Received: 9 September 2023 | Revised: 14 December 2023 | Accepted: 8 January 2024

DOI: 10.1002/ima.23028

RESEARCH ARTICLE

WILEY

Differentiation of COVID-19 pneumonia from other lung diseases using CT radiomic features and machine learning: A large multicentric cohort study

Isaac Shiri[1] | Yazdan Salimi[1] | Abdollah Saberi[1] | Masoumeh Pakbin[2] | Ghasem Hajianfar[1] | Atlas Haddadi Avval[3] | Amirhossein Sanaat[1] | Azadeh Akhavanallaf[1] | Shayan Mostafaei[4,5] | Zahra Mansouri[1] | Dariush Askari[6] | Mohammadreza Ghasemian[7] | Ehsan Sharifipour[8] | Saleh Sandoughdaran[9] | Ahmad Sohrabi[10] | Elham Sadati[11] | Somayeh Livani[12] | Pooya Iranpour[13] | Shahriar Kolahi[14] | Bardia Khosravi[15] | Maziar Khateri[16] | Salar Bijari[11] | Mohammad Reza Atashzar[17] | Sajad P. Shayesteh[18] | Mohammad Reza Babaei[19] | Elnaz Jenabi[20] | Mohammad Hasanian[21] | Alireza Shahhamzeh[22] | Seyed Yaser Foroghi Ghomi[22] | Abolfazl Mozafari[23] | Hesamaddin Shirzad-Aski[24] | Fatemeh Movaseghi[23] | Rama Bozorgmehr[25] | Neda Goharpey[26] | Hamid Abdollahi[27,28] | Parham Geramifar[20] | Amir Reza Radmard[29] | Hossein Arabi[1] | Kiara Rezaei-Kalantari[30] | Mehrdad Oveisi[31,32] | Arman Rahmim[27,28,33] | Habib Zaidi[1,34,35,36]

Correspondence
Habib Zaidi, Geneva University Hospital, Division of Nuclear Medicine and Molecular Imaging, CH-1211 Geneva, Switzerland.
Email: habib.zaidi@hcuge.ch

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: SNSF 320030_176052

Abstract

To derive and validate an effective machine learning and radiomics-based model to differentiate COVID-19 pneumonia from other lung diseases using a large multi-centric dataset. In this retrospective study, we collected 19 private and five public datasets of chest CT images, accumulating to 26 307 images (15 148 COVID-19; 9657 other lung diseases including non-COVID-19 pneumonia, lung cancer, pulmonary embolism; 1502 normal cases). We tested 96 machine learning-based models by cross-combining four feature selectors (FSs) and eight dimensionality reduction techniques with eight classifiers. We trained and evaluated our models using three different strategies: #1, the whole dataset (15 148 COVID-19 and 11 159 other); #2, a new dataset after excluding healthy individuals and COVID-19 patients who did not have RT-PCR results (12 419 COVID-19 and 8278 other); and #3 only non-COVID-19 pneumonia patients and a

For affiliations refer to page 14

random sample of COVID-19 patients (3000 COVID-19 and 2582 others) to provide balanced classes. The best models were chosen by one-standard-deviation rule in 10-fold cross-validation and evaluated on the hold out test sets for reporting. In strategy#1, Relief FS combined with random forest (RF) classifier resulted in the highest performance (accuracy = 0.96, AUC = 0.99, sensitivity = 0.98, specificity = 0.94, PPV = 0.96, and NPV = 0.96). In strategy#2, Recursive Feature Elimination (RFE) FS and RF classifier combination resulted in the highest performance (accuracy = 0.97, AUC = 0.99, sensitivity = 0.98, specificity = 0.95, PPV = 0.96, NPV = 0.98). Finally, in strategy #3, the ANOVA FS and RF classifier combination resulted in the highest performance (accuracy = 0.94, AUC =0.98, sensitivity = 0.96, specificity = 0.93, PPV = 0.93, NPV = 0.96). Lung radiomic features combined with machine learning algorithms can enable the effective diagnosis of COVID-19 pneumonia in CT images without the use of additional tests.

KEYWORDS

computed tomography, COVID-19, differential diagnosis, machine learning, radiomics

1 | INTRODUCTION

Different diagnostic methods have been proposed for SARS-CoV-2, the virus that causes coronavirus disease 2019, also known as COVID-19.[1–3] Medical imaging modalities, including chest x-rays (CXR) and computed tomography (CT) scanning, have a key role in the diagnosis, prognosis, and treatment planning of COVID-19 pneumonia.[1–3] CXR is regarded as the fastest, non-invasive, widely available, and cost-effective modality for the assessment of pulmonary lesions and diseases.[4] In the context of COVID-19, some characteristics, such as ground glass or patchy opacities, can be linked with COVID-19 pneumonia in a CXR.[4,5] Nonetheless, discovering these findings requires expertise as these changes are subtle, leading to the low sensitivity of CXR.[5] CT images, on the other hand, better demonstrate the opacities and are useful even for the early diagnosis of COVID-19 pneumonia in asymptomatic/pre-symptomatic patients.[6] Nevertheless, CT shows significantly higher sensitivity and diagnostic accuracy compared to CXR (0.85 and 0.56, respectively) while being similar in terms of low specificity (0.50 and 0.60, respectively).[7,8] Hence, it is better suited in the triage of patients if a scanner is available and the higher dose to the patient from the CT scan is carefully considered.[8]

A wide range of studies reported on the clinical use of chest CT and associated qualitative assessment and quantitative analysis for the diagnosis and management of COVID-19 patients.[9,10] A number of qualitative/semi-quantitative findings in chest CT, such as the presence and laterality of ground-glass opacities and consolidation, the number of lobes affected, and the extent of lung involvement, are used as acceptable features for COVID-19 diagnosis.[11] Although the sensitivity of CT imaging might be high for the assessment of COVID-19 pneumonia, it lacks the ability to make a confident diagnosis as it suffers from a low specificity since these findings rely on physician's inference and are most likely to be subjective.[12] Computerized tools have been proposed as a solution to address the aforementioned limitations of CT with aiming to provide tools that can visualize and extract the most subtle and minute characteristics of images.

Radiomics, a high-level image analysis technique that mines multi-dimensional data from medical images,[13–16] has emerged in the past decade and has been used more recently in evaluating COVID-19 based on CT images.[17] Studies reported on the feasibility of deep learning (DL)/radiomics applied to CT images toward classification (e.g., COVID-19 pneumonia, non-COVID-19 pneumonia, other lung diseases, or normal images). Harmon et al.[18] trained a DL-based neural network and validated it on a cohort of 1337 patients. Bai et al.[19] included 1186 patients and evaluated an artificial intelligence (AI)-based model to classify CT images into COVID-19 and non-COVID-19 pneumonia. Zhang et al.[20] conducted a study on 3777 patients and evaluated an AI system with the addition of clinical data to determine whether an image reflects COVID-19 pneumonia, other pneumonia, or normal. Di et al.[21] utilized radiomic features to assess the differentiation of COVID-19 pneumonia and community-acquired pneumonia in 3330 patients. Xie et al.[22] evaluated the ability of radiomics features and ground-glass opacities in 301 patients for the discrimination of COVID-19 and non-COVID-19 patients.

In addition, several studies were conducted at the intersection era of COVID-19 to develop CT-based radiomic signatures for diagnostic and predictive purposes with the aim to improve clinical decision-making.[17,23–32] Most aforementioned studies utilized datasets consisting of the dichotomy of COVID-19 pneumonia versus other non-COVID-19 pneumonia or COVID-19 vs. normal patients. However, some CXR or CT studies included other lung diseases as well. For instance, Albahli et al.[23] included 14 classes of diseases (e.g., cancer, pneumothorax, fibrosis, edema, atelectasis, etc.) in their dataset and trained their model on CXR images. Das et al.[33] also included tuberculosis patients and added their images to COVID-19 pneumonia, other pneumonia, and healthy CXR images of their dataset. Wang et al.[34] studied a dataset consisting of COVID-19 and other lung disease images in addition to normal lung images.

Several other AI-based algorithms, such as machine learning (ML) and deep learning (DL), have been proposed for the tasks of COVID-19 detection, diagnosis, or prognostication; however, the majority of these studies face limitations according to a systematic review by Roberts et al.,[35] which included more than 2000 original articles focusing on the development of different DL/ML-based algorithms in the diagnosis/prognosis of COVID-19 patients. First and foremost is data bias; many articles have used datasets with small sample sizes, duplicate samples, and low-quality or non-standardized medical image formats. Moreover, many researchers have studied the so-called Frankenstein[35] and Toy[36] datasets, utilizing small and/or low-quality images assembled and redistributed from other datasets.[35] In addition, most studies have not provided sufficient information regarding data preprocessing and demographics of training/testing cohorts and did not provide public access to code/data. Roberts et al.[35] reported that most articles in the diagnostic era failed to balance the number of COVID-19 and other classes of diseases in the training and testing datasets. For example, one study may have included considerably fewer COVID-19 cases compared to other cases.[35] Regarding the methodology, most of the studies failed to elucidate an exact methodology which is a must for conducting reproducible research.[35] Hence, few studies are practical in real clinical situations.[35]

In the present study, we have gathered multi-centric CT image datasets of 26 307 patients containing COVID-19, non-COVID-19 pneumonia, and other lung pathologies.[37] We aimed to assess the value of systematically utilizing CT-based radiomics and multiple dimensionality reductions, feature selection (FS), and classification algorithms to distinguish COVID-19 pneumonia from non-COVID-19 pneumonia and other lung diseases. To this end, we have assembled and utilized a very large, curated dataset and applied different ML models consistent with prognostic/diagnostic modeling guidelines.

