Turkish Journal of Science & Technology
18(1), 183-198, 2023

Research Paper
https://doi.org/10.55525/tjst.1237103

A Hybrid Model Based on Deep Features and Ensemble Learning for the Diagnosis of COVID-19: DeepFeat-E

Berivan ÖZAYDIN[1], Ramazan TEKİN[*2]
[1]Electrical and Electronics Engineering, Graduate Education Institute, Batman University
[2]Computer Engineering, Faculty of Engineering and Architecture, Batman University
[1] berzaydin@gmail.com, [*2] ramazan.tekin@batman.edu.tr

(Geliş/Received: 17/01/2023; Kabul/Accepted: 15/03/2023)

Abstract: COVID-19, which has been declared a pandemic disease, has affected the lives of millions of people and caused a major epidemic. Despite the development of vaccines and vaccination to prevent the transmission of the disease, COVID-19 case rates fluctuate worldwide. Therefore, rapid and reliable diagnosis of COVID-19 disease is of critical importance. For this purpose, a hybrid model based on transfer learning methods and ensemble classifiers is proposed in this study. In this hybrid approach, called DeepFeat-E, the diagnosis process is performed by using deep features obtained from transfer learning models and ensemble classifiers consisting of classical machine learning methods. To test the proposed approach, a dataset of 21,165 X-ray images including 10,192 Normal, 6012 Lung Opacity, 1345 Viral Pneumonia and 3616 COVID-19 were used. With the proposed approach, the highest accuracy was achieved with the deep features of the DenseNet201 transfer learning model and the Stacking ensemble learning method. Accordingly, the test accuracy was 90.17%, 94.99% and 94.93% for four, three and two class applications, respectively. According to the results obtained in this study, it is seen that the proposed hybrid system can be used quickly and reliably in the diagnosis of COVID-19 and lower respiratory tract infections.

Key words: COVID-19, Deep Features, Transfer Learning, Ensemble Classifier.

COVID-19 Teşhisi için Derin Özniteliklere ve Topluluk Öğrenmeye Dayalı Hibrit bir Model: DeepFeat-E

Öz: Pandemik hastalık olarak ilan edilen COVID-19, milyonlarca insanın hayatını etkilemiş ve büyük bir salgına neden olmuştur. Hastalığın bulaşmasını önlemek amacıyla aşılar geliştirilmesine ve aşılama yapılmasına rağmen dünya genelinde COVID-19 vaka oranları dalgalı bir seyir göstermektedir. Dolayısıyla COVID-19 hastalığının hızlı ve güvenilir teşhisi kritik bir öneme sahiptir. Bu amaçla, bu çalışmada transfer öğrenme yöntemlerine ve topluluk sınıflandırıcılara dayalı hibrit bir model önerilmiştir. DeepFeat-E olarak isimlendirilen bu hibrit yaklaşımda, transfer öğrenme modellerinden elde edilen derin öznitelikler ile klasik makine öğrenme yöntemlerinden oluşan topluluk sınıflandırıcılar kullanılarak teşhis işlemi gerçekleştirilmektedir. Önerilen yaklaşımı test etmek için 10.192 Normal, 6012 Akciğer Opaklığı, 1345 Viral Pnömoni ve 3616 COVID-19 toplamda 21.165 X-ray görüntüsünden oluşan veri seti kullanılmıştır. Önerilen yaklaşım ile en yüksek başarı DenseNet201 transfer öğrenme modeline ait derin öznitelikler ve İstifleme topluluk öğrenme yöntemiyle elde edildiği görülmüştür. Buna göre dört, üç ve iki sınıflı uygulamalarda sırasıyla test doğruluğu 90,17%, 94,99% ve 94,93% olarak elde edilmiştir. Bu çalışma kapsamında elde edilen sonuçlara göre, önerilen hibrit sistemin COVID-19 ve alt solunum yolu enfeksiyonlarının teşhisinde hızlı ve güvenilir bir şekilde kullanılabileceğini görülmektedir.

Anahtar Kelimeler: COVID-19, Derin Öznitelikler, Transfer Öğrenme, Topluluk Sınıflandırıcı.

1. Introduction

COVID-19, the new virus of the coronavirus family called severe acute, affects millions of people worldwide. The World Health Organization (WHO) declared the outbreak caused by this virus, which causes severe acute respiratory syndrome, a global pandemic in March 2020. The first case of COVID-19 was registered in Wuhan, China [1,2]. This severe acute respiratory syndrome coronavirus (Sars-Cov-2) has caused approximately 6.2 million deaths and over 500 million cases of infection worldwide [3]. This puts a serious burden on healthcare workers. Despite advances in health and technology, the impact of the virus still persists and new cases and deaths continue.

The first symptoms of COVID-19 are often similar to the common cold or flu. This makes it difficult to detect the first stage of COVID-19 cases [4]. Due to the long incubation period, people with the virus continue their routine daily lives and interact with other people until they realize that they have the virus. As a result, this leads to more infections, making COVID-19 more contagious [5]. Most people have mild to moderate symptoms and do not require hospitalization. However, if acute respiratory distress progresses too far, it can cause cytokine

[*] Corresponding Author: ramazan.tekin@batman.edu.tr. ORCID Number of authors: 1 0000-0001-8935-6684, 2 0000-0003-4325-6922

respiratory syndrome and consequences such as multiple organ failure and death can occur [6]. This virus can be very dangerous, especially in people with chronic diseases such as diabetes, cardiovascular disease and asthma. In addition, due to the large size of the number of cases, it also creates serious problems in the country's economies [6].

The presymptomatic infection rate of Covid-19, the duration of viral transmission and the variability of characteristics in different countries make the COVID-19 pandemic more unpredictable and difficult to control [3]. Being able to contain Covid-19 depends on carriers being identified and quarantined, as well as being tested frequently. Rapid and accurate detection of the COVID-19 disease is critical to slowing the spread of the virus and saving the lives of vulnerable groups. Specialized vaccines have been developed to reduce the rates of COVID-19 infection. However, waves or fluctuations of infection are still experienced due to the difficulty of accessing vaccines in many poor countries and the lack of adequate vaccination.

COVID-19 can be detected by polymerase chain reaction (PCR) testing by taking throat and nose samples from patients [6]. However, since these test kits are not reliable enough, alternative rapid methods are needed to detect COVID-19 and other lower respiratory diseases. Today, artificial intelligence and especially deep learning methods are reliably used to diagnose COVID-19 and other lower respiratory diseases [7]. In addition, artificial intelligence systems can be successfully applied in a wide range of fields from lip-reading applications [8] to document language [9] and gesture recognition [10], epileptic seizures [11] and heart disease detection [12]. Similarly, artificial intelligence systems based on cough sounds [13,14] and especially chest images (X-Ray and CT scan) are widely used in COVID-19 diagnosis [15,16].

For this purpose, in this study, a hybrid artificial intelligence system named DeepFeat-E is proposed for COVID-19 diagnosis. The proposed system is based on deep features extracted from a given pre-trained Transfer Learning (TL) model. Using these features of the training data set, the best five Machine Learning (ML) methods are selected and used as classifiers in ensemble learning algorithms. In order to reduce the computational cost, the deep features were reduced by the Extra Tree Classifier (EAS) method before using them. Then, the reduced features were applied to the five best ML methods selected according to the 10-fold cross validation result. The final decisions of these classifiers are obtained using Stacking, Soft and Hard voting ensemble learning methods. The performance of the proposed hybrid model is analyzed using a large open access dataset of 21,165 chest X-Ray images, including 10,192 Normal, 6012 Lung Opacity, 1345 Viral Pneumonia and 3616 COVID-19 images. In the first application, TL models were applied directly to the dataset for comparison purposes. In the other three applications, separate applications were carried out with three different ensemble learning methods within the scope of the proposed hybrid system. In each application, two-class (COVID-19 and Normal), three-class (COVID-19, Viral-Pneumonia and Normal) and four-class (COVID-19, Lung-Opacity, Viral-Pneumonia and Normal) applications were performed separately.

