Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 12 NO: 4 PAGE: 1015-1027 YEAR: 2023
DOI:10.17798/bitlisfen.1312360

Multi Deep Learning Based Approaches for COVID-19 Diagnosis Using Class Resampling on Chest X-Ray Images

Talha Burak ALAKUŞ[1*], Muhammet BAYKARA[2]

[1]*Kırklareli University, Engineering Faculty, Department of Software Engineering, Kırklareli, Turkey*
[2]*Fırat University, Technology Faculty, Department of Software Engineering, Elazığ, Turkey*
(ORCID: 0000-0003-3136-3341) (ORCID: 0000-0001-5223-1343)

Keywords: Deep learning, Artificial intelligence, Medical imaging, Transfer learning, Classification

Abstract

Nowadays, current medical imaging techniques provide means of diagnosing disorders like the recent COVID-19 and pneumonia due to technological advancements in medicine. However, the lack of sufficient medical experts, particularly amidst the breakout of the epidemic, poses severe challenges in early diagnoses and treatments, resulting in complications and unexpected fatalities. In this study, a CNN (Convolutional Neural Network) model, VGG16 + XGBoost and VGG16 + SVM hybrid models, were used for three-class image classification on a generated dataset named Dataset-A with 6,432 chest CXR (Computed X-Ray) images (containing Normal, Covid-19, and Pneumonia classes). Then, pre-trained ResNet50, Xception, and DenseNet201 models were employed for binary classification on Dataset-B with 7,000 images (consisting of Normal and Covid-19). The suggested CNN model achieved a test accuracy of 98.91%. Then the hybrid models (VGG16 + XGBoost and VGG16 + SVM) gained accuracies of 98.44% and 95.60%, respectively. In our experiments, accuracy rates of 98.90%, 99.14%, and 99.00% were achieved for the fine-tuned ResNet50, Xception, and DenseNet201 models, respectively. Finally, the models were further evaluated and tested, yielding impressive results. These outcomes demonstrate that the models can aid radiologists with robust tools for early lungs related disease diagnoses and treatment.

1. Introduction

Just recently, the entire world was in a state of extreme fear and trepidation due to the outburst of the novel COVID-19 (Corona Virus Disease 2019) epidemic that started in Wuhan, Hubei province of China, in December 2019 [1], [2]. This virus, also known as SARS-CoV-2 or the severe acute respiratory syndrome coronavirus 2, holds an unknown etiology and is zoonotic, meaning it can propagate from animals to human beings [3]. Because of the novel virus's airborne nature, any infected person can spread it to people around them by merely breathing, speaking, sneezing, or coughing [4]. The above-mentioned nature of the virus made it easy to circulate rapidly. Seeing the rate at which the virus spread, the WHO (World Health Organization) assessed and characterized it all together as a global epidemic on March 11, 2020 [5]. The virus poses serious illness, particularly to the elderly, children, and persons with underlying medical conditions [6–8]. According to WHO, Worldometer, and Statista websites, there have been over 6.5 million reported deaths and over 635 million documented cases of SARS-CoV-2 infections worldwide [9]–[11].

The most effective means of combating such diseases being early testing and diagnosis are challenging, especially in the event of a worldwide epidemic, when the number of infected people exceeds the capacity of hospitals and healthcare professionals. Moreover, most of the widely accepted methods of testing the COVID-19 virus are not without caveats, making them unsuited for application in early testing and diagnosis. Even the

*Corresponding author: talhaburakalakus@klu.edu.tr
Received: 12.06.2023, Accepted: 11.10.2023

widely used and validated method, RT-PCR (Reverse Transcription - Polymerase Chain Reaction), is not entirely effective, especially in patients with low viral loads. Also, the cost of running laboratories that require expensive instruments, affirmative examinations, and trained personnel remains a significant drawback [12]. Due to the advancement and diversity of various medical imaging and CAD (Computer Aided Diagnoses) techniques, mainly in the field of radiology and oncology, AI-based models provide a quick and inexpensive means of diagnosing COVID-19 and other related diseases [13]. The primary cause of this significant milestone is the remarkable breakthrough these models have accomplished, particularly in image classification and segmentation tasks for both binary and multi-class with high accuracy [14]. It is now simpler to employ the power of these models to support radiologists with early diagnostic techniques even amid a global pandemic, because of the abundance of medical images from technologies like MRI (Magnetic Resonance Imaging), CT (Computed Tomography) scans, and X-ray [15]. Because of their improved classification performance, robustness, and enormous data processing power, various artificial intelligence algorithms significantly reduce the limitations of human medical experts, thereby mitigating the occurrence of severe illnesses and deaths [16].

Several recent studies employed deep learning models based on conventional CNN to class medical chest images into various classes. In study [17], researchers proposed a CNN model that classified CXR images as either COVID-19 cases or normal, achieving 96.71% accuracy. In another study researchers proposed other sets of CNN models for image segmentation and classification, reaching the highest accuracy of 93.2% [18]. Also, fine-tuning pre-trained models for detecting COVID-19 from several medical images is another widely deployed method by many researchers. In study [19] researchers utilized the ResNet18 model on CXR to identify COVID-19 with 93% accuracy. In study [20] ResNet50V2, DenseNet201, and InceptionV3 models were employed on CXR images attaining an accuracy of 91.62%. The hybridization of deep learning models with machine learning classifiers provides a fast and efficient means of classifying medical images, particularly in the presence of limited training data. In another study authors combined DenseNet201 with a Random Forest (RF) classifier to detect COVID-19 on CXR images with an accuracy of 94.55% [21]. Lastly, in study [22] researchers applied a hybrid of several pre-trained models with the SVM (Support Vector Machine) classifier to detect COVID-19 with 94.7% accuracy.

