal machine learning algorithm with the highest AUC value for the respective dataset model. Bold text highlights the best-performing algorithm for each specific evaluation criterion within the respective models.

an AUC value of 0.95. In the Model 2 dataset, the RF algorithm was the most successful prediction method, with 94 % accuracy and 0.97 AUC value. When the Model 3 dataset was evaluated, it was determined that XGBoost was the most optimal classification algorithm with 94 % accuracy and 0.98 AUC value. Although the performance metrics of the algorithms were numerically close across the three different dataset models, it was observed that XGBoost performed the best malignant/benign classification with 94 % accuracy and a 0.98 AUC value when utilizing the comprehensive Model 3 dataset. Alternatively, it was determined that the RF algorithm could still perform successful classification in machine learning through the Model 2 data set.

Table 4 shows the results of the feature importance level searches according to the classification algorithms showing optimal results for each of the three datasets.

Although there are publications in the literature that highlight the superiority of MTV and TLG over SUV_{max} in the analysis of PET-CT quantitative parameters in other cancer groups, such as lung cancer or spinal cord tumor cases [10, 12, 13], in our study, SUV_{max} [4, 16, 17], – widely accepted and established in clinical practice – was chosen as the common feature across three different dataset models in which various parameter combinations were tested.

Table 4: Results of three different datasets for feature importance search with optimal algorithm.

Feature names	Importances		
	Model 1 ETC algorithm	Model 2 RF algorithm	Model 3 XGBoost algorithm
SUV_{max}	0.48	0.44	0.60
MTV	0.20	0.12	–
TLG	0.32	0.23	–
Age	–	0.09	0.08
Sex	–	0.03	0.05
Smoking	–	0.09	0.27

In addition, only PET/CT anatomical-metabolic parameters were included in Model 1, all available parameters in Model 2, and parameters such as age, gender, and smoking were included in Model 3. In the feature significance analysis based on ETC, RF, and XGBoost algorithm, which provided optimal results in Model 1, Model 2, and Model 3, the SUV_{max} feature contributed the most to malignant/benign classification with 48 , 44, and 60 %, respectively.

Based on the results of Model 1 and Model 2, TLG contributed second (32 % for ETC, 23 % for RF) and MTV contributed third and last (20 % for ETC, 12 % for RF) to the classification of PET/CT anatomical-metabolic parameters. Ohri et al. [13] evaluated MTV among pretreatment FDG-PET metrics as a predictive parameter for clinical outcomes of patients treated with chemoradiotherapy in stage III non-small cell lung cancer. Matsimoto et al. [12] concluded that MTV can predict the clinical outcome of primary malignant spine/spinal tumors. Im et al. [10], evaluated MTV comprehensively and stated that it shows efficacy in many types of malignancies and is a comprehensive parameter that better reflects metabolic tumor burden. Our results differ from the literature in that TLG ranked second, and MTV ranked third on malignant/benign classification. While previous studies have highlighted the prognostic value of MTV and TLG for predicting survival and treatment response, our study specifically targets the diagnostic differentiation between benign and malignant lesions. In this classification task, SUVmax (representing peak metabolic activity) emerged as the most dominant feature. This suggests that while volumetric parameters are critical for assessing total tumor burden and prognosis, peak intensity may be more specific for the initial binary classification of malignancy, independent of the patient cohort demographics.

In our study, when Model 2 was evaluated within itself, gender was found to have a meager contribution to classification with 3 %. In comparison, age and smoking (9 %) contributed almost as much as MTV (12 %). When Model 3 was evaluated in itself, gender was found to have a very low contribution to classification with 5 %, while smoking was found to have a high effect with 27 %. Lim et al. [17] developed a ML approach using PET/CT-based radiomics for the prediction of PD-L1 expression in non-small cell lung cancer. Accordingly, optimal parameters were determined from radiomics features through the Gini index. Consistent with these previous findings, SUVmax emerged as the most

critical feature across all our models; however, relying solely on anatomical-metabolic parameters (Model 1) proved insufficient, yielding the lowest accuracy and highlighting the necessity of integrating clinical data, such as age and smoking, to achieve optimal performance (Models 2 and 3). It was concluded that gender, the parameter with the lowest contribution in all cases, is a negligible scale, in line with the literature [17].

In our study, RF, DT, ETC, and XGBoost algorithms were used for binary classification using tabular data sets in Excel format (.xlsx). XGBoost algorithm was found to provide optimal binary classification with 94 % accuracy and 0.98 AUC. In similar binary classification studies in the literature; Liu et al. [8] obtained 94 % accuracy and 0.981 AUC using multi-view convolutional neural networks for lung nodule classification, while other studies calculated accuracy between 77 and 88 % and AUC between 0.92 and 0.94 [18–22]. This comparison highlights that, for structured tabular data involving PET/CT and clinical features, traditional ML algorithms can achieve high classification performance, offering advantages in interpretability and computational efficiency without the need for complex image-level deep learning architectures.