2. Related Work

In this section, we present the current deep learning architecture studies on COVID-19 diagnosis, especially the studies based on the ensemble classification method. Researchers generally preferred transfer methods or improved (Fine-Tuning) transfer methods with their own small size datasets [17]. Using pre-trained models of these approaches reduces the need for labeled datasets for their training and the physical resources required during training [18]. There are quite a number of studies on COVID-19 diagnosis in the literature. Below, studies that include deep learning-based methods, especially ensemble learning methods, which is one of the focal points of this study, are presented.

Chowdhury et al. [19] proposed an EfficientNet-based an ensemble of deep convolutional neural networks (CNN) named ECOVNet to detect COVID-19 using a dataset of 16493 X-Ray images. This method has a CNN architecture embedded in a pre-trained EfficientNet. Using the proposed model, snapshots of several training predictions are entered into the ensemble classifier. In the study, X-Ray images containing COVID-19, normal and pneumonia samples were used and the highest three-class accuracy was %97.

In another study, Mahmud et al. [20] proposed a CNN-based architecture called CovXNet. In this study, the predictions of different forms of this architecture called CovXNet were evaluated together with the stacking ensemble classifier method. In the applications performed in the study, %97.4, %89.6 and %90.2 accuracy was obtained for two-class (COVID/Normal), three-class (COVID, Viral and bacterial pneumonia) and multi-class (COVID, normal, Viral and bacterial pneumonia), respectively.

In another study, Karim et al. [21] trained DenseNet, ResNet and VGGNet architectures and obtained snapshots of these models during the training. They proposed a framework called DeepCOVIDExplainer, which uses Softmax class posterior averaging (SCPA) and prediction maximization (PM) to integrate these models into ensemble classifiers. Tang et al. [22] proposed a hybrid model based on deep learning and ensemble learning called EDL-COVID. The EDL-COVID model uses multiple snapshots of the architecture named COVID-Net, which was developed for COVID-19 detection. It evaluates these snapshots in an ensemble classifier based on the weighted average method. In the study, X-Ray images containing examples of COVID-19, normal and pneumonia were used and the test accuracy was %95 for three classes.

Banerjee et al. [23] developed a model with ensemble structure by taking snapshots of the network from local minimum points in a single training process of a model with DenseNet-201 architecture. They used Blending method as an ensemble classifier and Random Forest (RF) method as a meta-model. Two different datasets with three categories (COVID-19, Pneumonia, and Normal) were used in the study, one consisting of a large number of X-Ray images and the other a low number of X-Ray images. The accuracy values of the proposed model for the three classes in large and small datasets were %94.55 and %98.13, respectively.

Gour and Jain [24] proposed a new stacked convolutional neural network model based on Xception and Vgg19. In the proposed approach, they tried to predict diagnosis with softmax classifier and Stacking ensemble learning methods, using sub-models obtained from VGG19 and Xception models during training. In the study, CT (Computerized Tomography) images with two categories (COVID-19 and No-Findings) and X-Ray images with three categories (COVID-19, Pneumonia and Normal) were used and %98.30 and %97.27 accuracy was achieved respectively.

In the diagnosis of COVID-19, it is seen that models based on transfer learning or original convolutional architectures are used extensively in the literature, as well as ensemble classifier-based studies. Apostolopoulos and Mpesiana [25] examined some deep architectures such as VGG, Inception, MobilNet on X-ray images. In the study, a dataset consisting of X-Ray images with the categories of COVID-19, pneumonia and normal was used. Among the Transfer Learning models, the highest accuracy was obtained in the VGG19 model with %93.48 and %98.75 for 3 classes and 2 classes, respectively.

Ozturk et al. [26] proposed a new model called CovidDarkNet for automatic COVID-19 detection using raw X-ray images for Covid-19 diagnosis. The CovidDarkNet model, which is used as a classifier for the YOLO (You Only Look Once) real-time object detection system, includes 17 convolutional layers. The method proposed in this study was applied to datasets with two classes (COVID and No-Findings) and three classes (COVID, No-Findings, Pneumonia). It was reported that %98.08 and %87.02 classification accuracy was obtained for 2 classes and 3 classes , respectively.

Khan et al. [27] proposed a deep convolutional neural network model called CoroNet based on Xception architecture for the diagnosis of COVID-19 from X-ray images. The proposed model is based on the Xception architecture pre-trained on the ImageNet dataset. With the proposed model, %89.5, %94.59 and %99 accuracy was achieved on 4, 3 and 2 class datasets, respectively.

Huang and Liao [28] performed Covid-19 diagnosis using InceptionV3, ResNet50V2, Xception, DenseNet121, MobileNetV2, EfficientNet and EfficientNetV2 transfer methods. In this study, they also proposed a new architecture called LightEfficientNetV2. The applications were performed on 2 different datasets consisting of X-Rays and CT images with three classes. The three-class highest accuracy in X-Ray and CT images was obtained in the LightEfficientNetV2 model with %98.33 and %97.48, respectively [28]. Similarly, Ahamed et al. [15] developed a modified ResNet50V2 architecture for COVID-19 detection with datasets consisting of X-Ray and CT images. The datasets used in the study include four classes: COVID-19, Normal, viral pneumonia and bacterial pneumonia. Using the proposed model, an accuracy of %96.45, %97.24 and %98.95 was obtained for X-Ray images with four (COVID-19/Normal/Bacterial pneumonia/Viral pneumonia), three (COVID-19/Normal/Bacterial pneumonia) and two (COVID-19/Viral pneumonia) classes, respectively. In CT images, %99.01 and %99.99 accuracy was obtained in images with three (COVID-19/Normal/Pneumonia) and two (Normal/COVID-19) classes, respectively.

Islam et al. [29] proposed a new convolutional neural network model called Cov-RADNet for the diagnosis of COVID-19 from CT and X-ray images. The first dataset used in this study consists of X-ray images with four categories: COVID-19, viral pneumonia, lung-opacity, and normal. The other dataset, CT, contains images with COVID-19 and non-COVID categories. The prediction accuracy of their proposed model was %97, %99.5 and %99.72 for four classes (COVID-19, Viral Pneumonia, Lung-Plaque and Normal), three classes (COVID-19, Viral Pneumonia, Normal) and two classes (COVID-19 and non-COVID), respectively. In CT images, the prediction accuracy was %99.25.

3. The Proposed Method

In the hybrid system called DeepFeat-E proposed in this study, deep features obtained from pre-trained Transfer Learning models are entered into classical ML methods and COVID-19 disease diagnosis is performed with the help of ensemble classifiers. The proposed system generally consists of four basic stages as shown in Figure 1. The dataset used in the study is a very large dataset containing a total of 21,135 chest X-Ray images in four categories (Normal, Lung Opacity, Viral Pneumonia and COVID-19), and this dataset was selected as the best COVID-19 dataset by the Kaggle committee [19].

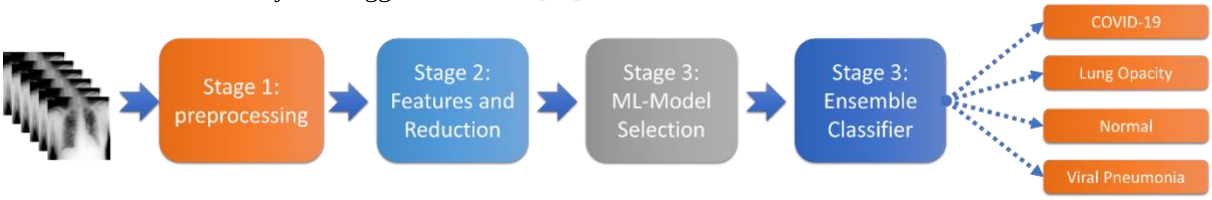

Figure 1. Stages of the proposed hybrid system.