This study aims to demonstrate the performance of three categories of AI-based models on CXR images for binary and three-class image classification tasks utilizing two openly accessible datasets from Kaggle [23]-[24]. After applying random oversampling to address class-imbalance, a CNN and two hybrid models (using VGG16 to extract deep-level features followed by XGBoost and SVM classifiers to categorize the obtained features) were employed for three-class image classification on our generated Dataset-A. There are 6,342 images in total in Dataset-A, and there are three classes in this dataset: normal, pneumonia and COVID-19. Then, Dataset-B was used for binary image classification using three pre-trained models: Xception, ResNet50, and DenseNet201, using transfer learning and fine-tuning techniques. There are 7,000 images in total in the Dataset-B, of which 3,500 are normal and 3,500 are COVID-19. In this data set, there are two class labels. Finally, the study's outcome was analyzed and compared with recent works on this area. To fully appreciate the relevance of this study, we precisely summarized its contributions as follows:

- Employing the random oversampling technique to handle the issue of class imbalance and limited input data significantly improves the performance of the models. High test accuracies on unseen images, indicating less overfitting and more generalizable models, are evidence of these improvements.
- Utilizing proper image pre-processing techniques, like resizing and cropping the images to remove unnecessary features and enable the models to learn the appropriate ones, has significantly improved their performance and conserved memory and resources during training.
- Training different models, including standard convolutional neural networks, hybrid models, and pre-trained models that achieve high test accuracies, demonstrates that this research can undoubtedly aid radiologists in making accurate early diagnoses, even during pandemics. In turn, this could reduce severe illness and save lives.

The remainder of the article is organized as follows. In the second section, information about the data sets and methods used in the study were given. Moreover, the study encompassed an examination of the deep learning models that were considered for evaluation. In the third section, the results of the classifiers were given, and the evaluations of the classifiers were made. This section also involved the examination and comparison of studies from existing

literature with the findings obtained in this current study. In the fourth section, the advantages and disadvantages of this study were mentioned by discussion. In the last section, the study was summarized and its contributions to the literature were mentioned.

2. Material and Method

In this section, the data set, the image processing steps, and the deep learning models were mentioned.

2.1. Dataset A

Dataset-A consists of 6,432 CXR images from the following categories: Normal, pneumonia, and COVID-19 [23]. Table 1 illustrates the image allocation in the train and test folders. As shown in Table 1, the dataset comprises 1,266 and 317 normal images, 3,418 and 855 pneumonia-infected images, and 460 and 116 COVID-19 positive images in the train and test folders, respectively.

Table 1. Distribution of the unprocessed and unbalanced CXR images in Dataset A

Data	Normal	Pneumonia	COVID-19	Total
Train	1,266	3,418	460	5,144
Test	317	855	116	1,288
Total	1,583	4,273	576	6,432

2.2. Dataset B

The initial dataset is created in partnerships between medical professionals and researchers from various universities from Qatar, Bangladesh, Malaysia, and Pakistan [24]. Dataset-B, our created dataset in this study, is composed by randomly selecting 3,500 CXR images from the COVID-19 positive and the normal CXR image folders. A total of 7,000 images were formed and used for training and validation after applying several data augmentation techniques like rescaling, zooming, rotation, etc. Table 2 and Table 3 show the breakdown of the fully pre-processed and balanced data distributed among the classes for the binary and three-class image classifications for Dataset-A and Dataset-B in train, test, and validation categories.

Table 2. Distribution of the preprocessed and balanced CXR images in Dataset-A

Data	Normal	Pneumonia	COVID-19	Total
Train	3,418	3,418	3,418	10,254
Test	428	428	428	1,284
Val.	427	427	427	1,281
Total	4,273	4,273	4,273	12,819

Table 3. Distribution of the preprocessed and balanced CXR images in Dataset-B

Data	Normal	COVID-19	Total
Train	2,800	2,800	5,600
Test	350	350	700
Val.	350	350	700
Total	3,500	3,500	7,000

2.3. Random Oversampling

The random oversampling technique offers a naive and straightforward approach to harmonizing the class allocations of the imbalanced dataset. The Random Over Sampler [25] is employed in this study to generate new samples of the under-represented classes by arbitrarily sampling and replacing the existing images. The issue of overfitting is appropriately addressed by using data augmentation, dropout, batch normalization, and callbacks. Also, the data were adequately distributed into training, validation, and testing, with the testing data being withheld and then utilized to evaluate the performance of the models on the never-before-seen data.

2.4. Data Augmentation

Data augmentation involves creating additional data samples from existing data to boost the extent and diversity of the training data. Augmentation may include making minimal modifications to data or utilizing machine learning models to create extra data samples in the underlying space of the initial data to augment the dataset and reduce overfitting. Several geometric transformations, like translation, flipping, rescaling, rotation, and so on, are used on the original images to generate numerous variations of each image. These rendered images will appear different to the classifying algorithms, thereby increasing the volume of data and curtailing the over-memorization issue that learning algorithms fall short of due to imbalance or limited available data. It differs from synthetic data generation, in which data is created artificially. Several geometric transformations were made to these images. These transformations consist

of a horizontal flip, the nearest fill mode, a channel shift range of 10, a rotation range of 40 degrees, and a width shift, height shift, shear, and zoom range values of 0.2. Data augmentation was carried out by considering geometric transformations. These formations cover the meta learning. Meta learning is also an effective approach in convolution based neural networks [26]. For this reason, geometric transformations were employed in this study. Figure 1 provides random samples of images that were created because of using numerous data augmentation approaches in the study.

Figure 1. A random sample of the augmented images.

In the study, the original data was first augmented and then separated as training, validation, and test data. To construct effective deep learning models, it is vital that the validation error consistently diminishes along with the training error. Data augmentation stands as a potent technique for accomplishing this objective. By generating augmented data, a broader array of potential data instances is captured, thereby reducing the gap between the training and validation sets, as well as any subsequent test sets. In this way, the problem of overfitting can be avoided [26].

2.5. The Proposed CNN Model

After applying the random oversampling and necessary data pre-processing to Dataset A, the data types were converted from 8 bits unsigned integers to 32 bits floating point numbers using the assign-type python command. This step is crucial because all the mathematical operations in the deep learning process involve continuous rather than discrete values. Then these values were further normalized to the range [0, 1] by dividing by the highest image pixel value of 255 to aid with faster computations and reduce exhausting unnecessary computer resources. The final step before building the proposed model involved splitting the pre-processed and normalized images into 80:10:10 ratios for training, validation, and testing. These ratios signify that the model will use 80 % of the input images during the training phase, 10 % for the validation phase, and the remaining 10 % for the testing phase. Figure 2 shows a visualized depiction of the proposed architect to aid with the quick assimilation.