Finally, 10-fold cross-validation was applied not to enhance performance, but to verify model stability. The fact that cross-validation results were consistent with the hold-out test set accuracy confirms the robustness of our models and indicates that the high classification performance is not due to a specific random split of the data. Although the performance difference between XGBoost and other classifiers (like RF) was not statistically significant, XGBoost was selected as the representative optimal model due to its consistent performance and highest numerical AUC value. This lack of significant difference also highlights that the combined dataset (Model 3) provides strong predictive power independent of the specific algorithm chosen.

Considering the scope of this study, several limitations should be acknowledged. First, the dataset was derived from a relatively small cohort of 73 patients from a single medical center. Although promising results were achieved in classification performance using multiple ML algorithms and cross-validation techniques, the limited sample size may affect the generalizability of the models. Furthermore, the single-center nature of the data restricts the variability in patient demographics and clinical protocols. Future research should aim to validate these findings in larger, multi-center datasets to enhance model robustness and ensure broader applicability across diverse clinical settings.

Moreover, the study did not include additional biological parameters such as genetic or inflammatory markers, which may also contribute to the malignant/benign classification. Parameters like PD-L1 expression, circulating tumor DNA, or C-reactive protein (CRP) could potentially enhance the diagnostic model when combined with PET/CT and clinical data. However, these were not routinely available for all patients in our cohort. Future studies should consider incorporating a broader range of biological indicators to develop more comprehensive and biologically informative models. In this context, adapting methods for robust biomarker screening, typically employed in gene expression data analysis, would be valuable for selecting the most predictive features from high-dimensional radiomic and clinical datasets.

Additionally, it is important to note that the feature importance scores reported in this study were derived from tree-based models (such as Gini importance). In small datasets, these native importance metrics can sometimes be biased or unstable. Therefore, while SUVmax emerged as a dominant predictive feature in our current models, future studies employing permutation-based feature importance or rigorous stability analyses on larger cohorts are necessary to confirm the absolute hierarchy of these diagnostic parameters.

Conclusions

This study systematically assessed the effectiveness of four ML algorithms – RF, DT, ETC, and XGBoost – in categorizing instances of malignant and benign lung cancer using various combinations of diagnostic criteria. The study demonstrated the effectiveness of the XGBoost algorithm, which achieved the highest classification accuracy of 94 % and an AUC value of 0.98. This was particularly evident when the dataset included patient demographics and lifestyle factors, such as age and smoking history, in addition to PET/CT anatomical-metabolic parameters.

Significantly, the results showed that although SUVmax, a widely used clinical metric, was consistently necessary for categorization in all models, the predictive accuracy was much improved by adding other variables, including age and smoking. Additionally, the study discovered that, in this specific patient group, parameters like TLG and MTV – which have been heavily stressed in earlier research – contributed less to classification accuracy than anticipated.

The study also highlighted the drawbacks of relying solely on anatomical-metabolic factors (as in Model 1), which produced the lowest accuracy and AUC values. This shows that classification models can be significantly enhanced by including readily available clinical data. The cross-validation findings were utilized to estimate the generalization performance of our models, confirming that the classification stability remained acceptable across different data splits, despite the inherent limitations of the small cohort size.

Overall, this work confirms that ML methods – particularly XGBoost – are highly effective in achieving robust accuracy in binary classification tasks related to lung cancer when utilizing structured diagnostic and clinical data. While these preliminary results support the potential of incorporating additional clinical data into prediction models to enhance lung cancer diagnosis, the limited sample size and single-center nature of this study necessitate cautious interpretation. Further validation using larger, multi-center cohorts is required before these machine learning models can be reliably integrated into routine clinical practice.

Acknowledgments: This study was supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) 1001 project number 123S718. The authors thank TÜBİTAK for its support of the project.

Research ethics: The study was approved by the local Institutional Ethics Committee of Altinbas University (Document ID/Approval Number: 2022/157).

Informed consent: Informed consent was obtained from all individuals included in this study.

Author contributions: All authors contributed equally to this work. All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: The authors state no conflict of interest.

Research funding: This study was financially supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Grant number: [123S718]).