Stage 1 (Preprocess): In the first stage, the preprocessing stage, after all images are brought to a standard size (224x224x3); Data augmentation is performed by methods such as rotation, scaling, horizontal/vertical translation, horizontal flip, brightness adjustment. Then, %80 of the data set is divided into two as training and %20 as testing, and the next stage is passed.
Stage 2 (Features and Reduction): In the second stage, the deep features of the training and test images obtained separately from the valid pre-trained TL method are reduced and transferred to the next stage. This step is performed separately for each TL model.
Stage 3 (ML-Model Selection): In the third stage, the best five ML methods are selected by using the reduced deep features of the training dataset. The selection process is performed according to the AUC metric after 10 cross validations. Then, the five best ML models selected are transferred to the next stage for use in the ensemble classifier.
Stage 4 (Ensemble Classifier): In the fourth stage, which is the final stage, predictions are made with the current ensemble learning method using the best five ML models. For each of the Stacking, Soft Voting and Hard Voting ensemble learning methods used in the study, this stage was carried out separately with four-class, three-class and two-class data sets.

In this study, all applications were implemented using Python programming language and Tensorflow, Keras and scikit-learn packages. A personal computer with AMD Ryzen 7 5800H (16 CPU, ~3.2GHz) processor, Nvidia GeForce RTX 3050 GPU (4GB GDDR6, ~1.5GHz) and 16GB RAM was used for models and analysis.

3.1. Datasets

The dataset used in this study consists of X-Ray images collected by a group of researchers from different countries (Qatar, Bangladesh, Pakistan, etc.) with four categories: COVID, Viral Pneumonia, Non-COVID Lung Opacity and Normal [19,30]. The dataset "COVID-19 Radiography Database" used in the study can be downloaded from Kaggle's website.

Table 1. Number of images in training and test dataset.

Datasets	COVID-19	Lung-Opacity	Normal	Viral-Pneumonia
Total	3616	6012	10192	1345
Train	2893	4809	8154	1076
Test	723	1203	2038	269

The dataset is updated periodically by the working group and the latest version used in this study contains 10,192 Normal, 6012 Lung Opacity (Non-COVID lung infection), 1345 Viral Pneumonia and 3616 chest X-Ray images of COVID-19 [30]. Table 1 shows the distribution of training and test datasets. Contrary to similar studies in the literature, the number of images for each category in the dataset is not equalized. However, data augmentation techniques such as rotation, scaling, horizontal/vertical shift, horizontal rotation, etc. were applied to prevent overlearning and to ensure that the models produce stable results. Examples of the images in the dataset are given in Figure 2.

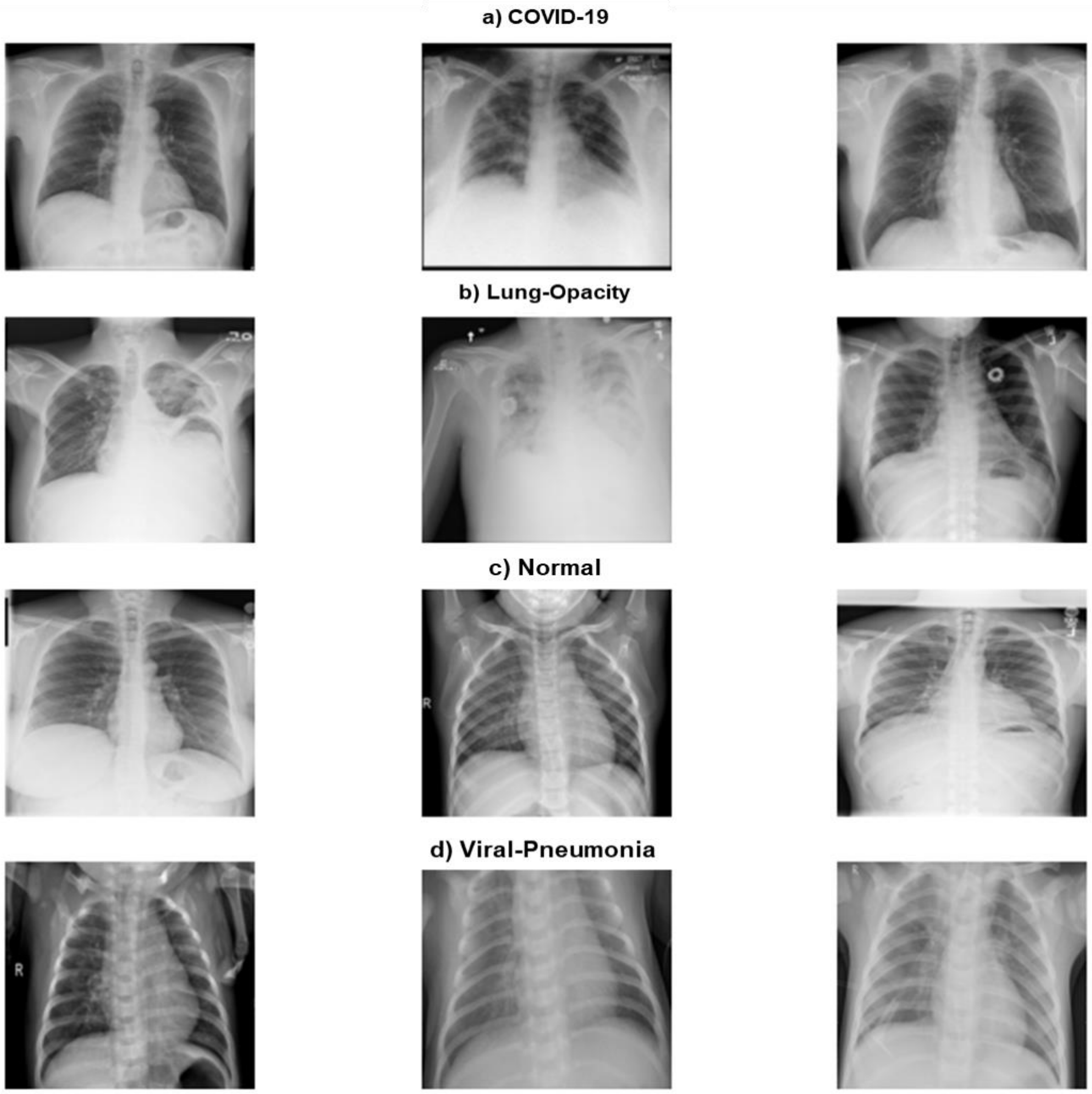

Figure 2. Example images. a) COVID-19, b) Lung-Opacity, c) Normal/Healthy, d) Viral-Pneumonia.

3.2. Proposed Hybrid System

Within the scope of the study, applications were carried out for a total of four different systems in which pre-trained TL architectures were used directly without community learning and three different ensemble learning techniques were used. In the ensemble learning systems, the deep features obtained from pre-trained TL architectures were analyzed by Stacking, Soft and Hard Voting ensemble methods by entering the five best ML models selected. For each TL model, datasets with two (COVID-19 and Normal), three (COVID-19, Viral Pneumonia and Normal) and four (COVID-19, Viral Pneumonia, Normal, and Lung-Opacity) classes were created and analyzed separately. The pre-trained TL models used in the study are listed in Table 2 and the applications were performed separately for each of them. Since these models are known general TL models, detailed explanations are not given under a separate heading. Detailed information about these models can be accessed through the references given next to the relevant model in the table. In the proposed system, the five best ML (S.ML) methods selected for the ensemble classifier are different for each TL, and the selected methods of the four-class applications are listed in Table 2 by ordering them according to the UAC value. In this table, the number of parameters (Prm), the number of deep features before reduction (BR.Fet.) and after reduction (AR.Fet.) are presented for each TL.

Table 2. Used TL methods and deep features.