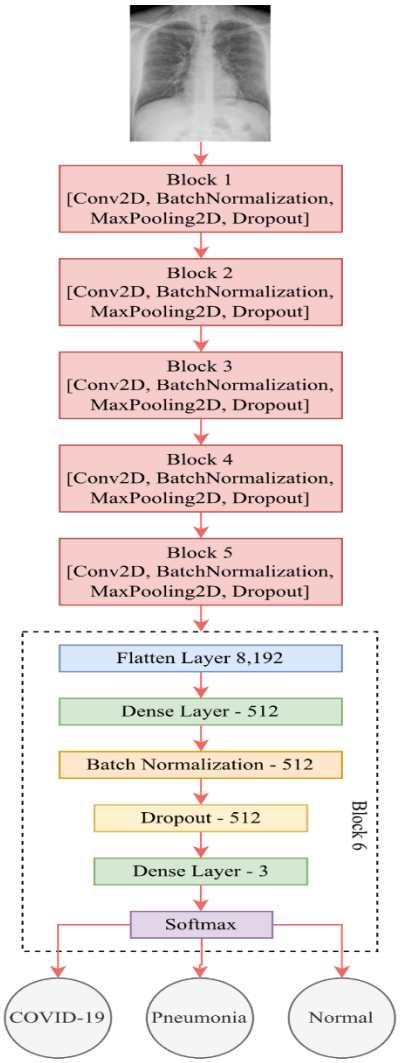

Figure 2. The proposed CNN architecture.

The model was later compiled by an RMSprop optimizer with a 0.001 learning rate, sparse categorical cross-entropy loss function since the input data were not hot encoded, and an accuracy metric. The compiled model generated 6,571.651 parameters due to the weights and biases, of which only 3,840 are not trainable. Furthermore, two callback functions, namely early stopping and reduced learning rate, were used during the training process to aid with quick convergence and mitigate the overfitting of the suggested model. In Table 4, the tune parameters of the CNN model proposed in the study are given.

Table 4. Hyperparameters of the proposed CNN model

Hyperparameters	Value
Number of epoch	60
Learning rate	0.001
Batch size	32
Loss function	Sparse categorical cross entropy (multi-class), binary cross entropy (binary class)
Optimizer	RMSProp

2.6. Hybrid Models

The second sets of architecture proposed in the paper is the hybrid model. These hybrid models are an example of the late fusion method of deep hybrid learning techniques. Late fusion because only at the last part of the process a particular classifier is used to make predictions. Deep learning methods, like the pre-trained models have remarkable feature-extracting power (VGG16 in this case), which are leveraged to perform automatic feature extraction, where classical machine learning classifiers are employed to make predictions from the generated features of Dataset A. Initially, the VGG16 pre-trained model was loaded without the top fully connected layers with an input shape of 150 × 150 × 3 using the ImageNet weights and making the loaded layers non-trainable to allow exclusively feature extraction and avoid retraining the model from scratch. Then the same images from Dataset A in the earlier CNN model were passed through the VGG16 feature extractor. Two classifiers, XGBoost and SVM, were utilized to provide predictions for three-class image categorization based on the derived features from the training and validation sets of data. Figure 3 reveals a graphical illustration of the proposed model architecture for more comprehension of the whole process. Later, the test set of images was utilized to evaluate the implementation of the classifiers, resulting in more accurate classifications. Other performance indicators were applied for further analysis, including the confusion matrix, F1-score, precision, and recall.

2.6.1. VGG16

This architecture is a prominent CNN model presented by [27], which enhances its predecessor, the AlexNet model, by substituting the 11 × 11 and 5 × 5 kernels in the first two convolution layers with several 3 × 3 ones in succession. The model is about 528 MB in size with a recorded 90.1% top-5 accuracy on ImageNet data and about 138.4 million parameters. The ImageNet dataset possesses around 14 million images from 1,000 categories. VGG16 was trained on powerful GPUs over a period of several weeks.

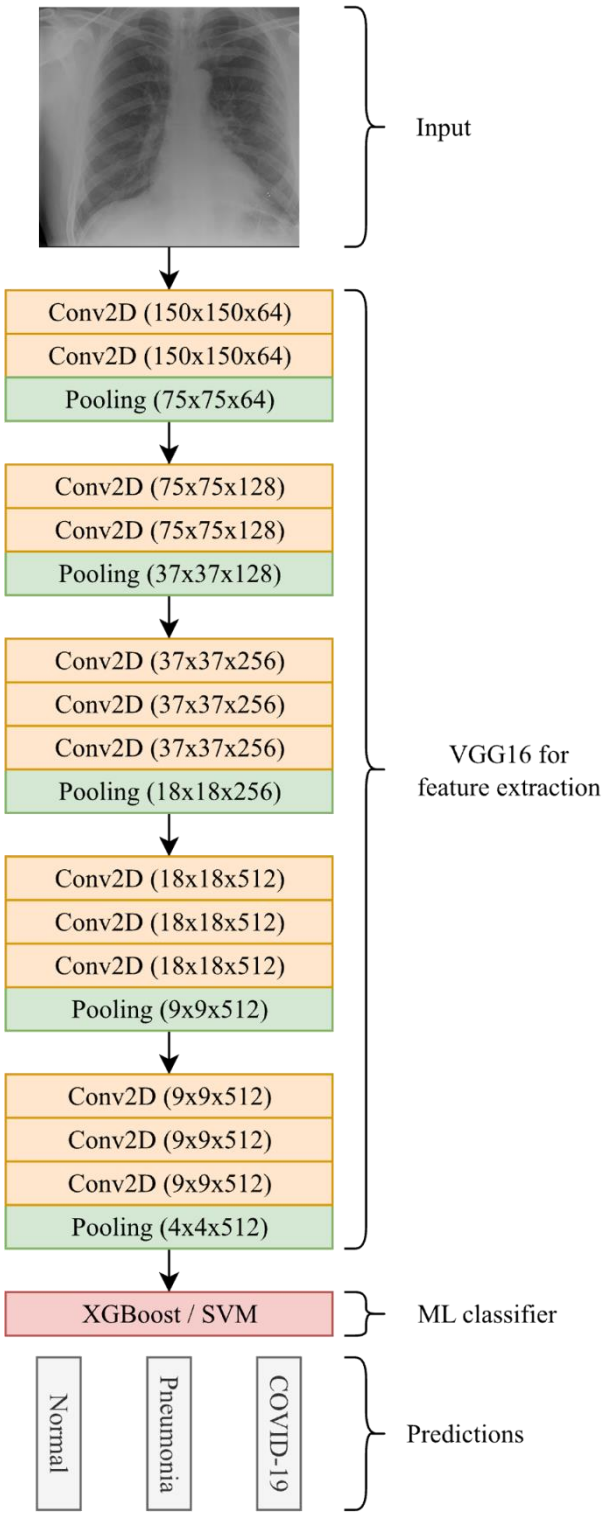

Figure 3. The architecture of the hybrid model.