2 | MATERIALS AND METHODS

The methodological steps adopted in this study can be found in Figure 1. We filled out several standardized

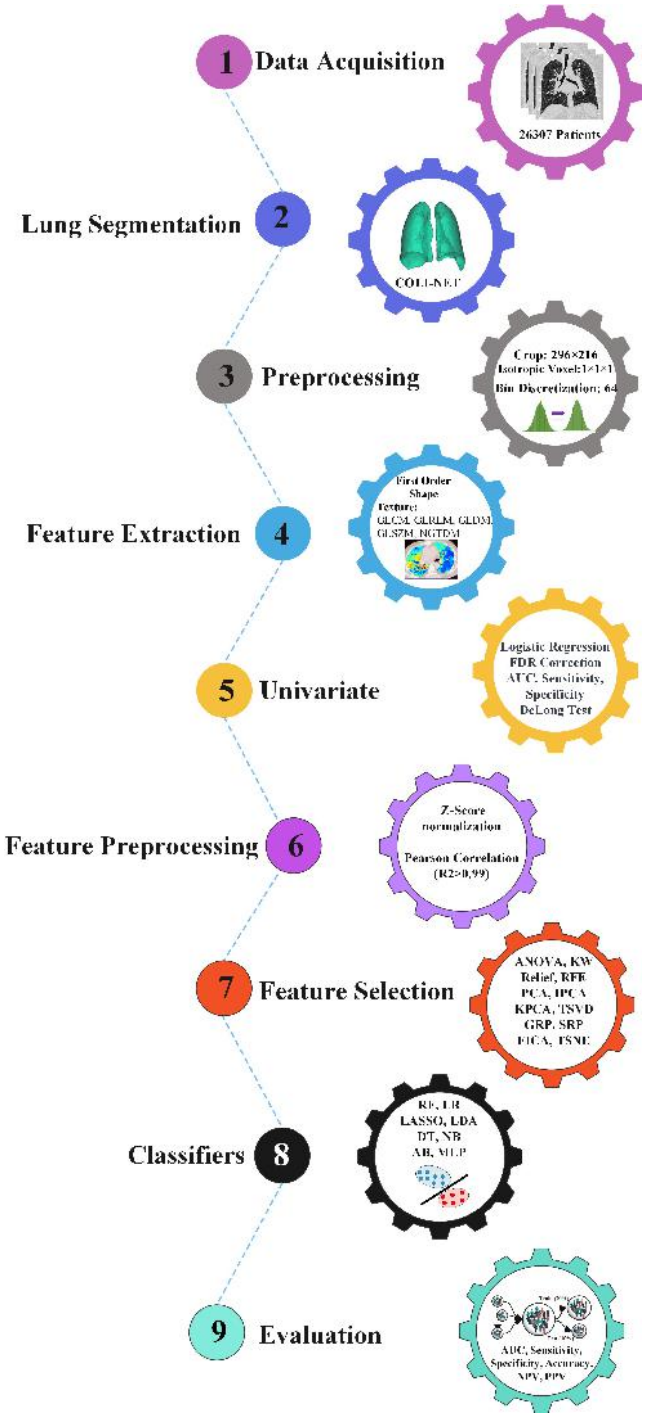

FIGURE 1 Flowchart of the different steps implemented in this study, including data acquisition, segmentation using COLI-NET pipeline, image preprocessing, and feature extraction. Univariate analysis was performed for each feature, followed by feature preprocessing performed on feature sets, and feature selection and classifier algorithms applied to these features.

checklists regarding diagnostic modeling (TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis[38]) and AI in medical image analysis (CLAIM, Checklist for Artificial Intelligence in Medical imaging[39]) to ensure the reproducibility and decency of our study. The complete checklist of standardizations was provided in the supplementary files.

2.1 | Datasets

The data of this study consisted of 19 local and five public datasets containing both COVID-19 and other lung diseases, arriving at 26307 images (15 148 COVID-19; 9657 other lung diseases including non-COVID-19 pneumonia, lung cancer, and pulmonary embolism; and 1502 normal). Our public dataset consisted of five separate datasets, including COVID-19 ($n = 1744$),[18,40,41] pulmonary embolism (PE, $n = 5696$),[42] and lung cancer ($n = 1379$)[40] CT images. Additional information about public datasets is provided in.[18,40–43] Our private datasets were assembled in this effort from 19 clinical centers in Iran, totaling 13 404 COVID-19 patients with the same number of CT images (one image per patient). The approval for our research was granted by the institutional ethical review boards, and due to the study's retrospective design, the requirement for obtaining written informed consent was exempted. All public and private datasets consisted of 3D CT images.

In this study, we included patients with confirmed COVID-19 pneumonia ($n = 15\ 148$),[44] patients with confirmed other types of pneumonia ($n = 2582$), pulmonary embolism (PE) ($n = 5696$), lung cancer ($n = 1379$), and normal healthy population ($n = 1502$). The confirmation of COVID-19 pneumonia was either based on positive RT-PCR or a consensus opinion of two experienced radiologists regarding COVID-19 manifestation in CT images (in compliance with CO-RADS).[2,44] If the two radiologists reached a discrepancy, a third experienced radiologist gave the final judgment.[44]

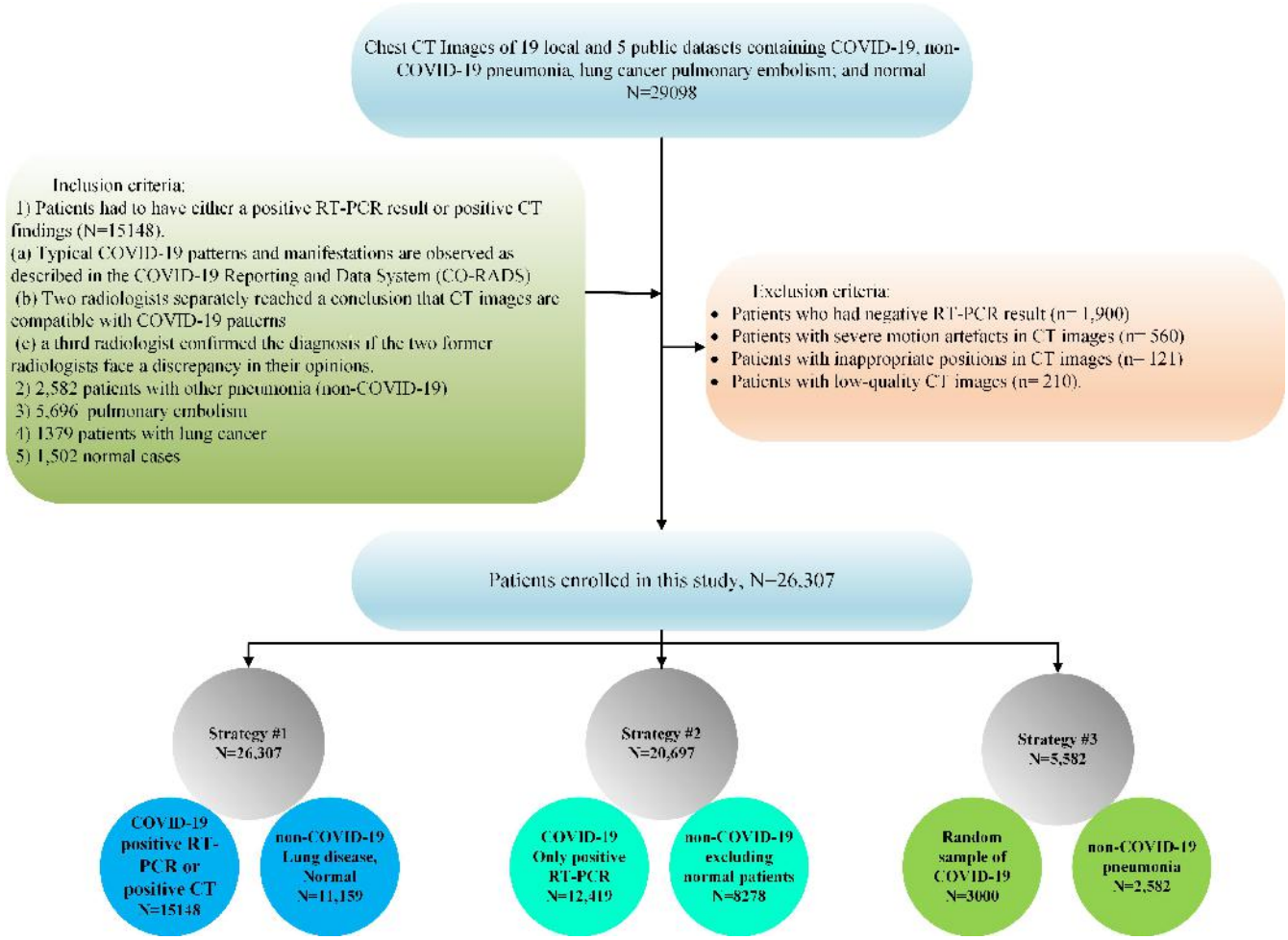

FIGURE 2 Flowchart of patient enrollment along with their inclusion and exclusion criteria. Altogether, 29 098 patients with various lung diseases were enrolled in this study, 2791 patients were excluded, and 26 307 patients were used for further analysis.

Altogether, 29 098 patients with various lung diseases were enrolled in this study. The inclusion and exclusion criteria are summarized in Figure 2. The exclusion criteria for patients were as follows: (1) patients who had negative RT-PCR results ($n = 1900$), (2) patients presenting with severe motion artifacts in CT images confirmed by an experienced medical physicist ($n = 560$), (3) patients with inappropriate positions in CT images ($n = 121$), and (4) patients with low-quality CT images ($n = 210$). Overall, we excluded 2791 patients because of the above-mentioned reasons, and 26 307 patients were used for further analysis.