#	TL Model	S.ML Methods	Prm	BR.Fet.	AR.Fet.
1	Xception [31]	[lr, lda, xgb, lgbm, gb]	20,869,676	2048	550
2	NASNet [32]	[lr, lda, xgb, lgbm, gb]	84,932,950	4032	1163
3	MobileNet [33]	[xgb, lgbm, lr, lda, gb]	3,232,964	1024	263
4	DenseNet169 [34]	[xgb, lda, lgbm, gb, lr]	12,649,540	1664	398
5	DenseNet201 [34]	[xgb, lgbm, lda, gb, rf]	18,329,668	1920	461
6	VGG16 [35]	[xgb, lgbm, lr, lda, gb]	14,716,740	512	189
7	InceptionV3 [36]	[xgb, lr, lgbm, lda, gb]	21,810,980	2048	520
8	ResNet50V2 [37]	[xgb, lgbm, lr, lda, gb]	23,572,996	2048	508
9	ResNet101V2 [37]	[xgb, lgbm, lr, lda, gb]	42,634,756	2048	521

In the first model without ensemble learning, the training and test datasets were applied directly to the TL models in order to compare the performance of the proposed ensemble classifier model. The TL models were used without retraining (pre-trained) with their current configurations and weights. The block diagram of the hybrid system named DeepFeat-E based on the ensemble learning methods proposed in this study is presented in Figure 3. As shown in the *Feature Selection* block, the deep features obtained from each of the pre-trained TL models were reduced by trying various feature selection methods (principal component analysis, linear discriminant analysis and extra tree classifier). Since the most successful results were obtained with the Extra Tree Classifier, after the reduction with this feature selection method, diagnosis was tried to be estimated with Stacking, Soft Voting and Hard Voting ensemble learning techniques, respectively. In ensemble learning models, as shown in the *Model Selection* block, the top 5 methods with the best performance (according to UAC) according to 10 cross-validation results are selected among 14 different known ML methods. Although the priority order of the selected ML methods varies for each TL model, it is generally seen that lr (Logistic Regression), lda (Linear Discriminant Analysis), xgb (eXtreme Gradient Boosting), lgbm (Light Gradient Boosting Machine), gb (Gradient Boosting), rf (Random Forest) ML models were selected. Then, these five best ML methods selected were used as classifiers as seen in the *Ensemble Learning* block. The top five ML models for each TL are presented in Table 2.

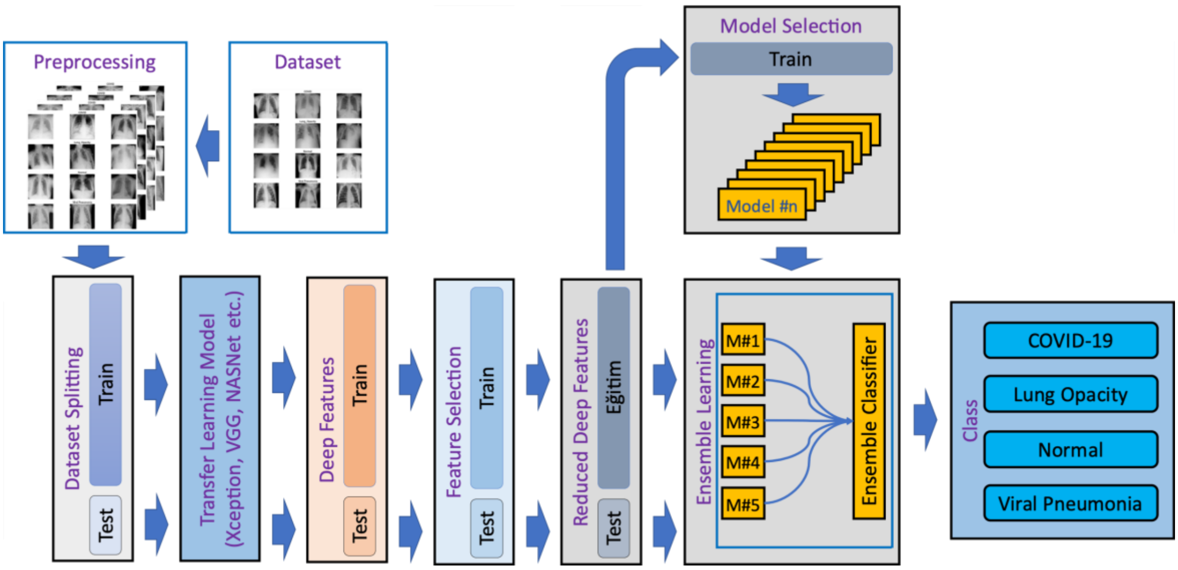

Figure 3. Block diagram structure of the proposed system.

3.3. Transfer Learning (TL) Methods

TL models are used to solve complex problems in various fields. In terms of machine learning, TL is the use of architectures and trained models designed for a specific problem by pre-training or retraining them on a different problem [38]. Since there is no need for retraining in new problems for which it is desired to use TL models, it saves both resources and time [39]. In this study, 9 TL models listed in Table 2 were used.

3.4. Feature Selection with Extra Tree Classifier (ETC)

ETC (Extra Trees Classifier) feature selection method was used to extract appropriate features from the deep features obtained from transfer deep learning methods [40]. Feature reduction is a data preprocessing process that enables the selection of important features that will make the maximum contribution to the estimation method, especially in high-dimensional data. Elimination of unimportant features makes the problem simpler, reducing the computational cost of models and increasing model accuracy [41].

ETC is a decision-based method that provides a common framework between feature selection and classification. It is very similar to the Random Forest classifier due to its characteristics such as generating many subtrees and selecting random subsets. A random partitioning is performed at the parent node and then at the child nodes up to the leaf. The predictions of all trees are then combined to determine the majority decision. During the creation of the forest for feature selection, the Gini importance value is calculated for each attribute. For feature selection, the features are sorted in descending order according to their Gini importance, and as many features as desired can be selected among the top attributes.

$$Gini = 1 - \sum_{i=0}^{c-1} p_i(t)^2 \tag{1}$$

where $p_i(t)$ is the frequency of class i at node t, and c is the number of unique classes at this node.

3.5. Ensemble Learning Methods

Ensemble learning techniques aim to increase success by using multiple decision makers (classifiers) instead of a single decision maker in the decision-making phase. For classification, after 10 cross-validations according to the training set from a group of classical ML models, the best five were selected according to the AUC metric and used in the ensemble classifiers. The Deep features obtained from TL methods were entered into these classifiers and classified using Stacking and Voting (Soft and Hard) ensemble learning techniques. In the voting method, Hard Voting and Soft Voting techniques, which are based on majority vote and the average of class prediction probabilities, respectively, are used to predict class labels.

3.5.1. Voting Ensemble Learning Methods

The mainstay of voting ensemble learning methods is to combine the predictions of different ML methods to obtain a common ensemble decision. In this way, it is predicted that better performance can be obtained by balancing the wrong predictions caused by a single model. Suppose that for a sample x, we want to predict the class among k classes $\{s_1, s_2, s_3, \ldots, s_k\}$, with n separate classifiers $\{h_1, h_2, h_3, \ldots, h_n\}$ [42].
In the Hard Voting method, the prediction result of the majority of the classifiers that make up the ensemble for a given sample is considered to be the class of the sample. Accordingly, the predicted class of sample x according to the hard voting method is expressed as follows [42],

$$H(x) = s_{\underset{j}{\operatorname{argmax}} \sum_{i=1}^{n} h_i^j(x)} \tag{2}$$

In the soft voting method, the class prediction is made by averaging the weighted probabilistic predictions. The class with the highest average is considered as the prediction of the ensemble. Accordingly, the predicted class of sample x according to the soft voting method is expressed as follows [42],

$$H(x) = s_{\underset{j}{\operatorname{argmax}} \sum_{i=1}^{n} w_i h_i^j(x)} \tag{3}$$

where w_i represents the weight of the h_i classifier in the ensemble.

3.5.2. Stacking Ensemble Learning

The stacking ensemble learning method was developed by Wolpert [43] and is a two-step method. In the first stage, predictions are generated by different classifiers using the same data set (training) and in the second stage, these predictions are processed by a meta-classifier to obtain the ensemble prediction. The aim here is for the ensemble classifier to obtain predictions with higher accuracy [43].

3.6. Performance Evaluation

In this study, 10-fold cross-validation was used for performance evaluation. Accuracy (Acc.), Precision (Pre.), Sensitivity (Sen.), f1-score (f1-sc.) and AUC (area under the ROC) metrics were used to evaluate the performance of both TL models and ensemble learning classifiers [44]. These metrics are obtained in the light of the data in the confusion matrix in Table 3.