2.6.2. XGBoost

Coined after extreme gradient boosting, it is a collection of scalable and distributed gradient boosting that are prominent for their high efficiency, flexibility, and portability, which executes machine learning algorithms under the gradient-boosted frameworks. XGBoost presents a similar tree-boosting approach known as GBDT (Gradient-Boosted Decision Tree) or GBM (Gradient Boosting Machine) that efficiently and precisely solves classification and regression problems involving billions of cases or more [28].

2.6.3. SVM

SVM is among the most prominent algorithms for classification and regression issues along with outlier detection in supervised learning circumstances. SVM is a machine learning technique that strives to classify data points operating a hyperplane in the space of N-dimensions with N as the number of features. It is a memory-efficient method of applying varying kernel operations for the decision function to solve issues where the number of dimensions surpasses the number of samples [29].

2.7. The Pre-Trained Models

Like the pre-processing steps conducted in the first dataset, the images in Dataset B were also reshaped, normalized, augmented, and split into an 80:10:10 ratio for training, validation, and testing. Then ResNet50, Xception, and DenseNet201 models were loaded with the ImageNet pre-trained weights without their respective top layers. The same layers comprising global average pooling, dropout with the probabilities of 0.2, batch normalization, and dense layers with two units using the Softmax activation functions were later added to the models to aid with the classification. Then we compiled them using the Adam optimizer, each with a learning rate of 0.00001 and a decay of 0.00001/60. Other hyperparameters used in the compilation process include categorical cross-entropy loss function and accuracy metrics. The model's training process involves using epochs of 60, a batch size of 128, along with early stopping and reduced learning rate callbacks. The models were evaluated individually using accuracy, F1-score, precision, recall, and confusion matrix.

2.7.1. ResNet50

It is a kind of ANN (Artificial Neural Network) that produces networks by piling blocks of residual connections, as demonstrated by the 50-layer ConvNet featuring 48 convolution layers, a maximum, and an average pooling layer [30]. Trained on ImageNet data, the 224 × 224 input-sized networks can categorize images into a thousand object groups.

2.7.2. Xception

This architecture was created to improve the fundamentals of the Inception model, where 1 × 1 convolution operations were used to compress input data and further utilize varying filters on each data's depth space. Extreme Inception, short for Xception, essentially reverses this mentioned process. Conversely, it applies the filters to each depth map before performing the exact convolution processes across the depth to shrink the input space. This technique is known as the depth-wise separable convolution operation [31].

2.7.3. DenseNet201

This network has a 201-layer design in which directly linked channels form dense connections. Each layer obtains extra inputs from previous ones and sends its feature maps to succeeding layers utilizing a concatenated strategy. Splitting the 3 × 3 convolution operation into a 1 × 1 and a 3 × 3 minimizes the quantity of the model parameters since each layer gets feature maps from subsequent ones. The growth rate denoted by *k* signifies that the amount of feature maps rises by *k* each moment the dense block is traversed, which helps to lower the parameters. The dense blocks are joined by a transitional layer that significantly minimizes the number of features and nets out the most effective ones in each layer [32].

With the mentioned models and designed models, chest x-ray images were used in the study and the classification process was carried out. The flow chart of the study is given in Figure 4.

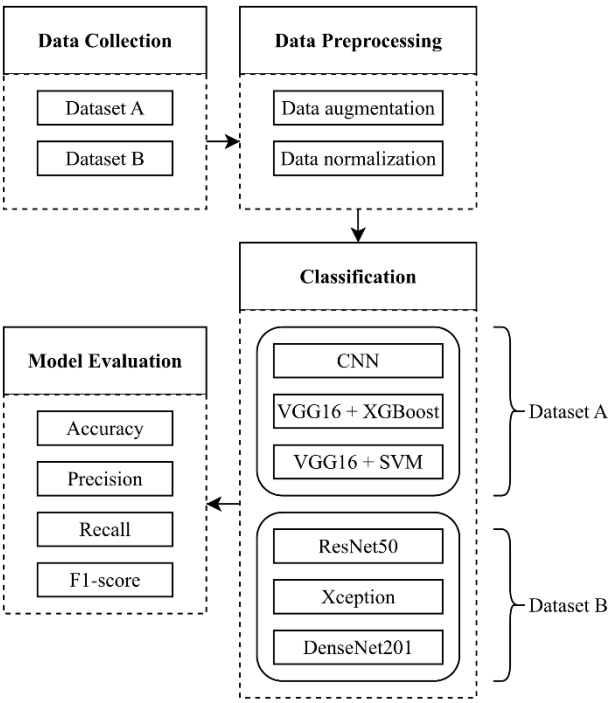

Figure 4. Flow chart of the study.

As can be seen in Figure 4, the study consists of four stages. In the first stage **(Data Collection)**, two different data sets containing CXR images were obtained, and the data sets were named as Data Set A and Data Set B. In the second stage **(Data Preprocessing)**, data augmentation was performed on the images in the data sets and the size of the data sets was increased. Afterwards, the normalization process was performed, and the pixels of the images were reduced to the range of [0, 1]. In the third stage **(Classification)**, classification was carried out according to the data sets. For Data Set A CNN, VGG16+XGBoost and VGG16+SVM classifiers were employed. On the Data Set B, deep learning models ResNet50, Xception and DenseNet201 were considered, and the images were classified. In the last stage **(Model Evaluation)**, the performances of the classifiers were evaluated and for this, accuracy (acc), precision (pre), recall (rec) and F1-score (f1s) values were calculated.

3. Results and Discussion

3.1. Experimental Results

After training the models in this study with the two distinct datasets, we evaluated them using test portions of the preserved data to measure their test loss and accuracy. The plots of these losses and accuracies against the number of training epochs are known as learning curves. Figure 5 (a) and (b), and Figure 6 (a) and (b) depict the learning curves for our suggested ConvNets model, ResNet50, Xception, and DenseNet201 deep learning models, respectively. Moreover, Table 5 and Table 6 contain the summary of overall results of the experimental analysis.

