Ensemble deep models for covid-19 pandemic classification using chest x-ray images via different fusion techniques

Lamiaa Menshawy [a,1,*], Ahmad H. Eid [b,2], Rehab F. Abdel-Kader [b,3]

[a] Department of Technology and Information systems, Port Said University, Port Said, Egypt
[b] Department of Electrical Engineering, Port Said University, Port Said, Egypt
[1] lamyaa_menshawy@himc.psu.edu.eg ; [2] ahmad.eid@eng.psu.edu.eg; [3] rehabfarouk@eng.psu.edu.eg
* corresponding author

ARTICLE INFO

Article history
Received September 19, 2022
Revised January 19, 2023
Accepted January 30, 2023
Available online March 31, 2023

Keywords
Coronavirus (covid-19)
Deep learning
Machine learning
Decision fusion
Feature fusion

ABSTRACT

A pandemic epidemic called the coronavirus (COVID-19) has already afflicted people all across the world. Radiologists can visually detect coronavirus infection using a chest X-ray. This study examines two methods for categorizing COVID-19 patients based on chest x-rays: pure deep learning and traditional machine learning. In the first model, three deep learning classifiers' decisions are combined using two distinct decision fusion strategies (majority voting and Bayes optimal). To enhance classification performance, the second model merges the ideas of decision and feature fusion. Using the fusion procedure, feature vectors from deep learning models generate a feature set. The classification metrics of conventional machine learning classifiers were then optimized using a voting classifier. The first proposed model performs better than the second model when it concerns diagnosing binary and multiclass classification. The first model obtains an AUC of 0.998 for multi-class classification and 0.9755 for binary classification. The second model obtains a binary classification AUC of 0.9563 and a multiclass classification AUC of 0.968. The suggested models perform better than both the standard learners and state-of-the-art and state-of-the-art methods.

1. Introduction

Daily life, public health, and the global economy are all being impacted by COVID-19. To stop the pandemic from spreading further and to treat the affected individuals as soon as feasible, it is essential to identify positive cases as soon as possible. Diagnostic tools are now more necessary than ever because there aren't any reliable automated toolkits [1]. The favourite option for diagnosing COVID-19 is currently real-time reverse transcription-polymerase chain reaction (RT-PCR) testing, with radiographic imaging of the chest, including computed tomography (CT) and X-rays, playing a significant role in the early diagnosis and management of this disease. Even with a negative result, symptoms can be found by looking at a patient's radiographs because RT-PCR has a low sensitivity of 60%–70% [1],[2].

Information fusion is one of the newest methodologies for processing data. The three levels of information fusion are pixel level, feature level, and decision level. Image categorization uses machine learning (ML) and deep learning (DL) approaches to investigate feature- and decision-level fusion techniques for model learning. Since they improve scalability, processing speed, dependability, and even human performance in some healthcare tasks, ML and in particular DL are promising technologies that are being implemented by numerous healthcare providers [3]. To increase the speed and accuracy of

diagnosis, better understand the patterns of COVID-19 virus propagation, and create novel, efficient therapeutic strategies, ML and DL algorithms are being applied [4].

Information fusion in the form of "decision fusion" merges the decision taken by various classifiers into a general conclusion. For the same classification task, multiple classifiers may respond differently. Better performance than would be feasible with either source alone can be achieved by combining the results from other approaches, algorithms, sources, or classifiers [5],[6]. Classifiers might have the same or separate feature sets and can be of the same type [7]. The outputs of base classifiers can be combined using a variety of techniques, including the majority voting method, the weighted majority voting approach by allocating weights to classifier support, the behavior knowledge space (BKS) method, and the Naive-Bayes method [6],[7].

The data input for feature-level data fusion is either already data or features that have been extracted. As a result, we can produce sophisticated traits or features in the form of additional patterns that can be used to achieve various objectives, including higher-level decisions [8]. The two major techniques for feature fusion are concatenating and adding. If the order is a crucial component, the adding approach is not the best option. Deep learning frequently uses concatenation.

To create a composite image from numerous input images that each include complementary details about the same scene, pixel-level image fusion is used. Source images, also known as input images, are taken by various imaging devices or by a single device with various parameter settings. A fused image is a composite image that is meant to be more sensitive to human or machine sensitivity than individual input [9].