Table 3. Representation of the Confusion Matrix

	Predicted Label	
True Label	True Positive (TP)	False Negative (FN)
	False Positive (FP)	True Negative (TN)

Accuracy is calculated as the ratio of correctly predicted samples to the total number of samples in the model.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} * 100 \tag{4}$$

Sensitivity is the generic name given to the metric that indicates how many of the samples we should have predicted as positive we predicted as positive.

$$Sensitivity = \frac{TP}{TP + FN} * 100 \tag{5}$$

Precision is the metric that allows us to check how many of the values we predict as positive are actually positive.

$$Precision = \frac{TP}{TP + FP} * 100 \tag{6}$$

The f1-score is the harmonic mean of sensitivity and precision measures. The f1-score reveals realistic results including all error costs in the case of non-uniformly distributed data sets, or data sets with unknown distribution.

$$f1 - score = 2 * \frac{TP}{2TP + FN + FP} * 100 \tag{7}$$

4. Experimental Results

In order to compare the performance of the proposed system, firstly, direct classification was performed with the TL model without using an ensemble classifier. The other three applications were performed with Stacking, Soft and Hard Voting ensemble classifiers and making a total of four separate applications. The nine TL models used within the scope of the study were tested separately in each application. In addition, in each of these four applications, two-class (COVID-19 and Normal), three-class (COVID-19, Viral-Pneumonia and Normal) and four-class (COVID-19, Lung-Opacity, Viral-Pneumonia and Normal) applications were performed. Performance measures for the applications are presented separately in tables.

For this purpose, the performance analysis of the four-, three- and two-class applications for the training and test datasets are presented separately. Table 4 shows the accuracy of the TL models without ensemble classifiers for the training and test datasets. Table 5 presents the accuracies obtained for the test dataset using deep features and the proposed ensemble classifier system. The highest accuracies for training and test datasets in four-class applications when the ensemble classifier is not used are %84.83 (Precision: %85.67, Sensitivity: %83.49, f1-score: %84.49) and %85.09 (Precision: %87.18, Sensitivity: %82.05, f1- score: %84.14) obtained with MobileNet. The lowest estimation accuracy was obtained with the VGG16 TL model as %63.17 and %63.17 for the training and test data sets, respectively. In the other application group, the tri-class applications, the highest accuracies were obtained with the DenseNet201 model as %91.26 (Precision: %92.29, Sensitivity: %87.60, f1-score: %89.69) and with the MobileNet as %90.53 (Precision: %93.10, Sensitivity: %84.39, f1-score: %88.09). Similarly, the lowest prediction accuracy was obtained as 73.07% and 72.42% for the training and test data sets in the VGG16 BL model, respectively. In the last application group, two-class applications, the highest accuracies for both training and test datasets were obtained with MobileNet as %90.98 (Precision: %90.10, Sensitivity: %85.94, f1-score: %87.73) and %91.53 (Precision: %92.35, Sensitivity: %85.38, f1-score: %88.12), respectively. The lowest prediction accuracy was obtained in the VGG16 TL model, as %78.52 and %78.93 for the training and test datasets, as in the previous four and three-class applications.

Table 4. Training and test accuracies of TL models.

	Train (%)			**Test (%)**		
Model	**4-class**	**3-class**	**2-class**	**4-class**	**3-class**	**2-class**
Xception	80.86	87.72	87.43	80.6	88.06	88.09
NASNet	80.74	88.04	87.9	79.09	87.73	87.91
MobileNet	**84.83**	90.99	**90.98**	**85.09**	**90.53**	**91.53**
DenseNet169	84.33	90.75	89.9	84.88	90.2	90.55
DenseNet201	83.75	**91.26**	90.74	82.45	90.47	91.09
VGG16	63.17	73.07	78.52	63.17	72.42	78.93
InceptionV3	81.22	88.12	88.19	80.75	86.34	86.82
ResNet50V2	84.37	90.4	90.52	84.17	89.64	90.7
ResNet101V2	83.41	90.29	89.29	83.7	88.95	89.86

According to the proposed hybrid model named DeepFeat-E, in four-class applications with deep features and ensemble classifiers, the highest accuracy of %90.17 (Precision: %92.29, Sensitivity: %89.65, f1-score: %90.88) for the Stacking method was obtained with the DenseNet201 TL model. For Soft and Hard voting methods, the highest accuracy values were obtained with DenseNet169 Transfer Learning model as %88.45 (Precision: %90.84, Sensitivity: %87.05, f1-score: %88.72) and %87.93 (Precision: %90.15, Sensitivity: %86.35, f1-score: %88.03), respectively. On the other hand, the lowest accuracy values were obtained with the Xception TL model for Stacking, Soft and Hard Voting as %82.35, %81.53 and %80.75, respectively. In the other group of applications with three classes, the highest accuracy for the Stacking model was %94.99 (Precision: %95.99, Sensitivity: %91.65, f1-score: %93.68) with the DenseNet201 TL model, and for the Soft and Hard voting methods as %93.17 (Precision: %95.00, Sensitivity: %88.70, f1-score: %91.55) and %92.91 (Precision: %94.59, Sensitivity: %88.34 , f1-score: %91.17) with the DenseNet169 TL model, respectively. On the other hand, the lowest accuracy values were obtained in the Xception TL model as %89.84 for the Stacking method, and for Soft and Hard Voting as %88.52 and %88.29, respectively, in the InceptionV3 TL model. In the last group of applications, two-class applications, the highest accuracy in the Stacking method and Soft and Hard voting methods is %94.82 (Precision: %94.43, Sensitivity: %91.99, f1-score: %93.12), %94.21 (Precision: %94.93, Sensitivity: %89.96, f1- score: %92.09) and %94.17 (Precision: %95.13, Sensitivity: %89.71, f1- score: %92.00) was obtained with the DenseNet201 TL model. On the other hand, the lowest accuracy values were obtained with the InceptionV3 TL model as %90.15, %89.68 and %89.46 for Stacking, Soft and Hard Voting methods, respectively.

Table 5. Test accuracy values of ensemble classifiers.

	Stacking Method (%)			Soft Voting (%)			Hard Voting (%)		
Model	**4-class**	**3-class**	**2-class**	**4-class**	**3-class**	**2-class**	**4-class**	**3-class**	**2-class**
Xception	82.35	89.84	90.88	81.53	89.15	90.51	80.75	88.49	90.33
NASNet	84.08	90.7	91.42	81.74	89.05	91.31	81.72	88.55	90.59
MobileNet	88.09	93.67	93.08	86.16	91.95	92.72	85.64	91.45	92.83
DenseNet169	89.63	94.52	93.05	**88.45**	**93.17**	93.56	**87.93**	**92.91**	93.52
DenseNet201	**90.17**	**94.99**	**94.82**	87.81	92.97	**94.21**	87.24	92.35	**94.17**
VGG16	87.15	94.09	92.61	84.08	90.83	92.14	84.03	90.37	91.82
InceptionV3	83.13	90.1	90.15	82.07	88.52	89.68	81.29	88.29	89.46
ResNet50	88.33	93.27	93.63	86.51	91.42	92.47	85.99	91.09	92.58
ResNet101	86.96	92.51	93.12	84.74	90.37	92.65	84.57	90.1	92.43

Table 6 shows the improvement amounts by listing the differences between the accuracy values obtained directly with TL models without using ensemble classifiers (Test column in Table 4) and the accuracy values obtained with the proposed hybrid system (Table 5). Figure 4(a, b and c) shows the improvement of the proposed system for each of the Stacking, Soft and Hard voting ensemble classifiers respectively. The amount of improvement in the TL models used in each graph is presented by grouping according to the different number of classes performed. Accordingly, in the proposed system, ensemble classifiers achieved higher prediction accuracy than TL models in all applications and significantly increased the accuracy of TL models.

Table 6. Amount of improvement in test accuracies of the proposed hybrid system.