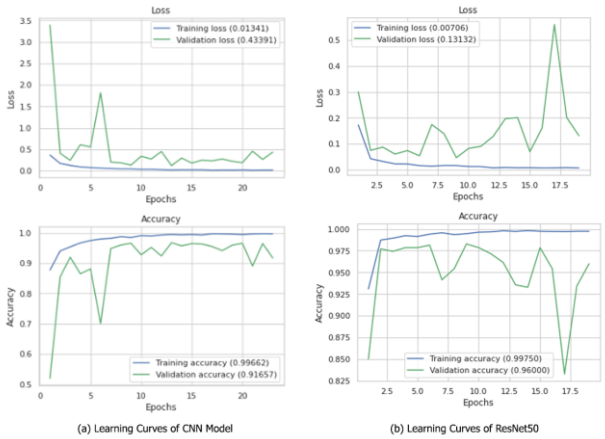

Figure 5. Learning curves of losses and accuracies of (a) CNN model, (b) ResNet50 model.

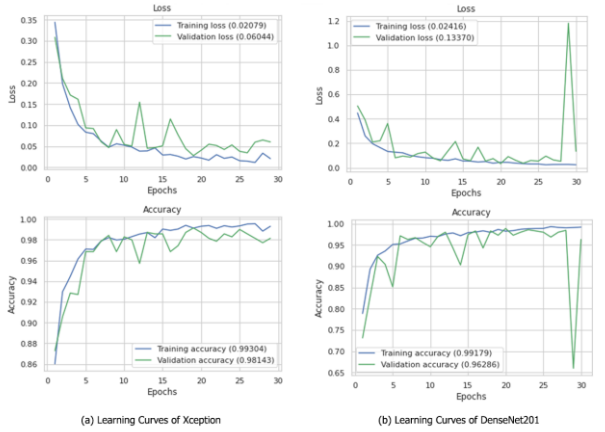

Figure 6. Learning curves of losses and accuracies of (a) Xception model, (b) DenseNet201 model.

Table 5. Results of the models on Dataset A

Models	Pre	Rec	F1s	Acc
CNN	98.92%	98.90%	98.90%	98.91%
VGG16 + XGBoost	96.00%	96.00%	96.00%	98.44%
VGG16 + SVM	96.00%	96.00%	96.00%	95.60%

Table 6. Results of the models on Dataset B

Models	Pre	Rec	F1s	Acc
ResNet50	98.86%	98.86%	98.86%	98.90%
Xception	99.14%	99.14%	99.14%	99.14%
DenseNet201	99.00%	99.00%	99.00%	99.00%

3.2. Evaluation Metrics

The metrics employed for evaluating the obtained results include accuracy, precision, recall, and F1-score. The accuracy of an artificial intelligence model is represented as the proportion of accurate predictions to all other classifications generated by the model using the input data. Precision is the ratio of the correct predictions to all the positive predictions in its dataset. The recall is expressed as the ratio of the correct predictions to the complete number of accurate entities in the given dataset. Finally, F1-score is the weighted harmonic mean of the recall and precision, ranging between zero (0) and one (1). When the value of the f-measure is increased, the model's classification performance also increases [33]. The formulas of the evaluation metrics used in the study are given between Equation 1 and Equation 4.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 * \left(\frac{precision * recall}{precision + recall}\right) \quad (4)$$

The equations portrayal of TP expression signifies True Positive values, whereas the TN expression represents True Negative values. FN corresponds to False Negative values, while FP corresponds to False Positive values.

3.3. Confusion Matrix

A confusion matrix is a table-like plot that is widely employed to describe the performance of classification models on a collection of known-true test data. It offers far more information than a straightforward accuracy score. For the multi-class case, it illustrates how inaccurate a prediction maybe when the output classes are ordinal. When dealing with imbalanced data, where there is a considerable gap between the different categories of groups, accuracy alone is insufficient. Figure 7 (a), (b), and (c) presents the confusion matrices for the ConvNet and hybrid models obtained using Dataset A.

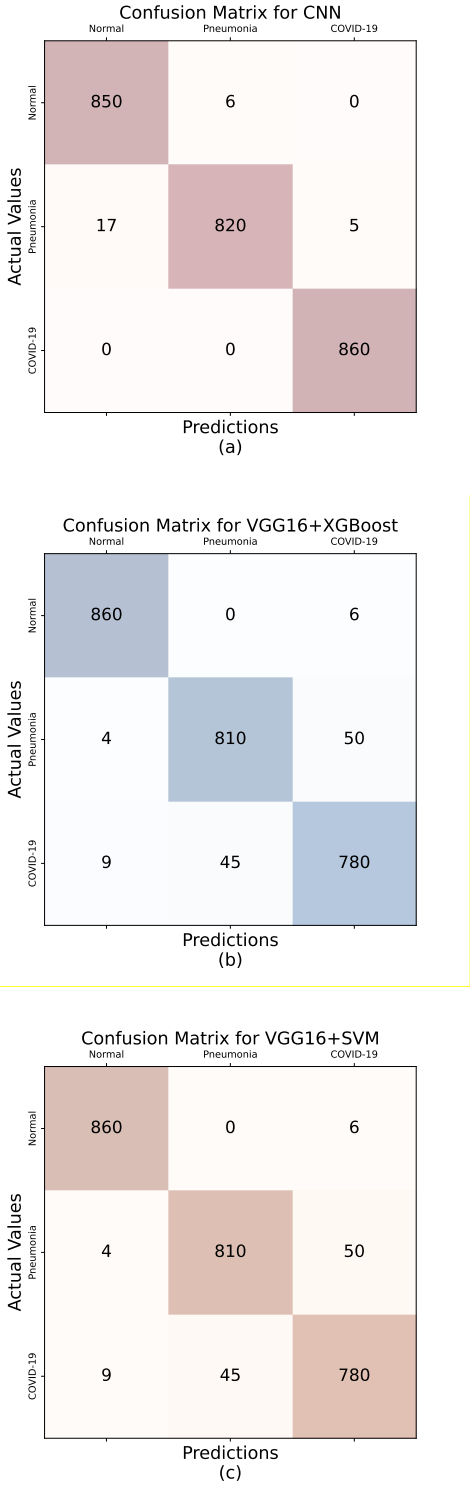

Figure 7. The confusion matrices of (a) CNN model, (b) VGG16+XGBoost model and (c) VGG16+SVM model on Dataset A.