The private centers acquired all chest CT images using an institutional variant of COVID-19 imaging criteria provided by an Iran national society of Radiology.[45] To decrease motion artifacts, image acquisition was performed in breath-hold mode. Each center's image acquisition features are provided in.[44]

2.2 | Image segmentation, preprocessing, and feature extraction

The images were segmented automatically using a previously developed deep learning algorithm called COLI-Net[46] and validated[29,44,47,48] DL algorithm. Images and segmentations were reviewed and edited in case of mis-segmentation. The images were first cropped to attain the lung-only region box, then resized to 296×216. Next, the image voxel size was resized to $1 \times 1 \times 1\ mm^3$ followed by 64-bin size discretization. Radiomics feature extraction was performed using the Pyradiomcis library,[49] which has been standardized according to the image biomarker standardization initiative (IBSI).[50,51] The 105 extracted features included shape ($n = 14$), first-order histogram ($n = 16$), second-order gray level co-occurrence matrix (GLCM, $n = 24$), and higher-order features including gray level dependence matrix (GLDM, $n = 14$), gray level size zone matrix (GLSZM, $n = 16$), gray level run length matrix (GLRLM, $n = 16$), and neighboring gray tone difference matrix (NGTDM, $n = 5$).

2.3 | Univariate analysis

In our univariate analysis, we calculated the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity for test sets using the logistic regression model, which was performed on each feature in three strategies. The Delong test was implemented to show if any significant difference exists between the AUCs of train and test sets.

2.4 | Feature preprocessing, feature selection, and classifiers

For Z-Score normalization, the mean and standard deviation were calculated in the training sets and then applied to testing sets for each feature. Features with high correlation ($R^2 > 0.99$) were eliminated using Pearson correlation. Four feature selection (FS) algorithms, including analysis of variance (ANOVA), Kruskal–Wallis (KW), recursive feature elimination (RFE), and relief, alongside eight dimensionality reduction techniques, including principal component analysis (PCA), incremental PCA (IPCA), Kernel PCA (KPCA), truncated SVD (TSVD), Gaussian random projection (GRP), sparse random projection (SRP), fast ICA (FICA), and t-distributed stochastic neighbor embedding (TSNE) were implemented. Eight classifiers, including logistic regression (LR), least absolute shrinkage and selection operator (LASSO), linear discriminant analysis (LDA), decision tree (DT), random forest (RF), AdaBoost (AB), Naïve Bayes (NB), and multilayer perceptron (MLP) were implemented for classification. In total, we tested 96 different combinations via cross-combining four FSs and eight dimensionality reduction techniques (12 in total) with eight classifiers.

2.5 | Evaluation

We trained and evaluated our models in three different strategies. First of all, the entire dataset (26 307 patients, including 15 148 COVID-19 and 11 159 non-COVID-19) being randomly split into 70% (18 414 patients) and 30% (7893 patients) for the training and test sets, respectively. Our dataset encompassed both categories of patients: those with a confirmed diagnosis of COVID-19 via reverse transcription polymerase chain reaction (RT-PCR) and those whose diagnoses were through CT imaging. Second, excluding normal patients in other lung disease classes and only including RT-PCR-positive COVID-19 cases in the COVID-19 class, the resulting dataset (20 697 patients, including 12 419 COVID-19 and 8278 non-COVID-19) was randomly split into 70% (14488 patients) and 30% (6209 patients) for the training and test sets, respectively. In the third strategy, only including non-COVID-19 pneumonia patients and a random sample of COVID-19 patients (5582 patients including 3000 COVID-19 and 2582 non-COVID-19) to provide a balanced dataset, and then randomly split the dataset into 70% (3907 patients) and 30% (1675 patients) for the training and test sets, respectively. Figure 2 summarizes the inclusion, exclusion, and each strategy dataset.

The steps for multivariate analysis, including feature preprocessing, FS, and classification, were performed

separately for each strategy. The model's parameters and hyper-parameters were optimized during the training phase using grid search algorithms in 70% of the data sets for three strategies as training sets. A one-standard-deviation rule chose the best models in 10-fold cross-validation on train sets (70% of datasets in three strategies) and the final results reported in the hold-out test sets (untouched during preprocessing, feature selection, and classifier training). Accuracy, sensitivity, specificity, AUC, positive predictive value (PPV), and negative predicted value (NPV) were reported for the test set. Statistical comparison of AUCs between the different models in the test sets was performed by the DeLong test[52] to ascertain the best performances. The significance level of 0.05 was considered for statistical tests. All multivariate machine learning analyses were performed in the Python open-source library Scikit-Learn[53].

3 | RESULTS

Supplemental Figures S1–3 and S4–6 show the hierarchical clustering heat map and TSNE visualization of the radiomic features distribution for COVID-19 and lung

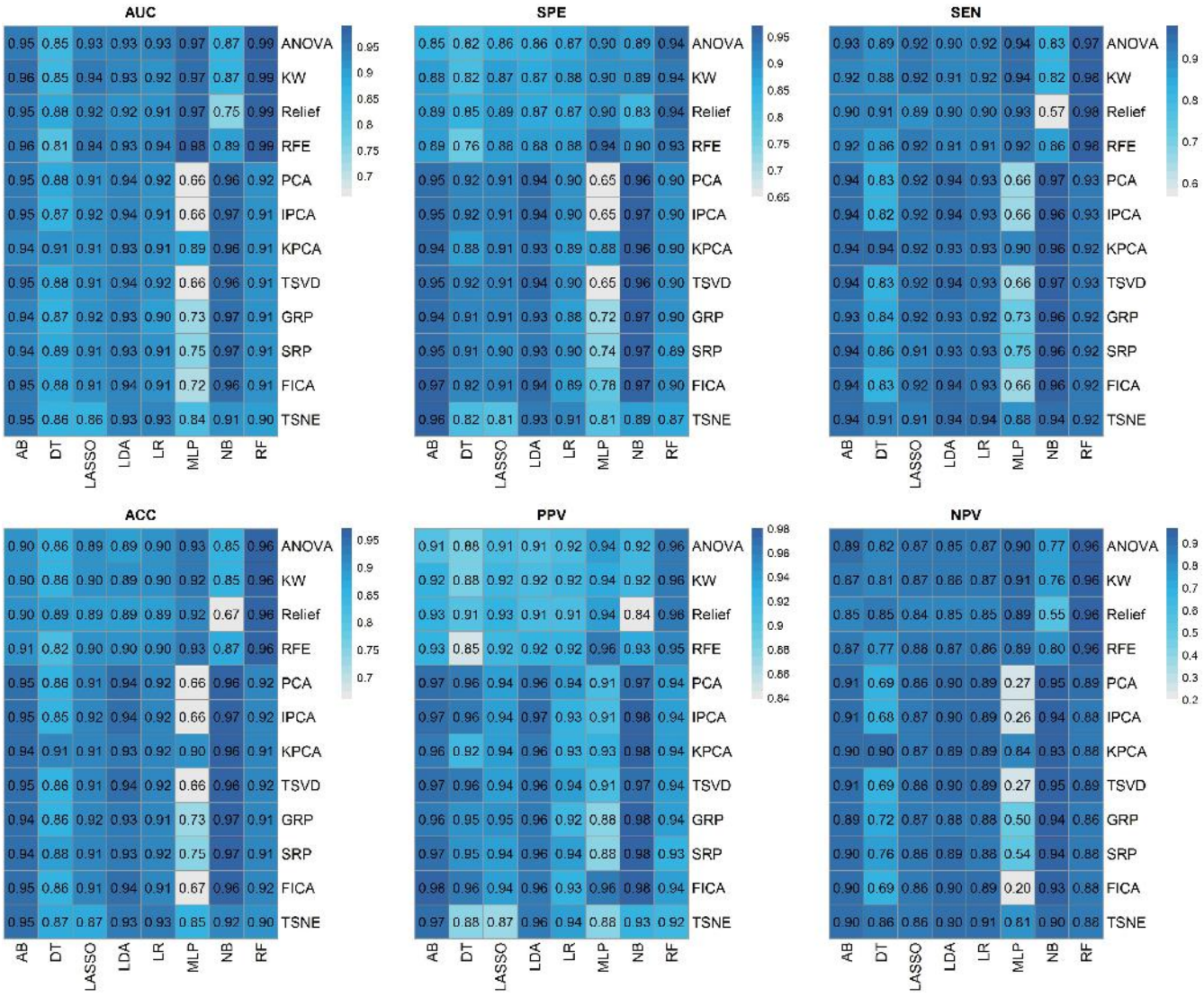

FIGURE 3 Heatmaps of the cross-combinations of feature selectors (12 rows) and classifiers (8 columns) for AUC, sensitivity, specificity, accuracy, PPV, and NPV in strategy #1. ACC: Accuracy, SEN: sensitivity, SPE: Specificity, AUC: area under the receiver operating characteristic curve positive predictive value (PPV), and negative predicted value (NPV). The feature selectors including, Analysis of Variance (ANOVA), Kruskal-Wallis (KW), Recursive Feature Elimination (RFE), Relief, Principal Component Analysis (PCA), Incremental PCA (IPCA), Kernel PCA (KPCA), Truncated SVD (TSVD), Gaussian Random Projection (GRP), Sparse Random Projection (SRP), Fast ICA (FICA), and t-Distributed Stochastic Neighbor Embedding (TSNE). Classifiers include Logistic Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), Linear Discriminant Analysis (LDA), Decision Tree (DT), Random Forest (RF), AdaBoost (AB), Naïve Bayes (NB), and Multilayer Perceptron (MLP).

disease classes for strategies #1, #2, and #3, respectively. According to these results, we did not find any cluster in an unsupervised manner based on centers/dataset. In addition, TSNE visualization did not show any meaningful cluster of datasets. The correlation of radiomic features is depicted in Supplemental Figures S7–9 for strategies #1, #2, and #3, respectively. Prior to multivariate analysis, highly correlated features (R^2 >0.99) were identified, and redundant features were eliminated, enabling dimensionality reduction in an unsupervised manner via machine learning algorithms.