	Stacking Method (%)			Soft Voting (%)			Hard Voting (%)		
Model	**4-class**	**3-class**	**2-class**	**4-class**	**3-class**	**2-class**	**4-class**	**3-class**	**2 sınıf**
Xception	1.75	1.78	2.79	0.92	1.09	2.43	0.14	0.43	2.24
NASNet	4.98	2.97	3.51	2.65	1.32	3.4	2.62	0.82	2.68
MobileNet	3	3.13	1.56	1.06	1.42	1.19	0.54	0.92	1.3
DenseNet169	4.75	4.32	2.5	**3.57**	**2.97**	3.01	**3.05**	**2.71**	2.97
DenseNet201	**7.73**	**4.52**	**3.73**	5.36	2.51	**3.11**	4.8	1.88	**3.08**
VGG16	23.98	21.68	13.69	20.91	18.41	13.22	20.86	17.95	12.89
InceptionV3	2.39	3.76	3.33	1.32	2.18	2.86	0.54	1.95	2.64
ResNet50	4.16	3.63	2.93	2.34	1.78	1.77	1.82	1.45	1.88
ResNet101	3.26	3.56	3.26	1.04	1.42	2.79	0.87	1.15	2.57

As stated before, the highest estimation accuracy in four, three and two-class applications was obtained with the hybrid model named DeepFeat-E. Accordingly, the accuracy values are 90.17%, 94.99% and 94.82% when the stacking ensemble classifier is used with the deep features obtained from the DenseNet201 TL model. As seen in Table 6 and Figure 4(a), the proposed system for DenseNet201 improves the accuracies of the TL model by +7.73, +4.52 and 3.73 points, respectively. Moreover, when all TL models are considered, it is seen that the highest accuracy improvements in four, three and two class applications are realized in the VGG16 TL model by +23.98,

+21.68 and +13.69 points, respectively, with the stacking ensemble learning method. In addition, the VGG16 TL model is also significantly improved in soft and hard voting techniques (Figure 4(b and c)).

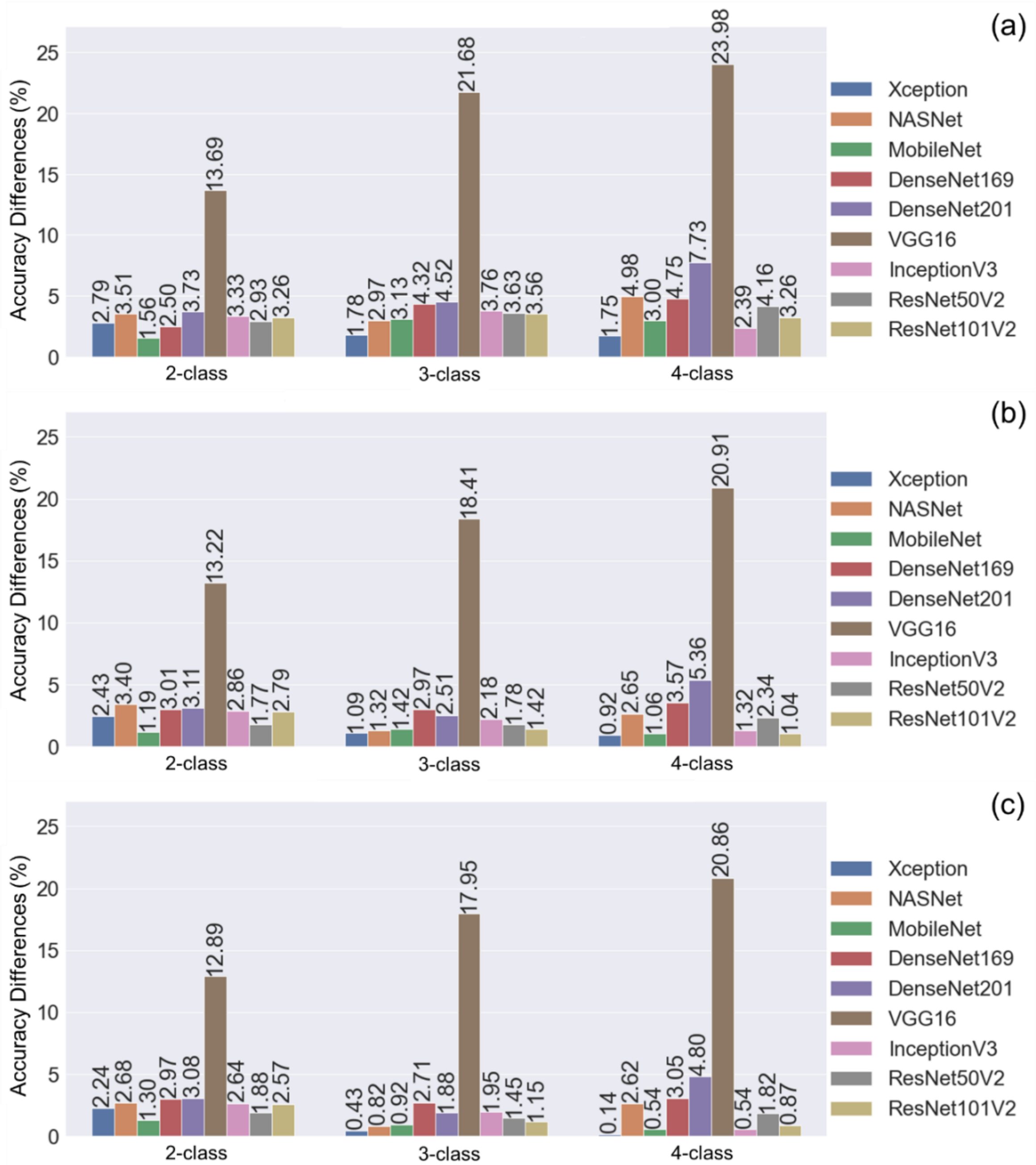

Figure 4. Amount of improvement in test accuracies of TL models grouped by number of classes.

Figure 5 (a, b, and c) shows the confusion matrices for the applications with four, three and two classes, respectively, where the highest accuracy values are obtained for the test dataset. In the proposed DeepFeat-E hybrid model, the highest accuracy values were obtained with the DenseNet201 TL model's deep features (DenseNet201F) and Stacking ensemble learning method.