Data availability: The data underlying this article will be shared on reasonable request to the corresponding author via e-mail.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
2. Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature 2018;553:446–54.
3. Dalar L, Ozdemir C, Sokucu SN, Nur Urer H, Altin S. Bronchoscopic treatment of benign endoluminal lung tumors. Cancer Res J 2019;2019: 5269728.
4. Bai B, Bading J, Conti PS. Tumor quantification in clinical positron emission tomography. Theranostics 2013;3:787–801.
5. Liao S, Penney BC, Wroblewski K, Zhang H, Simon CA, Kampalath R, et al. Prognostic value of metabolic tumor burden on 18F-FDG PET in nonsurgical patients with non-small cell lung cancer. Eur J Nucl Med Mol Imag 2012;39:27–38.
6. Choi JY. Treatment of bone metastasis with bone-targeting radiopharmaceuticals. Nucl Med Mol Imag 2018;52:200–7.
7. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JJ, Peng L, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 2019;25:954–61.
8. Liu K, Kang G. Multiview convolutional neural networks for lung nodule classification. Int J Imag Syst Technol 2017;27:12–22.
9. Roengvoraphoj O, Käsmann L, Eze C, Taugner J, Gjika A, Tufman A, et al. Maximum standardized uptake value of primary tumor (SUVmax_PT) and horizontal range between two most distant PET-positive lymph nodes predict patient outcome in inoperable stage III NSCLC patients after chemoradiotherapy. Transl Lung Cancer Res 2020;9:541–8.
10. Im HJ, Bradshaw T, Solaiyappan M, Cho SY. Current methods to define metabolic tumor volume in positron emission tomography: which one is better? Nucl Med Mol Imag 2018;52:5–15.
11. Liu DD, Li J, Li X, Xie L, Qin L, Peng F, et al. Prognostic value of metabolic tumor volume and total lesion glycolysis from 18F-FDG PET/CT in lymph node metastases and risk stratification of endometrial carcinoma. J Gynecol Oncol 2019;30:e89.
12. Matsumoto Y, Baba S, Endo M, Setsu N, Iida K, Fukushi JI, et al. Metabolic tumor volume by 18F-FDG PET/CT can predict the clinical outcome of primary malignant spine/spinal tumors. BioMed Res Int 2017;2017:8132676.
13. Ohri N, Duan F, Machtay M, Gorelick JJ, Snyder BS, Alavi A, et al. Pretreatment FDG-PET metrics in stage III non-small cell lung cancer: ACRIN 6668/RTOG 0235. J Natl Cancer Inst 2015;107:djv004.
14. Hyun SH, Ahn HK, Kim H, Ahn MJ, Park K, Ahn YC, et al. Volume-based assessment by (18)F-FDG PET/CT predicts survival in patients with stage III non-small-cell lung cancer. Eur J Nucl Med Mol Imag 2014;41:50–8.
15. Tougui I, Jilbab A, El Mhamdi J. Impact of the choice of cross-validation techniques on the results of machine learning-based diagnostic applications. Healthc Inform Res 2021;27:189–99.
16. Goksel S, Erdivanli OC, Bulbul O, Dursun E. The role of metabolic tumor parameters predicting cervical lymph node metastasis in patients with head and neck squamous cell carcinoma. J Cancer Res Therapeut 2022; 18:1045.
17. Lim CH, Koh YW, Hyun SH, Lee SJ. A machine learning approach using PET/CT-based radiomics for prediction of PD-L1 expression in non-small cell lung cancer. Anticancer Res 2022;42:5875–84.
18. Han F, Wang H, Zhang G, Han H, Song B, Li L, et al. Texture feature analysis for computer-aided diagnosis on pulmonary nodules. J Digit Imag 2015;28:99–115.
19. Shewaye TN, Mekonnen AA. Benign-malignant lung nodule classification with geometric and appearance histogram features. [Internet]. arXiv; 2016. [cited 2024 Aug 19]. Available from: http://arxiv.org/abs/1605.08350.
20. Kumar D, Shafiee MJ, Chung AG, Khalvati F, Haider MA, Wong A. Discovery radiomics for pathologically-proven computed tomography lung cancer prediction. [Internet]. arXiv; 2017. [cited 2024 Aug 19]. Available from: http://arxiv.org/abs/1509.00117.
21. Shen W, Zhou M, Yang F, Yang C, Tian J. Multi-scale convolutional neural networks for lung nodule classification. Inf Process Med Imaging 2015; 24:588–99.
22. Tanyildizi-Kokkulunk H, Alcin G, Cavdar I, Akyel R, Yigit S, Ciftci-Kusbeci T, et al. Machine learning-assisted classification of lung cancer: the role of sarcopenia, inflammatory biomarkers, and PET/CT anatomical-metabolic parameters. Phys Eng Sci Med 2026;49:61–73.