3.1 | Univariate outcome

Supplemental Tables S1–3 summarize univariate analysis for each feature in strategies #1, #2, and #3, respectively. Using binary logistic regression, we reported AUC, accuracy, sensitivity, specificity, and *p*-values (DeLong test) for each feature of the train and test sets. When comparing AUCs, most of the features had a *p*-value > 0.05 for the training and test sets using the DeLong test. In both strategies 1 and 2, two features, including Zone Variance and Large Area Emphasis from GLSZM (AUC = 0.85/0.84, accuracy = 0.79/0.79, sensitivity = 0.77/0.76/, specificity = 0.82/0.83), subsets had highest AUC, accuracy, sensitivity, and specificity. In strategy 3, Joint Entropy from GLRLM (AUC = 0.83, accuracy = 0.79, sensitivity = 0.84, and specificity = 0.73) and Gray Level Non-Uniformity Normalized from GLRLM (AUC = 0.83, accuracy = 0.78, sensitivity = 0.78, and specificity = 0.78) subsets had highest AUC, accuracy, sensitivity, and specificity.

3.2 | Multivariate outcome

3.2.1 | Strategy #1

For strategy #1, Figure 3 presents the heatmap of the cross-combinations of FSs and classifiers for different

TABLE 1 Mean value of different quantitative parameters in different feature selectors and classifiers for strategy 1.

		AUC	ACC	SEN	SPE	PPV	NPV
Classifier	LDA	0.93	0.92	0.93	0.91	0.95	0.88
	MLP	0.82	0.80	0.80	0.79	0.92	0.61
	RF	0.94	0.93	**0.94**	0.91	0.94	**0.91**
	LR	0.92	0.91	0.92	0.89	0.93	0.88
	LASSO	0.92	0.90	0.92	0.89	0.93	0.86
	AB	**0.95**	**0.93**	0.93	**0.93**	**0.95**	0.89
	DT	0.87	0.86	0.87	0.87	0.92	0.77
	NB	0.92	0.91	0.90	0.93	0.95	0.86
Feature selection	ANOVA	0.93	0.90	0.91	0.87	0.92	0.87
	KW	0.93	0.90	0.91	0.88	0.92	0.86
	Relief	0.91	0.87	0.87	0.88	0.92	0.83
	RFE	**0.93**	0.90	0.91	0.88	0.92	0.86
	PCA	0.89	0.89	0.89	0.89	0.95	0.80
	IPCA	0.89	0.89	0.89	0.89	0.95	0.79
	KPCA	0.92	**0.92**	**0.93**	**0.91**	0.95	**0.89**
	TSVD	0.89	0.89	0.89	0.89	0.95	0.80
	GRP	0.90	0.90	0.89	0.90	0.94	0.82
	SRP	0.90	0.90	0.90	0.90	0.94	0.83
	FICA	0.90	0.89	0.89	0.91	**0.96**	0.78
	TSNE	0.90	0.90	0.92	0.88	0.92	0.88

Note: ACC: accuracy, SEN: sensitivity, SPE: specificity, AUC: area under the receiver operating characteristic curve positive predictive value (PPV), and negative predicted value (NPV). Feature selectors including analysis of variance (ANOVA), Kruskal–Wallis (KW), recursive feature elimination (RFE), relief, principal component analysis (PCA), incremental PCA (IPCA), Kernel PCA (KPCA), truncated SVD (TSVD), Gaussian random projection (GRP), sparse random projection (SRP), fast ICA (FICA), and t-distributed stochastic neighbor embedding (TSNE). Classifiers, including logistic regression (LR), least absolute shrinkage and selection operator (LASSO), linear discriminant analysis (LDA), decision tree (DT), random forest (RF), AdaBoost (AB), Naïve Bayes (NB), and multilayer perceptron (MLP).

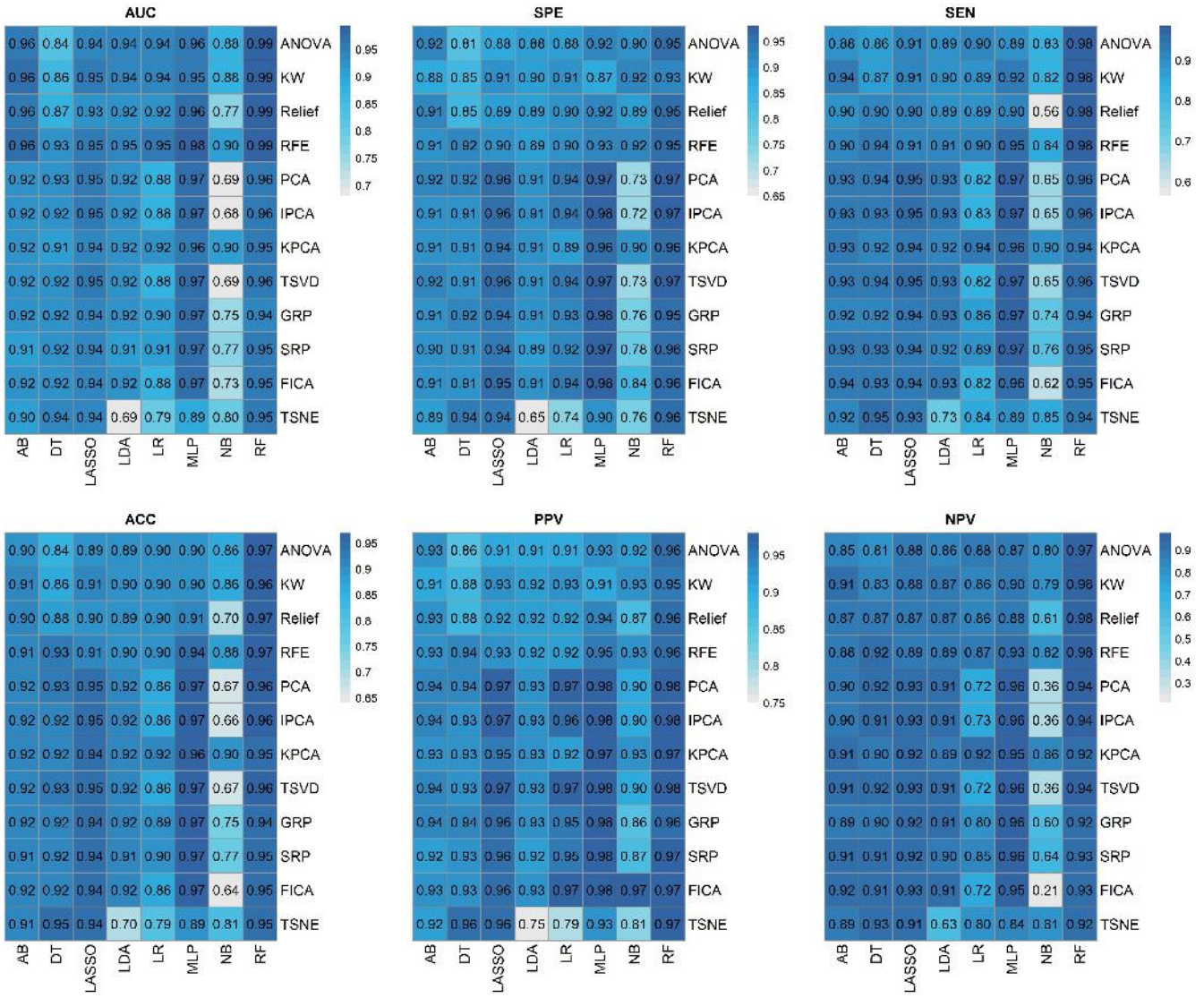

FIGURE 4 Heatmaps of the cross-combinations of feature selectors (12 rows) and classifiers (8 columns) for AUC, sensitivity, specificity, accuracy, PPV, and NPV in strategy #2. ACC: Accuracy, SEN: sensitivity, SPE: Specificity, AUC: area under the receiver operating characteristic curve positive predictive value (PPV), and negative predicted value (NPV). The feature selectors including, Analysis of Variance (ANOVA), Kruskal-Wallis (KW), Recursive Feature Elimination (RFE), Relief, Principal Component Analysis (PCA), Incremental PCA (IPCA), Kernel PCA (KPCA), Truncated SVD (TSVD), Gaussian Random Projection (GRP), Sparse Random Projection (SRP), Fast ICA (FICA), and t-Distributed Stochastic Neighbor Embedding (TSNE). Classifiers include Logistic Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), Linear Discriminant Analysis (LDA), Decision Tree (DT), Random Forest (RF), AdaBoost (AB), Naïve Bayes (NB), and Multilayer Perceptron (MLP).

quantitative parameters. Relief FS and RF classifier combination resulted in the highest performance (AUC = 0.99, sensitivity = 0.98, specificity = 0.94, accuracy = 0.96, PPV = 0.96, and NPV = 0.96). Table 1 presents the mean value of different quantitative parameters in different FSs and classifiers for strategy #1. Results show that the AB classifier across AUC, ACC, SPE, PPV, and RF classifier across SEN and NPV had higher mean performance (considering all metrics). The RFE (AUC), KPCA (ACC, SEN, SPE, and NPV), and FICA (PPV) had the highest mean performance among FS methods. Supplemental Figure S10 presents a box plot of FSs and classifiers for the different evaluation metrics in strategy #1.

3.2.2 | Strategy #2

For strategy #2, Figure 4 presents the heatmap of the cross-combinations of FSs and classifiers for different quantitative parameters. The RFE FS and RF classifier combination resulted in the highest performance

TABLE 2 Mean value of different quantitative parameters in different feature selectors and classifiers for strategy 2.