Figure 8 (a), (b), and (c) shows the confusion matrices for the pre-trained models acquired from Dataset B. Moreover, Table 7 compares the performances of our proposed models with recent state-of-the-art studies from the literature.

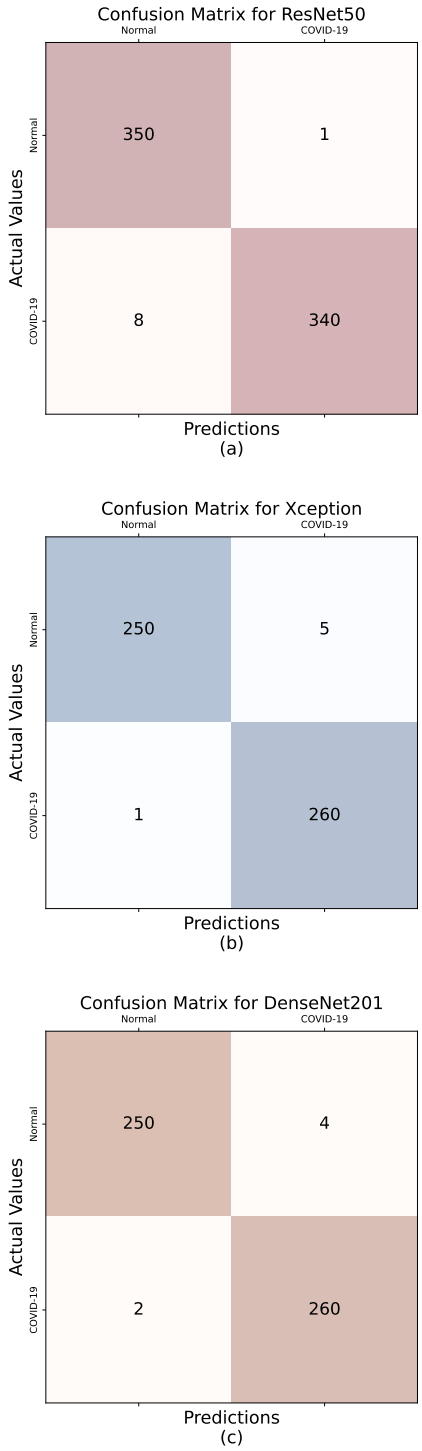

Figure 8. The confusion matrices of (a) ResNet50 model, (b) Xception model and (c) DenseNet201 model on Dataset B.

Table 7. Performance comparisons with some state-of-the-art studies from the literature

Reference	Top Accuracy
[34]	95.00%
[35]	96.00%
[36]	96.56%
[37]	87.00%
[38]	96.68%
[39]	97.40%
[40]	96.33%
[41]	97.00%
[42]	97.00%
[43]	98.05%
[44]	99.69%
[45]	96.73%
[46]	95.00%
[47]	97.00%
[48]	98.23%
This study	**99.10%**

The results of the experimental analysis conducted in this paper begins with the learning curves shown by Figure 5 and Figure 6. These plots are very significant as they indicate the performance of the models in terms of training and validation losses and accuracies. Typically, they are used to detect overfitting and underfitting in models. Based on these curves, the models perform well since the gap between average losses and accuracies is minimal. The variation in the number of epochs shown by various plots is another factor worthy of consideration concerning these graphs. These differences are due to the early stopping and reduced learning rate callback methods used during model training. These callbacks continuously monitor the models convergence to determine the optimal learning rates. Consequently, training was terminated early with epochs between 17 and 30, demonstrating the quality of the models. These factors also cause large models such as the ResNet50 and DensNet201 to have non-smooth learning curves while producing excellent results.

After obtaining the plots of the individual deep learning models learning curves, the average precision, recall, F1-score, and accuracy metrics from the test data was calculated. Table 5 and Table 6 show the summary of the experimental results obtained from the proposed models conducted on the two datasets for the three-class and binary-class image classification. The proposed CNN model achieved an estimated accuracy of 98.91%. It also attained an average precision value of 98.92%, 98.90% recall,

and 98.90% F1-score for COVID-19, pneumonia, and normal images. The VGG16 + XGBoost and VGG16 + SVM hybrid models obtained 98.44% and 95.60% accuracies, respectively. However, they recorded similar average precision, recall, and F1-score value of 96.00%. For the fine-tuned pre-trained models, the Xception model performed the best, reaching a test accuracy of 99.14%. The precision, recall, and F1-score values all showed consistent results at 99.14%. Following that, the DensNet201 model achieved an accuracy of 99.00%, with precision, F1-score, and recall also at 99.00%. Lastly, the pre-trained ResNet50 model achieved an accuracy of 98.90%, with identical values for precision, recall and F1-score, all at 98.86%.

The adoption of several essential image pre-processing techniques, such as reshaping, resizing, and cropping, as well as the class resampling techniques described previously, are among the main contributing factors for these vast models to achieve such performance with minimal computer resource consumption. In addition, the inclusion of callback functions has significantly aided in the prevention of overfitting and early convergence, as the reduced learning rate class, for instance, monitors the loss and decreases the learning rate when there is no discernible improvement. Moreover, these results are very impressive in comparison to many recent published state-of-the-arts work. Performance comparison presented by Table 7 is the evidence of this assertion since it was created by picking the best accuracies of some work conducted in the literature to the best of our knowledge and compared with the ones obtained by our proposed models. The comparison is organized into three sections to represent performance of the studies. The first section comprises [34] – [38] studies using the conventional CNN models to detect lungs diseases. The second category of studies [39] – [43] are the ones performed by fine-tuning the pre-trained models. Lastly, these studies [44] – [48] were based on the various hybrid models for COVID-19 and pneumonia detection. For the ConvNet and pre-trained models, our proposed models have the best accuracies compared to these novel studies to the best of our knowledge. Also, our proposed hybrid models are the second best in terms of accuracy. These accomplishments are only feasible due to the time and effort devoted to meticulously scrutinizing each experimental stage, including data collection, in-depth literature review, conceptualization, model training, experimental result gathering, etc. The adoption of this strategy stems from a desire to build upon recent work to create robust tools that will benefit the scientific community.