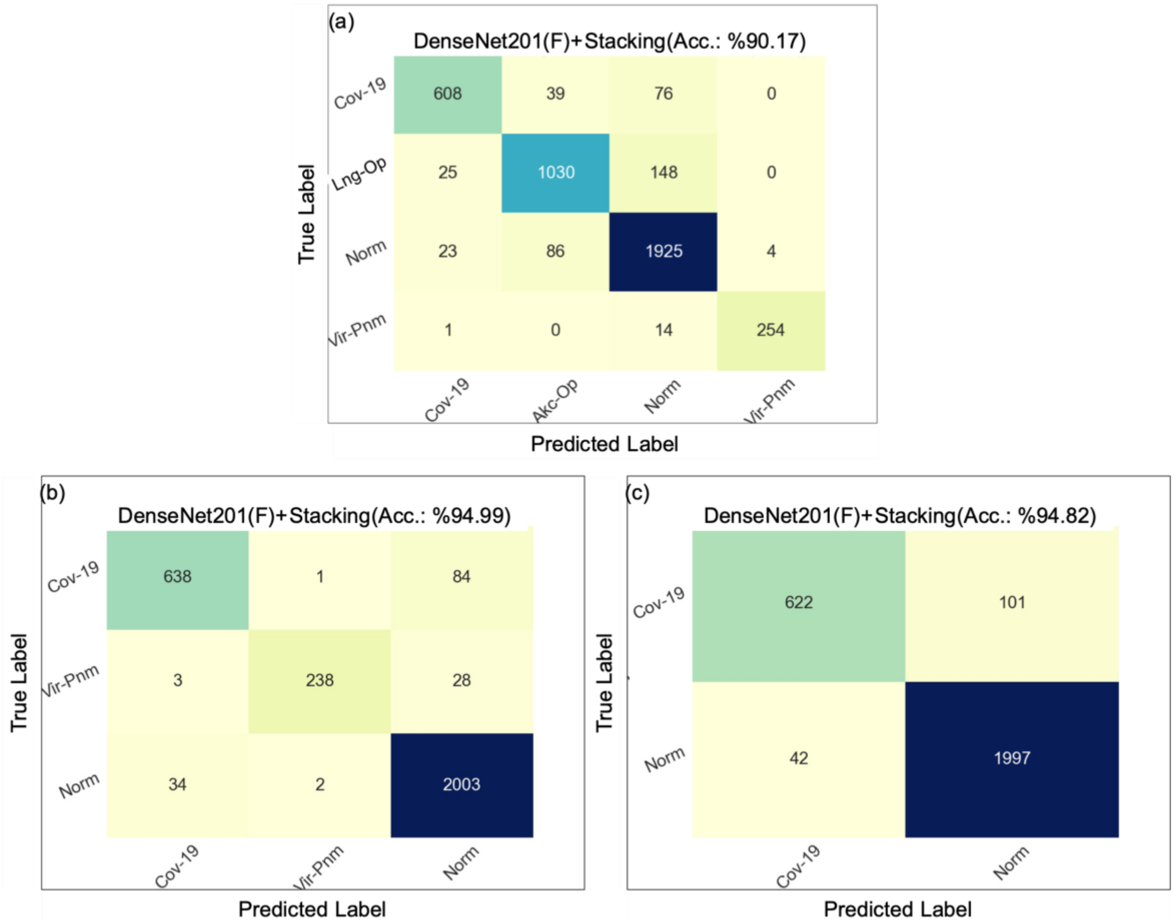

Figure 5. Best test accuracy confusion-matrixes for four-class (a), three-class (b) and two-class (c) applications.

5. Discussion

In this study, we propose a hybrid artificial intelligence system, DeepFeat-E, based on deep features extracted from X-Ray images using pre-trained TL models and an ensemble learning structure in which these features are processed by five selected best classical ML models. Although there are many artificial intelligence systems based on deep networks for COVID-19 diagnosis in the literature, this study differs from them since it is based on ensemble learning methods [6,15,25–29]. On the other hand, although there are similar studies using ensemble classifiers in the literature, it is seen that there are various studies in which snapshots of the same convolutional neural network or TL model during the training process are used as separate classifiers in the ensemble classifier [19–24]. However, in the system proposed in this study, deep features are extracted from pre-trained TL models and these features are classified in ensemble learning methods using classical ML methods. In the proposed system, there is only training in ensemble learning methods. In addition, other aspects that make the study valuable are the size of the dataset, the number of classes, the variety of the used TL models and the way in which the deep features are obtained and processed.

For comparison purposes, the performances of the hybrid model proposed in this study and other similar studies in the literature are listed in Table 7. The table shows the deep learning method used in the studies on X-Ray images, whether they are ensemble learning or not, the number of examples and classes in the datasets, and the accuracy values obtained as percentages. The table shows that the success rates vary according to the number of categorical classes of the images and the datasets used.

Table 7. Comparison of studies diagnosing COVID-19 from X-Ray images.

Author	**Method**	**Ensemble**	**Number of Samples**	**# Class**	**Acc. (%)**
Karim et al. [21]	DeepCOVIDExplainer	Yes	COVID-19: 358, Pneumonia: 5538, Normal: 8066	3	96.10
Mahmud et al. [20]	CovXNet	Yes	COVID-19: 305, Pneumonia-Bact.: 305, Pneumonia-Vir.: 305, Normal: 305	4 3 2	90.3 89.6 97.4
Chowdhur et al. [19]	ECOVNet	Yes	COVID-19: 589, Pneumonia: 6053, Normal: 8851	3	97.00
Apostolopoulos and Mpesiana [25]	VGG19	No	COVID-19: 224, Pneumonia-Bact.: 700, Normal: 504	3 2	93.48 98.75
Wang et al. [6]	COVID-Net	No	COVID-19: 358, Pneumonia: 5538, Normal: 8066	3	93.30
Ozturk et al. [26]	DarkCovidNet	No	COVID-19: 125, Pneumonia: 500, No-Findings: 500	3 2	87.02 98.08
Khan et al. [27]	CoroNet	No	COVID-19: 284, Pneumonia-Bact.: 330, Pneumonia-Vir.: 327, Normal: 310	4 3 2	89.65 94.59 99.00
Ahamed et al. [15]	Modified ResNet50V2	No	COVID-19: 1143, Pneumonia-Vir.: 1150, Pneumonia-Bact.: 1150, Normal: 1150	4 3 2	96.45 97.24 99.35
Tang et al. [22]	EDL-COVID	Yes	COVID-19: 573, Pneumonia: 6053, Normal: 8851	3	95.00
Huang and Liao [28]	LightEfficientNetV2	No	COVID-19: 600, Pneumonia: 600, Normal: 600	3	98.33
Islam et al. [29]	Cov-RADNet	No	COVID-19: 3616, Lung-Opacity: 6012, Pneumonia-Vir.: 1345, Normal: 10192	4 3 2	97.00 99.50 99.72
Banerjee et al. [23]	DenseNet-201+BlendingwRF	Yes	COVID-19: 568, Pneumonia: 6052, Normal: 8851	3	94.55
Banerjee et al. [23]	DenseNet-201+ Blending(RF)	Yes	COVID-19: 219, Pneumonia: 1345, Normal: 1341	3	94.13
Gour and Jain [24]	Stack CNN	Yes	COVID-19: 546, Pneumonia: 1355, Normal: 1139	3	97.27
This work	DeepFeat-E	Yes	COVID-19: 3616, Lung-Opacity: 6012, Pneumonia-Vir.: 1345, Normal: 10192	4 3 2	**90.17** **94.99** **94.82**

In the proposed system in this study, the highest accuracy values were obtained by using the deep features of the DenseNet201 TL model and the Stacking transfer learning method. It is seen that the accuracy values of %82.45, %90.47 and %91.09 are obtained in four, three and two class test datasets, respectively, in the applications where the CT model is used directly (Table 4). On the other hand, the proposed hybrid system improves the test accuracies to %90.17 (+7.73 points), %94.99 (+4.52 points) and %94.82 (+3.73 points) for four, three and two class datasets, respectively (Table 5 and Table 6). Therefore, it is understood that the proposed hybrid approach improves the success performance in all the applications performed in this study and has an acceptable accuracy. Accordingly, it is understood that in all four, three and two-class data sets, at least 9 out of 10 samples could be correctly diagnosed. Compared to other studies in the literature, there are studies with higher accuracy than the present study (Table 7). However, the fact that the data set used in this study is much larger and different data sets are used are thought to be the main reasons underlying this difference in success. Using the same dataset, Islam et al. [29] obtained more successful results in their study (Tablo 7). However, in their study, they performed balanced analyses by equalizing the number of class-based instances of this dataset, which contains different instances for each class. On the contrary, in this study, the number of class-based instances in the dataset was not equalized and the analyses were performed using an unbalanced number of class instances.

In addition, it has been observed that the proposed DeepFeat-E hybrid model is more successful and increases the diagnostic accuracy compared to the performances obtained when the TL models are used directly. Accordingly, for the four-class test dataset, the highest accuracy was %85.09 with MobileNet when using the TL models directly, and when the proposed hybrid system was used for the same model, the accuracy increased to %88.09 (+3 points) with the Stacking ensemble learning method. Likewise, when TL models are used directly for three and two-class test data sets, the highest accuracy was obtained with the MobileNet TL model as %90.53 and %91.53, respectively. Here too, the accuracy increases to %93.99 (+2.46 points) and %93.67 (+3.13 points) with the Stacking ensemble learning method when the proposed hybrid system is used. As seen in Table 5 and Table 6, all of the ensemble classifiers have higher accuracy values than the TL models, so it can be said that the proposed hybrid system significantly increases the success. As a result, it is seen that the proposed hybrid model is more successful than TL models in diagnosing COVID-19 and other lower respiratory tract infections and has an acceptable level of success with accuracy values of over %90 when compared to other studies in the literature.