		AUC	ACC	SEN	SPE	PPV	NPV
Classifier	LDA	0.91	0.89	0.90	0.88	0.91	0.87
	MLP	0.96	0.94	0.94	0.95	0.96	0.93
	RF	**0.97**	**0.96**	**0.96**	**0.96**	**0.97**	**0.95**
	LR	0.90	0.88	0.87	0.90	0.93	0.81
	LASSO	0.94	0.93	0.93	0.93	0.95	0.91
	AB	0.93	0.91	0.92	0.91	0.93	0.90
	DT	0.91	0.91	0.92	0.90	0.92	0.89
	NB	0.79	0.76	0.74	0.82	0.90	0.60
Feature selection	ANOVA	0.93	0.89	0.89	0.89	0.92	0.87
	KW	0.93	0.90	0.90	0.90	0.92	0.88
	Relief	0.92	0.88	0.87	0.90	0.92	0.85
	RFE	**0.95**	0.92	0.92	0.92	0.93	0.90
	PCA	0.90	0.90	0.89	0.92	0.95	0.83
	IPCA	0.90	0.90	0.89	0.91	0.95	0.83
	KPCA	0.93	**0.93**	**0.93**	0.92	0.94	**0.91**
	TSVD	0.90	0.90	0.89	0.91	0.95	0.83
	GRP	0.91	0.91	0.90	0.91	0.94	0.86
	SRP	0.91	0.91	0.91	0.91	0.94	0.88
	FICA	0.90	0.89	0.89	**0.93**	**0.96**	0.81
	TSNE	0.86	0.87	0.88	0.85	0.89	0.84

Note: ACC: accuracy, SEN: sensitivity, SPE: specificity, AUC: area under the receiver operating characteristic curve positive predictive value (PPV), and negative predicted value (NPV). Feature selectors including analysis of variance (ANOVA), Kruskal –Wallis (KW), recursive feature elimination (RFE), relief, principal component analysis (PCA), incremental PCA (IPCA), kernel PCA (KPCA), truncated SVD (TSVD), gaussian random projection (GRP), sparse random projection (SRP), fast ICA (FICA), and t-distributed stochastic neighbor embedding (TSNE). Classifiers, including logistic regression (LR), least absolute shrinkage and selection operator (LASSO), linear discriminant analysis (LDA), decision tree (DT), random forest (RF), AdaBoost (AB), Naïve Bayes (NB), and multilayer perceptron (MLP).

(AUC = 0.99, sensitivity = 0.98, specificity = 0.95, accuracy = 0.97, PPV = 0.96, NPV = 0.98). Table 2 presents the mean value of different quantitative parameters in different FSs and classifiers for strategy #2. Results indicated that the RF classifier had higher mean performance than other methods across all metrics (considering all metrics). Nevertheless, the best FS method was different based on each metric, including RFE (AUC), KPCA (ACC, SEN, and NPV), and FICA (SPE and PPV). Supplemental Figure S11 represents a box plot of FSs and classifiers for the different evaluation metrics in strategy #2.

3.2.3 | Strategy #3

For strategy #3, Figure 5 presents the heatmap of the cross-combinations of FSs and classifiers for different quantitative parameters. ANOVA FS and RF classifier combination resulted in the highest performance (AUC = 0.98, sensitivity = 0.96, specificity = 93, accuracy = 0.94, PPV = 0.93, NPV = 0.96). Table 3 presents the mean value of different quantitative parameters in different FSs and classifiers for strategy #3. RF (AUC, ACC, SEN, NPV) and MLP (SPE and PPV) classifiers and RFE FS across all metrics had higher mean performance than other methods. Supplemental Figure S12 represents a box plot of FSs and classifiers for the different evaluation metrics in strategy #3. Figure 6 depicts the receiver operating characteristic curve for three strategies (for the best models in each strategy).

3.3 | Supplemental data

Supplemental Figures S13–15 represent the statistical comparison of AUC between the different models using the DeLong test in strategies #1, #2, and #3, respectively. The combinations of RF classifiers with ANOVA, KW, Relief, and RFE FSs were significantly higher than other models in strategies #1 and #2 in terms of classification power. In strategy #3, several models, including the combinations of ANOVA FS with RF, LR, and

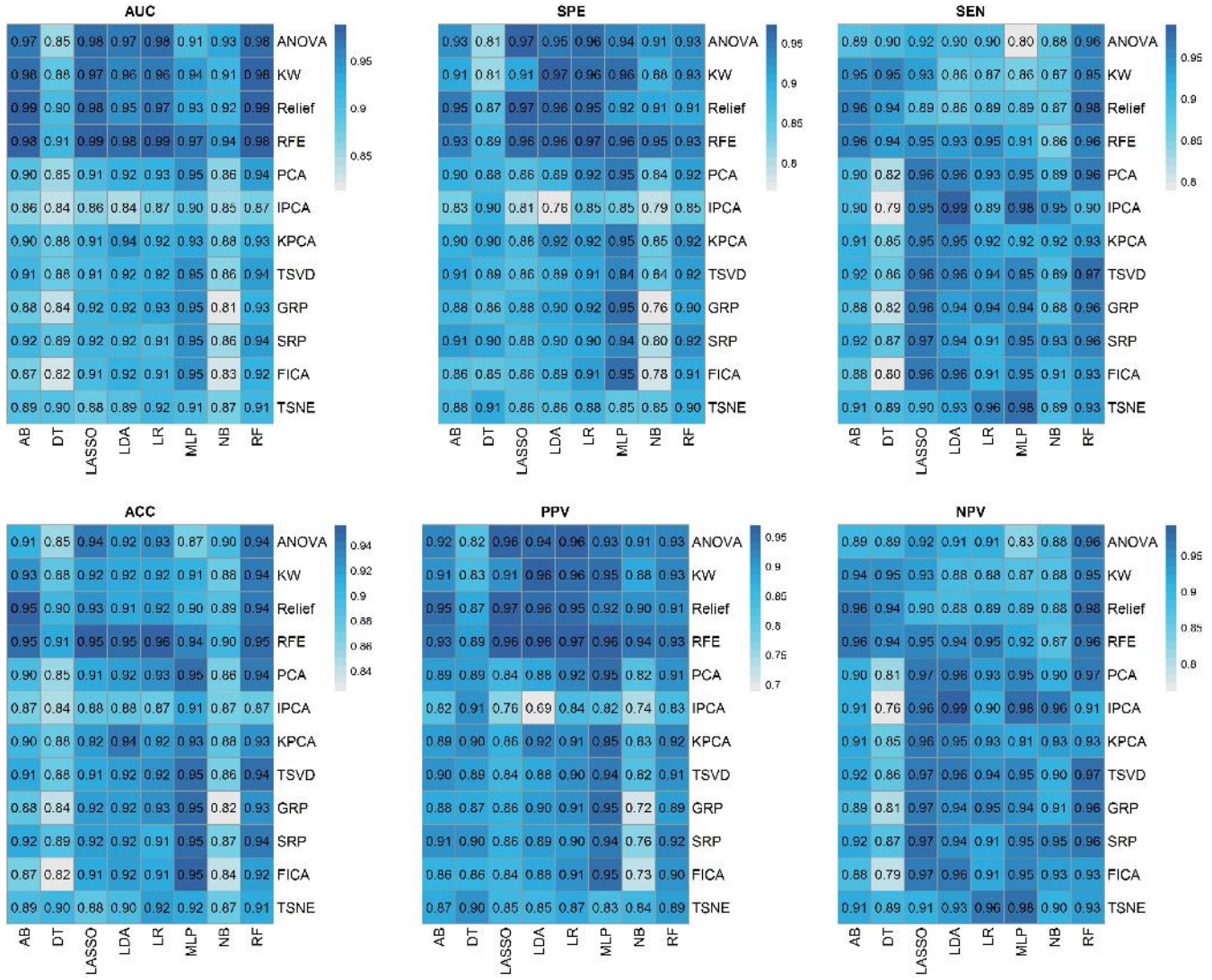

FIGURE 5 Heatmaps of the cross-combinations of feature selectors (12 rows) and classifiers (8 columns) for AUC, sensitivity, specificity, accuracy, PPV, and NPV in strategy #3. ACC: Accuracy, SEN: sensitivity, SPE: Specificity, AUC: area under the receiver operating characteristic curve positive predictive value (PPV), and negative predicted value (NPV). The feature selectors including, Analysis of Variance (ANOVA), Kruskal-Wallis (KW), Recursive Feature Elimination (RFE), Relief, Principal Component Analysis (PCA), Incremental PCA (IPCA), Kernel PCA (KPCA), Truncated SVD (TSVD), Gaussian Random Projection (GRP), Sparse Random Projection (SRP), Fast ICA (FICA), and t-Distributed Stochastic Neighbor Embedding (TSNE). Classifiers include Logistic Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), Linear Discriminant Analysis (LDA), Decision Tree (DT), Random Forest (RF), AdaBoost (AB), Naïve Bayes (NB), and Multilayer Perceptron (MLP).

LASSO classifiers, KW with AB and RF, Relief with AB, LASSO, and RF, and RFE with AB, LASSO, LDA, LR, and RF were significantly higher than other models.

4 | DISCUSSION

In this study, we implemented multiple machine learning algorithms to differentiate COVID-19 in a dataset consisting of a large number of patients (COVID-19, non-COVID-19 pneumonia, lung cancer, PE, and normal subjects). First, the entire lung was segmented using automated DL algorithms, followed by extracting radiomic features. Second, the features were normalized, redundant features were eliminated, and the remaining features were fed to FS algorithms and classifiers. Third, univariate results showed that some single features could differentiate between diseases when using different strategies with limited performance. Finally, we assessed if our models decently discriminated COVID-19 from other lung diseases and COVID-19 pneumonia from other

TABLE 3 Mean value of different quantitative parameters in different feature selectors and classifiers for strategy 3.