Several further studies have leveraged the might of deep learning algorithms to diagnose the novel COVID-19 disease using other techniques with high performances. For instance, to aid with fast and inexpensive methods in combating the recent pandemic, researchers developed a software that integrates the web design approach to create an interactive GUI (Graphical User Interface (GUI) [49]. Their software, CoviExpert, involves making combined independent predictions by training on 1,584 CXR images to detect COVID-19 cases with high classification accuracy. In another study, researchers aimed to minimize the performance reduction of models posed by obtaining datasets from various sources in image classification problems [50]. To achieve that, they offered a self-supervised block for feature standardization and optimization on three pre-trained models, VGG, Xception, and DenseNet, as a baseline to extract features from four CXR lung disease classification datasets and observed improved classification outcomes. Lastly, researchers in [51] suggested a hybrid multimodal framework that fused two separate models employing the weighted sum-rule fusion approach to segment and classify CXR images and collected cough samples. After using the necessary signal processing and Mel frequency cepstral coefficient to pre-process the cough samples, the fused model attained high classification accuracies for the CXR images and the cough samples. These alter-native methods to COVID-19 detection are a solid indication that one has no limit in employing deep learning approaches to develop robust models that can quickly and efficiently diagnose diseases in biomedical data with less cost and human errors.

4. Conclusion

The deep learning-based methods provide an efficient way of diagnosing chest-related diseases, particularly the recent COVID-19 virus. Three categories of models namely, conventional CNN, hybrid, and pre-trained models proposed in this study to classify CXR images into normal, pneumonia or COVID-19 and normal or COVID-19 classes. The suggested CNN and hybrid models achieved test accuracies of 98.91%, 98.44% and 95.60%, respectively. Also, the fine-tuned ResNet50, Xception, and DenseNet201

models achieved test accuracies of 98.90%, 99.14%, and 99.00%, respectively. These outcomes showed the significance of the study and, without a doubt, maked us confident that the models can indeed be deployed in real-life scenarios to aid radiologist with robust tools for early disease detection and diagnosis. For future research, we aim to use alternative resampling methods, such as combinations of under sampling and oversampling or synthetic minority oversampling techniques called SMOTE to address the class imbalance. We also wish to use additional hybrid models, such as merging CNN with ELM (Extreme Learning Machine) or convolution auto-encoders with deep learning classification using the encoded features to classify the images. Research in these domains will allow exploring other boundaries in image classification to create robust models for aiding medical experts in quick diagnoses and treatment of ailments.

This research work was supported by grants given to the MSc. dissertation project by the Scientific Research Projects Administration Unit of Fırat University, Elazığ, Turkey [grant number: TEKF.22.32]. The authors would like to express immense gratitude for this unit's financial support throughout the research study.

Contributions of the authors

The authors contributions to the paper are equal.

Conflict of Interest Statement

There is no conflict of interest between the authors.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

Acknowledgment

References

[1] L. Qun et al., "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *N Engl J Med.,* vol. 382, pp. 1199-1207, March 2020.
[2] M. A. Alah, S. Abdeen, and V. Kehyayan, "The first few cases and fatalities of corona virus disease 2019 (COVID-19) in the Eastern Mediterranean Region of the World Health Organization: a rapid review," *J Infect Public Health,* vol. 13, pp. 1367-1372, October 2020.
[3] J. S. Mackenzie, and D. W. Smith, "COVID-19: a novel zoonotic disease caused by a coronavirus from China: what we know and what we don't," *Microbiol Aust.,* pp. 1-14, March 2020.
[4] A. R. Rahmani et al., "Sampling and detection of corona viruses in air: a mini review," *Sci Total Environ.,* vol. 740, pp. 1-7, October 2020.
[5] J. She, L. Liu, and W. Liu, "COVID-19 epidemic: Disease characteristics in children," *J Med Virol.,* vol. 92, pp. 747-754, April 2020.
[6] L. Baroiu et al., "COVID-19 impact on the liver," *World J Clin Cases.,* vol. 9, pp. 3814-3825, June 2021.
[7] P. Sirohiya et al., "Airway management, procedural data, and in-hospital mortality records of patients undergoing surgery for mucormycosis associated with coronavirus disease (COVID-19)," *J Mycol Med.,* vol. 32, pp. 1-6, November 2020.
[8] Y. Rolland et al., "Coronavirus disease-2019 in older people with cognitive impairment," *Clin Geriatr Med.,* vol. 38, pp. 501-507, August 2022.
[9] M. Sachdeva et al., "Cutaneous manifestations of COVID-19: report of three cases and a review of literature," *J Dermatol Sci.,* vol. 98, pp. 75-81, May 2020.
[10] M. Bansal, "Cardiovascular disease and COVID-19," *Diabetes Metab Syndr.,* vol. 14, pp. 247-250, May 2020.
[11] B. P. Goodman et al., "COVID-19 dysautonomia," *Front Neurol.,* vol. 12, pp. 1-5, April 2021.
[12] L. Falzone et al., "Current and innovative methods for the diagnosis of COVID-19 infection (review)," *Int J Mol Med.,* vol. 47, pp. 1-23, June 2021.
[13] A. Barragan-Montero et al., "Artificial intelligence and machine learning for medical imaging: a technology review," *Phys Med.,* vol. 83, pp. 242-256, March 2021.
[14] M. Aljabri and M. AlGhamdi, "A review on the use of deep learning for medical images segmentation," *Neurocomputing,* vol. 506, pp. 311-335, September 2022.