6. Conclusion

Reducing the impact of both human health and economic damages caused by the COVID-19 pandemic is directly related to the rapid and accurate diagnosis of the disease. Therefore, rapid diagnosis, identification, treatment and isolation of COVID-19 are of utmost importance. Methods used in the diagnosis of the disease, such as PCR testing or manual interpretation of CT or X-Ray images, are known traditional methods. However, since they are faster and safer methods, medical image-based artificial intelligence systems and especially deep learning methods are successfully applied in the diagnosis of COVID-19.

Although there are similar artificial intelligence studies on COVID-19 diagnosis using ensemble learning methods in the literature, it is seen that snapshots obtained in the same training process are used as ensemble classifiers in these studies. With the hybrid diagnostic system named DeepFeat-E proposed in this study, it is attempted to diagnose COVID-19 from X-Ray images using deep features obtained from pre-trained TL models and classifiers consisting of classical machine learning methods. It was observed that the proposed system achieved the highest success with the deep features of DenseNet201 TL models and the Stacking ensemble learning method. Accordingly, the test accuracy was 90.17%, 94.99% and 94.82% for four, three and two class applications, respectively. It was also observed that the system increased the accuracy values obtained in all TL models by varying amounts (Table 6). The fact that the proposed system uses pre-trained TL models has significant advantages such as eliminating the need for big data for training the models and reducing resource and time costs. Therefore, the results obtained in this study show that the proposed DeepFeat-E hybrid system can be used quickly and reliably in the diagnosis of COVID-19 and lower respiratory tract infections.

References

[1] Kumari S, Ranjith E, Gujjar A, Narasimman S, Zeelani HSAS. Comparative analysis of deep learning models for COVID-19 detection. Global Transitions Proceedings 2021;2:559–65.

[2] Pang L, Liu S, Zhang X, Tian T, Zhao Z. Transmission dynamics and control strategies of COVID-19 in Wuhan, China. J Biol Syst 2020;28:543–60.

[3] Abir FF, Alyafei K, Chowdhury MEH, Khandakar A, Ahmed R, Hossain MM, et al. PCovNet: A presymptomatic COVID-19 detection framework using deep learning model using wearables data. Comput Biol Med 2022;147:105682.

[4] Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. Lancet Infect Dis 2020;20:631–3.

[5] Makris A, Kontopoulos I, Tserpes K. COVID-19 detection from chest X-Ray images using Deep Learning and Convolutional Neural Networks. 11th hellenic conference on artificial intelligence, 2020, p. 60–6.

[6] Wang Y, Kang H, Liu X, Tong Z. Combination of RT-qPCR testing and clinical features for diagnosis of COVID-19 facilitates management of SARS-CoV-2 outbreak. J Med Virol 2020.

[7] Verma SS, Prasad A, Kumar A. CovXmlc: High performance COVID-19 detection on X-ray images using Multi-Model classification. Biomed Signal Process Control 2022;71:103272.
[8] Tung H, Tekin R. New Feature Extraction Approaches Based on Spatial Points for Visual-Only Lip-Reading. Traitement Du Signal 2022;39.
[9] Noyan T, Kuncan F, Tekin R, Kaya Y. A new content-free approach to identification of document language: Angle patterns. Journal of the Faculty of Engineering and Architecture of Gazi University 2022;37:1277–92.
[10] Kuncan F, Kaya Y, Tekin R, Kuncan M. A new approach for physical human activity recognition based on co-occurrence matrices. J Supercomput 2022;78:1048–70.
[11] Kaya Y, Tekin R. A Novel Feature Extraction Method for Classification of Epileptic EEG Signals. Journal of Natural & Applied Sciences 2018.
[12] Kaya Y, Kuncan F, Tekin R. A New Approach for Congestive Heart Failure and Arrhythmia Classification Using Angle Transformation with LSTM. Arab J Sci Eng 2022:1–17.
[13] Sobahi N, Atila O, Deniz E, Sengur A, Acharya UR. Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. Biocybern Biomed Eng 2022;42:1066–80.
[14] Ren Z, Chang Y, Bartl-Pokorny KD, Pokorny FB, Schuller BW. The acoustic dissection of cough: diving into machine listening-based COVID-19 analysis and detection. MedRxiv 2022.
[15] Ahamed KU, Islam M, Uddin A, Akhter A, Paul BK, Yousuf MA, et al. A deep learning approach using effective preprocessing techniques to detect COVID-19 from chest CT-scan and X-ray images. Comput Biol Med 2021;139:105014.
[16] Ozcan T. A new composite approach for COVID-19 detection in X-ray images using deep features. Appl Soft Comput 2021;111:107669.
[17] Gilanie G, Bajwa UI, Waraich MM, Asghar M, Kousar R, Kashif A, et al. Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks. Biomed Signal Process Control 2021;66:102490.
[18] Basu A, Sheikh KH, Cuevas E, Sarkar R. COVID-19 detection from CT scans using a two-stage framework. Expert Syst Appl 2022;193:116377.
[19] Chowdhury MEH, Rahman T, Khandakar A, Mazhar R, Kadir MA, Mahbub Z bin, et al. Can AI help in screening viral and COVID-19 pneumonia? IEEE Access 2020;8:132665–76.
[20] Mahmud T, Rahman MA, Fattah SA. CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. Comput Biol Med 2020;122:103869.
[21] Karim MR, Döhmen T, Cochez M, Beyan O, Rebholz-Schuhmann D, Decker S. Deepcovidexplainer: explainable COVID-19 diagnosis from chest X-ray images. 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE; 2020, p. 1034–7.
[22] Tang S, Wang C, Nie J, Kumar N, Zhang Y, Xiong Z, et al. EDL-COVID: Ensemble deep learning for COVID-19 case detection from chest X-ray images. IEEE Trans Industr Inform 2021;17:6539–49.
[23] Banerjee A, Sarkar A, Roy S, Singh PK, Sarkar R. COVID-19 chest X-ray detection through blending ensemble of CNN snapshots. Biomed Signal Process Control 2022;78:104000.
[24] Gour M, Jain S. Automated COVID-19 detection from X-ray and CT images with stacked ensemble convolutional neural network. Biocybern Biomed Eng 2022;42:27–41.
[25] Apostolopoulos ID, Mpesiana TA. Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med 2020;43:635–40.
[26] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792.
[27] Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed 2020;196:105581.
[28] Huang M-L, Liao Y-C. A lightweight CNN-based network on COVID-19 detection using X-ray and CT images. Comput Biol Med 2022:105604.
[29] Islam MK, Habiba SU, Khan TA, Tasnim F. COV-RadNet: A Deep Convolutional Neural Network for Automatic Detection of COVID-19 from Chest X-Rays and CT Scans. Computer Methods and Programs in Biomedicine Update 2022;2:100064.
[30] Rahman T, Khandakar A, Qiblawey Y, Tahir A, Kiranyaz S, Kashem SBA, et al. Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images. Comput Biol Med 2021;132:104319.
[31] Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, p. 1251–8.
[32] Zoph B, Vasudevan V, Shlens J, Le Q v. Learning transferable architectures for scalable image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, p. 8697–710.

[33] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. ArXiv Preprint ArXiv:170404861 2017.
[34] Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, p. 4700–8.
[35] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. ArXiv Preprint ArXiv:14091556 2014.
[36] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, p. 2818–26.
[37] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. European conference on computer vision, Springer; 2016, p. 630–45.
[38] Sarkar D, Bali R, Ghosh T. Hands-On Transfer Learning with Python: Implement advanced deep learning and neural network models using TensorFlow and Keras. Packt Publishing Ltd; 2018.
[39] Zhou K, Greenspan H, Shen D. Deep learning for medical image analysis. Academic Press; 2017.
[40] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63:3–42.
[41] Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, et al. Feature selection: A data perspective. ACM Computing Surveys (CSUR) 2017;50:1–45.
[42] Zhou Z-H. Ensemble methods: foundations and algorithms. CRC press; 2012.
[43] Wolpert DH. Stacked generalization. Neural Networks 1992;5:241–59.
[44] Bohmrah MK, Kaur H. Classification of Covid-19 patients using efficient fine-tuned deep learning DenseNet model. Global Transitions Proceedings 2021;2:476–83.