		AUC	ACC	SEN	SPE	PPV	NPV
Classifier	AB	0.92	0.91	0.91	0.90	0.89	0.92
	DT	0.87	0.87	0.87	0.87	0.88	0.86
	LASSO	0.93	0.92	0.94	0.89	0.88	0.95
	LDA	0.93	0.92	0.93	0.91	0.89	0.94
	LR	0.93	0.92	0.92	0.92	0.92	0.92
	MLP	0.94	0.93	0.92	**0.93**	**0.92**	0.93
	NB	0.88	0.87	0.90	0.85	0.83	0.91
	RF	**0.94**	**0.93**	**0.95**	0.91	0.91	**0.95**
Feature selection	ANOVA	0.95	0.91	0.89	0.92	0.92	0.90
	KW	0.95	0.91	0.91	0.92	0.92	0.91
	RFE	**0.97**	**0.94**	**0.93**	**0.94**	**0.94**	**0.93**
	Relief	0.95	0.92	0.91	0.93	0.93	0.91
	SRP	0.91	0.91	0.93	0.90	0.89	0.93
	TSVD	0.91	0.91	0.93	0.90	0.89	0.93
	KPCA	0.91	0.91	0.92	0.90	0.90	0.92
	PCA	0.91	0.91	0.92	0.90	0.89	0.92
	TSNE	0.90	0.90	0.92	0.88	0.86	0.93
	GRP	0.90	0.90	0.92	0.88	0.87	0.92
	FICA	0.89	0.89	0.91	0.88	0.87	0.91
	IPCA	0.86	0.87	0.92	0.83	0.80	0.92

Note: ACC: accuracy, SEN: sensitivity, SPE: specificity, AUC: area under the receiver operating characteristic curve positive predictive value (PPV), and negative predicted value (NPV). Feature selectors including analysis of variance (ANOVA), Kruskal–Wallis (KW), recursive feature elimination (RFE), Relief, principal component analysis (PCA), incremental PCA (IPCA), Kernel PCA (KPCA), truncated SVD (TSVD), Gaussian random projection (GRP), sparse random projection (SRP), fast ICA (FICA), and t-distributed stochastic neighbor embedding (tsne). Classifiers, including logistic regression (LR), least absolute shrinkage and selection operator (LASSO), linear discriminant analysis (LDA), decision tree (DT), random forest (RF), AdaBoost (AB), Naïve Bayes (NB), and multilayer perceptron (MLP).

pneumonia. In this light, in strategy #1, Relief FS combined with the RF classifier resulted in the best performance (Accuracy = 0.96, AUC = 0.99, sensitivity = 0.98, specificity = 0.94, PPV = 0.96, and NPV = 0.96). In strategy #2, RFE FS and RF classifier combinations resulted in the best performance (Accuracy = 0.97, AUC = 0.99, sensitivity = 0.98, specificity = 0.95, PPV = 0.96, NPV = 0.98). Finally, in strategy #3, the ANOVA FS and RF classifier combination resulted in the highest performance (Accuracy = 0.94, AUC =0.98, sensitivity = 0.96, specificity = 0.93, PPV = 0.93, NPV = 0.96).

A study by Fang et al.[54] showed that a radiomics model based on CT imaging features could differentiate COVID-19 pneumonia from other types of pneumonia with an AUC of 0.95. Their results also suggested that radiomics performs better than the clinical-only model. In another study, Tan et al.[55] demonstrated the efficacy of a radiomics-based model in discovering whether a patient has COVID-19 pneumonia or other types of pneumonia by analyzing the non-infectious areas of their CT scan. Their model achieved an AUC as high as 0.95 in the training and test datasets. Di et al.[21] also studied the diagnostic accuracy of CT-based radiomics in patients and reported that their hypergraph model could distinguish between community-acquired pneumonia and COVID-19 disease pneumonia.

Similar to most radiomic studies of COVID-19, we used chest CT images.[21,22,56] However, there is ongoing research exploiting other imaging modalities as well. For example, Bae et al.[57] evaluated the prediction ability of radiomics on chest x-rays of 514 patients taking advantage of other DL methods (radiomics feature map + DL + clinical features). They achieved AUCs of 0.93 and 0.90 in the prediction of mortality and the need for mechanical ventilators, respectively. In another study by Chandra et al.,[58] chest x-rays of 2088 (training set) and 258 (testing set) patients taken at baseline were assessed, and radiomics analysis was performed.

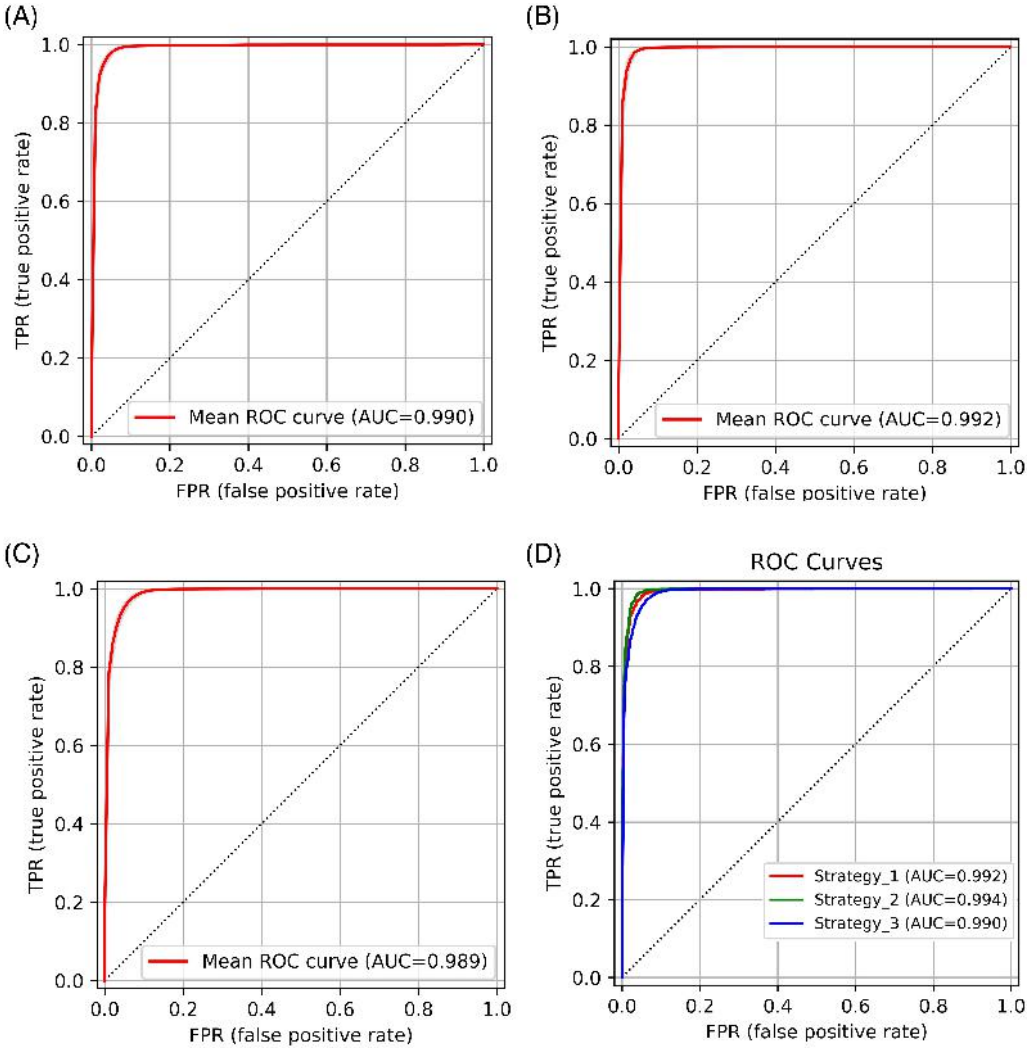

FIGURE 6 ROC curves for strategy #1 (A), strategy #2 (B), and strategy #3 (C). Confidence intervals are also plotted in gray with 10 000 bootstrapping. The confidence intervals are very small due to the large scale of our study. In addition, the results from strategies 1–3 are plotted all in one figure (D) to compare the different algorithms.

They reached an AUC of 0.95 in the test set for identifying normal, suspicious, and COVID-19 groups of patients.

Regarding the results of COVID-19 radiomics, literature review showed that multiple studies evaluated the ability of CT-based radiomic features to differentiate COVID-19 pneumonia from other types of pneumonia or other lung diseases. For instance, Velichko et al.[32] studied a total of 5759 chest CT image patches of confirmed COVID-19 pneumonia and 7958 CT images patches of pulmonary edema and proposed a model based on radiomic features to differentiate them. They achieved an AUC of 0.994, which performed better compared to other known neural networks. In another study, Peng et al.[31] assessed whether radiomic features derived from CT scans can aid in the differentiation of COVID-19 pneumonia from highly suspected but non-COVID-19 pneumonia in 145 patients. Their radiomics model performed well with an AUC of 0.99 and 0.97 on their training and testing dataset, which was significantly ($p < 0.001$) better than conventional CT parameters, such as the ratio of GGO.

Regarding the use of different lung pathologies in the dataset, a number of related studies have been conducted. Amyar et al.[30] dataset included healthy and COVID-19 patients with lung cancer and other lung pathologies. They achieved an AUC of 0.97 for the classification task. Wang et al.[34] utilized a PARL (prior attention residual learning)-based model for classifying CT images into COVID-19 pneumonia, other pneumonia, and non-pneumonic images and could achieve an AUC of 0.97 for COVID-19 discrimination. Chen et al.[59] included 422 patients who had COVID-19, other types of pneumonia, tuberculosis, and normal images in their dataset.