[15] P. Asha et al. "Artificial intelligence in medical Imaging: An analysis of innovative technique and its future promise," *Mater Today Proc.*, vol. 56, pp. 2236-2239, December 2021.
[16] A. Singhal et al., "Study of deep learning techniques for medical image analysis: a review," *Mater Today Proc.*, vol. 56, pp. 209-214, January 2022.
[17] M. E. Sahin, "Deep learning-based approach for detecting COVID-19 in chest X-rays," *Biomed Signal Process Control.*, vol. 78, pp. 1-10, September 2022.
[18] S. Hassantabar, M. Ahmadi, and A. Sharifi, "Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network approaches," *Chaos Solitons Fractals.*, vol. 140, pp. 1-11, November 2020.
[19] R. Malhotra, H. Patel, and B. D. Fataniya," Prediction of COVID-19 disease with chest X-Rays using convolutional neural network," in *Proc of the 3rd Int. Conf. on Inventive Research in Computing Applications, ICIRCA 2021, Coimbatore, India, September 2-4, 2021.*
[20] A. K. Das et al., "Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network," *Pattern Anal Appl.*, vol. 24, pp. 1111-1124, March 2021.
[21] A. Banerjee et al., "COVID-19 chest X-ray detection through blending ensemble of CNN snapshots," *Biomed Signal Process Control.*, vol. 78, pp. 1-9, September 2022.
[22] A. M. Ismael and A. Şengür, "Deep learning approaches for COVID-19 detection based on chest X-ray images," *Expert Syst Appl.*, vol. 164, pp. 1-11, February 2021.
[23] P. Patel, "Chest X-ray (Covid-19 & Pneumonia)," kaggle.com, 2020. [Online]. Available: https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia. [Accessed: Oct. 4, 2023].
[24] T. Rahman, M. Chowdhury, and A. Khandakar, "COVID-19 Radiography Database," Kaggle.com, 2021. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database. [Accessed: Oct. 4, 2023].
[25] R. Mohammed, J Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: overview study and experimental results," in *Proc of the 11th Int. Conf. on Information and Communication Systems, ICICS 2020, Irbid, Jordan, April 7-9, 2020,* pp. 243-248.
[26] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J Big Data*, vol. 6, pp. 1-48, July 2019.
[27] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc of the 3rd Int. Conf. on Learning Representations, ICLR 2015, San Diego, USA, May 7-9, 2015.*
[28] R. Mitchell and E. Frank, Accelerating the XGBoost algorithm using GPU computing," *PeerJ Comput Sci.*, vol. 3, pp. 1-37, July 2017.
[29] Y. Shihong, L. Ping, and H. Peiyi, "SVM classification:Its contents and challenges," *Appl Math J Chin Univ.*, vol. 18, pp. 332-342, September 2003.
[30] K. He et al., "Deep residual learning for image recognition," in *Proc of the Conf. on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, USA, June 27-30, 2016,* pp. 770-778.
[31] F. Chollet, "Xception: deep learning with depthwise separable convolutions", in *Proc of the Conf. on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, USA, July 21-26, 2017,* pp. 1800-1807.
[32] G. Huang et al., "Densely connected convolutional networks," in *Proc of the Conf. on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, USA, July 21-26, 2017,* pp. 2261-2269.
[33] T. Y. Chen, FC. Kuo, and R. Merkel, "On the statistical properties of the f-measure," in *Proc of the 4th Int. Conf. on Quality Software, QSIC 2004, Washington, USA, September 8-10, 2004,* pp. 146-153.
[34] G. Caseneuve et al., "Chest X-ray image preprocessing for disease classification," *Procedia Comput Sci.*, vol. 192, pp. 658-665, October 2021.
[35] G. Hussain and Y. Shiren, "Recognition of COVID-19 disease utilizing -ray imaging of the chest using CNN," in *Proc of the Int. Conf. on Computing, Electronics & Communications Engineering, iCCECE 2021, Southend, UK, August 16-17, 2021,* pp. 71-76.
[36] A. S. Musallam, A. S. Sherif, and M. K. Hussein, "Efficient framework for detecting COVID-19 and pneumonia from chest X-ray using deep convolutional network," *Egypt Inf J.*, vol. 23, pp. 247-257, July 2022.

[37] S. Singht et al., "CNN based Covid-aid: Covid 19 detection using chest X-ray," in *Proc of the 5th Int. Conf. on Computing Methodologies and Communication, ICCMC 2021, Erode, India, April 8-10, 2021,* pp. 1791-1797.

[38] G. Gilanie et al., "Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks," *Biomed Signal Process Control,* vol. 66, pp. 1-6, April 2021.

[39] T. Mahmud, A. Rahman, and S. A. Fattah, "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.,* vol. 122, pp. 1-10, July 2020.

[40] S. Dilshad et al., "Automated image classification of chest X-rays of COVID-19 using deep transfer learning," *Results Phys.,* vol. 28, pp. 1-10, September 2021.

[41] P. A. Vieria et al., "Classification of COVID-19 in X-ray images with genetic fine-tuning," *Comput Electr Eng.,* vol. 96, pp. 1-8, December 2021.

[42] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "4S-DT: self-supervised super sample decomposition for transfer learning with application to COVID-19 detection," *IEEE Trans Neural Netw Learn Syst.,* vol. 32, pp. 2798-2808, July 2021.

[43] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," *Comput Biol Med.,* vol. 132, pp. 1-13, May 2021.

[44] M. Toğaçar, "Disease type detection in lung and colon cancer images using the complement approach of inefficient sets," *Comput Biol Med.,* vol. 137, pp. 1-13, October 2021.

[45] D. K. Sharma et al., "Classification of COVID-19 by using supervised optimized machine learning technique," *Mater Today Proc.,* vol. 56, pp. 2058-2062, November 2021.

[46] S. S. Verma, A. Prasad, and A. Kumar, "CovXmlc: high performance COVID-19 detection on X-ray images using multi-model classification," *Biomed Signal Process Control,* vol. 71, pp. 1-7, January 2022.

[47] B. Prabha et al., "Intelligent predictions of Covid disease based on lung CT images using machine learning strategy," *Mayer Today Proc.,* vol. 80, pp. 3744-3750, July 2021.

[48] H. Nasiri and S. Hasani, "Automated detection of COVID-19 cases from chest X-ray images using deep neural network and XGBoost," *Radiography,* vol. 28, pp. 732-738, August 2022.

[49] A. Arivoli, D. Golwala, and R. Reddy, "CoviExpert: COVID-19 detection from chest X-ray using CNN," *Measur Sens.,* vol. 23, pp. 1-8, October 2022.

[50] X. Li et al., "A self-supervised feature-standardization-block for cross-domain lung disease classification," *Methods.,* vol. 202, pp. 70-77, June 2022.

[51] S. Kumar et al., "Chest X ray and cough sample based deep learning framework for accurate diagnosis of COVID-19," *Comput Electr Eng.,* vol. 103, pp. 1-19, October 2022.