Their ResNet model achieved an accuracy of 91.2% for classifying images. Our study also included several lung pathologies, including COVID-19, pneumonia of other types (viral and bacterial), lung cancer, and subjects with normal lungs. Overall, we utilized a large-sized dataset consisting of CT images of patients from multiple institutions (multiple scanners/protocols) from different countries. The large-scale dataset used in our work helped with the generalizability and reproducibility of our model in contrast to the small-sized datasets of most previous studies.[20,60,61]

Our efforts faced a number of limitations, some of which were addressed and considered. First, motion artifacts are inescapable when patients are undergoing CT scanning.[44] Hence, we excluded these patients as they had an overlapping region of pneumonia in their chest image. Second, not all patients were tested for COVID-19 RT-PCR; some were included in our study only by their positive CT symptoms.[44] Hence, we attempted to overcome this limitation using different model evaluation strategies (e.g., choosing patients with positive RT-PCR as the testing set) to reach a reproducible model for further studies. Third, we did not include clinical or laboratory data. However, these data have been shown to be linked with CT image features.[20,62] Forth, image acquisition parameters were undoubtedly distinct in each center, hence affecting radiomic features.[44] Therefore, further studies should focus on harmonized features between the different centers.[63]

In this study, different diseases were gathered from various centers/databases/time-points where some information (scanner name, injection of contrast enhance material,...) was unavailable (removed during anonymization) owing to privacy issues. Therefore, data distribution by disease and center/time of acquisition could potentially bias classification results. Yet, CT is a quantitative imaging modality in which pixel values in Hounsfield units are in the same range for the different tissues. In addition, we performed different image preprocessing steps before feature extraction to harmonize the feature extraction across different CT imaging protocols. Furthermore, unsupervised clustering and TSNE visualization based on features were performed to ensure no bias regarding the datasets. However, we should remain cautious about potential hidden patterns in the dataset that these methods may not have uncovered, and these should be investigated in future studies. Classification metrics based on the different centers should be reported in a multi-centric study to visualize the variability across different scanners and image acquisition and reconstruction protocols. Even though we achieved a high AUC and good performance on other metrics, it is important to consider that a few hundred cases were detected as false positives and false negatives using the proposed method in each strategy. This factor should be taken into account for future clinical implementation. Finally, exploring interpretable and explainable AI models should be considered in future studies focusing on the diagnosis and prognosis of COVID-19.

5 | CONCLUSION

Our results support radiomics and machine learning capability to differentiate COVID-19 pneumonia versus normal lungs and several lung pathologies, including other types of pneumonia, lung cancer, and pulmonary embolism. Our study was conducted on a large heterogeneous cohort of patients/individuals with three different data curation techniques. FS algorithms chose the most robust radiomic features extracted from CT images and fed them into machine learning classifiers in order to diagnose diseases if any exist. Our framework was successfully implemented, producing appropriate diagnostic results, and further emphasizing the ability of radiomic features to enhance lung pathologies discrimination.

AUTHOR CONTRIBUTIONS

Isaac Shiri and Yazdan Salimi played roles in conceptualization, methodology, software development, validation, formal analysis, investigation, resource management, data curation, original draft writing, review and editing, visualization, and supervision. Abdollah Saberi, Masoumeh Pakbin, Ghasem Hajianfar, Atlas Haddadi Avval, Amirhossein Sanaat, Azadeh Akhavanallaf, Shayan Mostafaei, Zahra Mansouri, Dariush Askari, Mohammadreza Ghasemian, Ehsan Sharifipour, Saleh Sandoughdaran, Ahmad Sohrabi, and Elham Sadati were involved in conceptualization, methodology, validation, investigation, resource management, data curation, and writing for review and editing, along with visualization. Somayeh Livani, Pooya Iranpour, Shahriar Kolahi, Bardia Khosravi, Maziar Khateri, Salar Bijari, Mohammad Reza Atashzar, Sajad P. Shayesteh, Mohammad Reza Babaei, Elnaz Jenabi, Mohammad Hasanian, Alireza Shahhamzeh, and Seyed Yaser Foroghi Ghomi contributed to conceptualization, software development, formal analysis, resource management, and writing for review and editing, as well as visualization. Abolfazl Mozafari, Hesamaddin Shirzad-Aski, Fatemeh Movaseghi, Rama Bozorgmehr, Neda Goharpey, Hamid Abdollahi, Parham Geramifar, Hossein Arabi, Kiara Rezaei-Kalantari, and Mehrdad Oveisi participated in conceptualization, methodology, validation, investigation, resource management, data curation, original draft writing, review and editing,

visualization, supervision, project administration, and funding acquisition. Amir Reza Radmard, Arman Rahmim, and Habib Zaidi contributed to conceptualization, methodology, validation, investigation, resource management, data curation, original draft writing, review and editing, visualization, supervision, project administration, and funding acquisition.

AFFILIATIONS

[1]Division of Nuclear Medicine and Molecular Imaging, Geneva University Hospital, Geneva, Switzerland
[2]Imaging Department, Qom University of Medical Sciences, Qom, Iran
[3]School of Medicine, Mashhad University of Medical Sciences, Mashhad, Iran
[4]Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden
[5]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden
[6]Department of Radiology Technology, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[7]Department of Radiology, Shahid Beheshti Hospital, Qom University of Medical Sciences, Qom, Iran
[8]Neuroscience Research Center, Qom University of Medical Sciences, Qom, Iran
[9]Clinical Oncology Department, Royal Surrey Hospital, Guildford, United Kingdom
[10]Radin Makian Azma Mehr Ltd. Radinmehr Veterinary Laboratory, Gorgan, Iran
[11]Department of Medical Physics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran
[12]Clinical Research Development Unit (CRDU), Sayad Shirazi Hospital, Golestan University of Medical Sciences, Gorgan, Iran
[13]Department of Radiology, Medical Imaging Research Center, Shiraz University of Medical Sciences, Shiraz, Iran
[14]Department of Radiology, School of Medicine, Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Imam Khomeini Hospital, Tehran University of Medical Sciences, Tehran, Iran
[15]Digestive Diseases Research Center, Digestive Diseases Research Institute, Tehran University of Medical Sciences, Tehran, Iran
[16]Department of Medical Radiation Engineering, Science and Research Branch, Islamic Azad University, Tehran, Tehran, Iran
[17]Department of Immunology, School of Medicine, Fasa University of Medical Sciences, Fasa, Iran
[18]Department of Physiology, Pharmacology and Medical Physics, Alborz University of Medical Sciences, Karaj, Iran
[19]Department of Interventional Radiology, Firouzgar Hospital, Iran University of Medical Sciences, Tehran, Iran
[20]Research Centre for Nuclear Medicine, Tehran University of Medical Sciences, Tehran, Iran
[21]Department of Radiology, Arak University of Medical Sciences, Arak, Iran
[22]Clinical Research Development Center, Qom University of Medical Sciences, Qom, Iran
[23]Department of Medical Sciences, Qom Branch, Islamic Azad University, Qom, Iran
[24]Infectious Diseases Research Center, Golestan University of Medical Sciences, Gorgan, Iran
[25]Clinical Research Development Unit, Shohada-e Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[26]Department of Radiation Oncology, Shohada-e Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran
[27]Department of Radiology, University of British Columbia, Vancouver, BC, Canada
[28]Department of Integrative Oncology, BC Cancer Research Institute, Vancouver, BC, Canada
[29]Department of Radiology, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran
[30]Rajaie Cardiovascular, Medical & Research Center, Iran University of Medical Sciences, Tehran, Iran
[31]Comprehensive Cancer Centre, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences & Medicine, King's College London, London, UK
[32]Department of Computer Science, University of British Columbia, Vancouver, BC, Canada
[33]Departments of Physics and Biomedical Engineering, University of British Columbia, Vancouver, BC, Canada
[34]Department of Nuclear Medicine and Molecular Imaging, University of Groningen, University Medical Center Groningen, Groningen, Netherlands
[35]Department of Nuclear Medicine, University of Southern Denmark, Odense, Denmark
[36]University Research and Innovation Center, Óbuda University, Budapest, Hungary

ACKNOWLEDGMENTS

This work was supported by the Swiss National Science Foundation under grant SNSF 320030_176052. Open access funding provided by Universite de Geneve.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ORCID

Hamid Abdollahi https://orcid.org/0000-0003-0761-1309

Habib Zaidi https://orcid.org/0000-0001-7559-5297

REFERENCES

1. Aljondi R, Alghamdi S. Diagnostic value of imaging modalities for COVID-19: scoping review. *J Med Internet Res*. 2020;22(8):e19673.
2. Prokop M, van Everdingen W, van Rees VT, et al. CO-RADS: a categorical CT assessment scheme for patients suspected of having COVID-19-definition and evaluation. *Radiology*. 2020; 296(2):E97-E104.
3. D'Andrea A, Radmilovic J, Carbone A, et al. Multimodality imaging in COVID-19 patients: a key role from diagnosis to prognosis. *World J Radiol*. 2020;12(11):261-271.
4. Ke Q, Zhang J, Wei W, et al. A neuro-heuristic approach for recognition of lung diseases from X-ray images. *Expert Syst Appl*. 2019;126:218-232.
5. Kanne JP, Little BP, Chung JH, Elicker BM, Ketai LH. Essentials for radiologists on COVID-19: An update-radiology scientific expert panel. *Radiology*. 2020;296(2):E113-E114.
6. Varble N, Blain M, Kassin M, et al. CT and clinical assessment in asymptomatic and pre-symptomatic patients with early SARS-CoV-2 in outbreak settings. *Eur Radiol*. 2020;1–12:4406.
7. Long C, Xu H, Shen Q, et al. Diagnosis of the coronavirus disease (COVID-19): rRT-PCR or CT? *Eur J Radiol*. 2020;126: 108961.
8. Borakati A, Perera A, Johnson J, Sood T. Diagnostic accuracy of X-ray versus CT in COVID-19: a propensity-matched database study. *BMJ Open*. 2020;10(11):e042946.
9. Li Y, Xia L. Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. *AJR Am J Roentgenol*. 2020; 214(6):1280-1286.
10. Awulachew E, Diriba K, Anja A, Getu E, Belayneh F. Computed tomography (CT) imaging features of patients with COVID-19: systematic review and meta-analysis. *Radiol Res Pract*. 2020;2020:1023506.
11. Yurdaisik I. Effectiveness of computed tomography in the diagnosis of novel Coronavirus-2019. *Cureus*. 2020;12(5):e8134.
12. Kovács A, Palásti P, Veréb D, Bozsik B, Palkó A, Kincses ZT. The sensitivity and specificity of chest CT in the diagnosis of COVID-19. *Eur Radiol*. 2021;31(5):2819-2824.
13. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48(4):441-446.
14. Yip SS, Aerts HJ. Applications and limitations of radiomics. *Phys Med Biol*. 2016;61(13):R150-R166.
15. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med*. 2017;38:122-139.
16. Hassani C, Varghese BA, Nieva J, Duddalwar V. Radiomics in pulmonary lesion imaging. *AJR Am J Roentgenol*. 2019;212(3): 497-504.
17. Bouchareb Y, Khaniabadi PM, Al Kindi F, et al. Artificial intelligence-driven assessment of radiological images for COVID-19. *Comput Biol Med*. 2021;136:104665.
18. Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun*. 2020;11(1):4080.
19. Bai HX, Wang R, Xiong Z, et al. Artificial intelligence augmentation of radiologist performance in distinguishing COVID-19 from pneumonia of other origin at chest CT. *Radiology*. 2020; 296(3):E156-E165.
20. Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. *Cell*. 2020;181(6):1423-1433.e11.
21. Di D, Shi F, Yan F, et al. Hypergraph learning for identification of COVID-19 with CT imaging. *Med Image Anal*. 2020;68: 101910.
22. Xie C, Ng MY, Ding J, et al. Discrimination of pulmonary ground-glass opacity changes in COVID-19 and non-COVID-19 patients using CT radiomics analysis. *Eur J Radiol Open*. 2020; 7:100271.
23. Albahli S, Yar G. Fast and accurate COVID-19 detection along with 14 other chest pathology using: multi-level classification. *J Med Internet Res*. 2021;23:e23693.
24. Shi H, Xu Z, Cheng G, et al. CT-based radiomic nomogram for predicting the severity of patients with COVID-19. *Eur J Med Res*. 2022;27(1):13.
25. Cai S, Chen Y, Zhao S, et al. Dynamic 3D radiomics analysis using artificial intelligence to assess the stage of COVID-19 on CT images. *Eur Radiol*. 2022;1–11:4760-4770.
26. Ke Z, Li L, Wang L, et al. Radiomics analysis enables fatal outcome prediction for hospitalized patients with coronavirus disease 2019 (COVID-19). *Acta Radiol*. 2022;63(3):284185121 994695.
27. Wang H, Wang L, Lee EH, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. *Eur J Nucl Med Mol Imaging*. 2021;48(5): 1697.
28. Moradi Khaniabadi P, Bouchareb Y, Al-Dhuhli H, et al. Two-step machine learning to diagnose and predict involvement of lungs in COVID-19 and pneumonia using CT radiomics. *Comput Biol Med*. 2022;150:106165.
29. Shiri I, Mostafaei S, Haddadi Avval A, et al. High-dimensional multinomial multiclass severity scoring of COVID-19 pneumonia using CT radiomics features and machine learning algorithms. *Sci Rep*. 2022;12(1):14817.
30. Amyar A, Modzelewski R, Li H, Ruan S. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation. *Comput Biol Med*. 2020;126: 104037.
31. Peng S, Pan L, Guo Y, et al. Quantitative CT imaging features for COVID-19 evaluation: the ability to differentiate COVID-19 from non- COVID-19 (highly suspected) pneumonia patients during the epidemic period. *PLoS One*. 2022; 17(1):e0256194.
32. Velichko E, Sharialy F, Orooji M, et al. Development of computer-aided model to differentiate COVID-19 from pulmonary edema in lung CT scan: EDECOVID-net. *Comput Biol Med*. 2022;141:105172.
33. Das D, Santosh KC, Pal U. Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Phys Eng Sci Med*. 2020; 43(3):915-925.
34. Wang J, Bao Y, Wen Y, et al. Prior-attention residual learning for more discriminative COVID-19 screening in CT images. *IEEE Trans Med Imaging*. 2020;39(8):2572-2583.

35. Roberts M, Driggs D, Thorpe M, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nat Mach Intell*. 2021;3(3):199-217.
36. Tizhoosh HR, Fratesi J. COVID-19, AI enthusiasts, and toy datasets: radiology without radiologists. *Eur Radiol*. 2021;31(5):3553-3554.
37. Shiri I, Salimi Y, Saberi A, et al. Diagnosis of COVID-19 using CT image radiomics features: a comprehensive machine learning study involving 26,307 patients. *medRxiv*. 2021;2021:2021.2012. 2007.21267367.
38. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ*. 2015;350:g7594.
39. Mongan J, Moy L, Charles E, Kahn J. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. 2020;2(2):e200029.
40. Clark K, Vendt B, Smith K, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045-1057.
41. Morozov S, Andreychenko A, Pavlov N, et al. Mosmeddata: chest ct scans with covid-19 related findings dataset. arXiv Preprint arXiv:200506465. 2020.
42. Colak E, Kitamura FC, Hobbs SB, et al. The RSNA pulmonary embolism CT dataset. *Radiol Artif Intell*. 2021;3(2):e200254.
43. Ning W, Lei S, Yang J, et al. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. *Nat Biomed Eng*. 2020;4(12):1197-1207.
44. Shiri I, Salimi Y, Pakbin M, et al. COVID-19 prognostic modeling using CT radiomic features and machine learning algorithms: analysis of a multi-institutional dataset of 14,339 patients. *Comput Biol Med*. 2022;145:105467.
45. Radpour A, Bahrami-Motlagh H, Taaghi MT, et al. COVID-19 evaluation by low-dose high resolution CT scans protocol. *Acad Radiol*. 2020;27(6):901.
46. Shiri I, Arabi H, Salimi Y, et al. COLI-net: deep learning-assisted fully automated COVID-19 lung and infection pneumonia lesion detection and segmentation from chest computed tomography images. *Int J Imaging Syst Technol*. 2022;32(1):12-25.
47. Salimi Y, Shiri I, Akhavanallaf A, et al. Deep learning-based fully automated Z-axis coverage range definition from scout scans to eliminate overscanning in chest CT imaging. *Insights Imaging*. 2021;12(1):1-16.
48. Salimi Y, Shiri I, Akhavanallaf A, et al. Deep learning-based calculation of patient size and attenuation surrogates from localizer image: toward personalized chest CT protocol optimization. *Eur J Radiol*. 2022;157:110602.
49. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104-e107.
50. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. 2020;295(2):328-338.
51. Depeursinge A, Andrearczyk V, Whybra P, et al. Standardised convolutional filtering for radiomics. arXiv Preprint arXiv:200605470. 2020.
52. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12(1):1-8.
53. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825-2830.
54. Fang X, Li X, Bian Y, Ji X, Lu J. Radiomics nomogram for the prediction of 2019 novel coronavirus pneumonia caused by SARS-CoV-2. *Eur Radiol*. 2020;30(12):6888-6901.
55. Tan HB, Xiong F, Jiang YL, et al. The study of automatic machine learning base on radiomics of non-focus area in the first chest CT of different clinical types of COVID-19 pneumonia. *Sci Rep*. 2020;10(1):18926.
56. Zeng QQ, Zheng KI, Chen J, et al. Radiomics-based model for accurately distinguishing between severe acute respiratory syndrome associated coronavirus 2 (SARS-CoV-2) and influenza a infected pneumonia. *MedComm (Beijing)*. 2020;1:240-248.
57. Bae J, Kapse S, Singh G, et al. Predicting mechanical ventilation requirement and mortality in COVID-19 using radiomics and deep learning on chest radiographs: a multi-institutional study. *Diagnostics*. 2021;11(10):1812.
58. Chandra TB, Verma K, Singh BK, Jain D, Netam SS. Coronavirus disease (COVID-19) detection in chest X-ray images using majority voting based classifier ensemble. *Expert Syst Appl*. 2021;165:113909.
59. Chen H, Guo S, Hao Y, et al. Auxiliary diagnosis for COVID-19 with deep transfer learning. *J Digit Imaging*. 2021;34:1-11.
60. Chao H, Fang X, Zhang J, et al. Integrative analysis for COVID-19 patient outcome prediction. *Med Image Anal*. 2020;67:101844.
61. Chassagnon G, Vakalopoulou M, Battistella E, et al. AI-driven quantification, staging and outcome prediction of COVID-19 pneumonia. *Med Image Anal*. 2020;67:101860.
62. Lassau N, Ammari S, Chouzenoux E, et al. Integrating deep learning CT-scan model, biological and clinical variables to predict severity of COVID-19 patients. *Nat Commun*. 2021;12(1):1-11.
63. Shiri I, Amini M, Nazari M, et al. Impact of feature harmonization on radiogenomics analysis: prediction of EGFR and KRAS mutations from non-small cell lung cancer PET/CT images. *Comput Biol Med*. 2022;105230:105230.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Shiri I, Salimi Y, Saberi A, et al. Differentiation of COVID-19 pneumonia from other lung diseases using CT radiomic features and machine learning: A large multicentric cohort study. *Int J Imaging Syst Technol*. 2024;34(2):e23028. doi:10.1002/ima.23028