Due to its significance in the diagnosis and treatment of many medical issues, medical imaging has attracted attention in recent years. Images from CT scans and X-rays are used in medical imaging technology to identify COVID-19. A summary of the different classification methods developed based on radiographic images is shown in Table 1. According to [10], [11], [12], [13], and [14], CT is now a quick way to find COVID-19 patients. [10] The ResNet-101 may be a promising model for diagnosing COVID-19 infections, according to this study, which used 10 convolutional neural networks (CNN) to separate COVID-19 infection from non-COVID-19 patients. [11] Produced two unique datasets using 150 CT scans. To obtain deep features, a pre-trained CNN model was used. A support vector machine was trained using these deep features that were combined and arranged (SVM). Chest CT images are used in [12] as part of a multi-center sparse learning and decision fusion approach for automating COVID-19 diagnosis. To enhance diagnostic performance, decision fusion is utilized to thoroughly examine the connections between multi-center data and compare the diagnostic outcomes of various classifiers. In [13], a COVID-19 diagnosis system was created using three deep learning models (CNN, a stacked autoencoder, and a deep neural network). According to test results, CNN outperforms Deep Neural Networks and Stacked Autoencoder in terms of performance. A total of 15 pre-trained CNN architectures were implemented and further improved upon in [14]. A majority voting-based ensemble approach was subsequently developed.

The use of CT to find COVID-19 has many problems. The equipment required to take pictures is expensive and not present in many institutions. The X-ray approach, in contrast, is quicker, simpler, less expensive, and less dangerous than CT. As a result, the issue of categorizing COVID-19 using X-ray pictures is one that many researchers are dealing with.

With accuracy rates of 98.08% and 87.02% for binary and multi-class tasks, respectively, the authors of [1] proposed a deep neural network-based approach. The study's use of only a few COVID-19 X-rays has some limitations. For automatically predicting COVID-19 patients, the authors of [15] proposed a deep transfer learning-based method. The pre-trained ResNet50 model surpasses the five models in terms of performance using three different datasets. The use of only a few COVID-19 X-rays is another drawback of this investigation. By fine-tuning four pre-trained convolutional models (ResNet18, ResNet50, Squeeze Net, and Dense Net–121) on the training data, a deep learning framework for COVID-19 detection was created [16]. A dataset of about 5000 photos was prepared. These models had

an average specificity rate of about 90% and were able to conduct binary classification with a sensitivity rate of 98%.

Table 1. The literature overview.

Reference	Type of Images	Dataset samples	Accuracy
[10]	CT	1020 CT scans from a set of 86 patients without COVID-19 and 108 patients with COVID-19.	AUC of 0.994
[11]	CT	150 CT scans were used to create two distinct datasets. Each dataset has 3000 COVID-19-tagged images with No findings.	Recall: 98.93%
[12]	CT	2298 normal patients and 1034 COVID-19 CT scans.	Mean Recall: 97.40%
[13]	CT	the dataset includes 349 CT scans of COVID-19 infection and 397 of uninfected individuals.	Recall:87.65%
[14]	CT	397 negative and 349 positive COVID-19 patients using CT scans	Recall: 85.4%
[1]	X-ray	500 cases of pneumonia, 500 cases of no findings, and 127 cases whose X-rays were positive.	Recall: 95.13 % of Binary classes Recall:85.35%on Multi-classes
[15]	X-ray	50 patients with COVID-19 and 50 of them had normal chest X-rays.	Accuracy: 1st Dataset: 96.1%, 2nd Dataset: 99.5% 3rd Dataset-3: 99.7%
[16]	X-ray	Set of 5,000 chest X-ray images for COVID-19 detection.	Sensitivity: 98% Specificity: 90%.
[17]	X-ray	There are 403 COVID-19 and 721 normal images.	Recall:69% using CNN alone Recall:90% using CovidGAN.
[18]	X-ray	15,000 images total, 5000 normal, 5000 pneumonia, and 5000 COVID-19.	Recall:99.64% on Binary classes Recall:95.41% on Multi-classes
[19]	X-ray	412 X-ray images (Normal: 206, Covid: 206)	Recall:93%
[20]	X-ray	4273 pneumonia cases, 1583 normal X-rays, and 576 COVID-19	Recall:100%
[21]	X-ray	3615 COVID-19 cases, 6012 cases of lung opacity, 5870 cases of pneumonia, 20,000 cases of lung cancer, 1400 cases of tuberculosis, and 10,192 cases of normal images.	Recall: 93.75 %
[22]	X-ray	1341 1Normal, 219 COVID-19, and 345 viral pneumonitis	Recall: 99.83%
[23]	X-ray	There are a total of 388 images in each of the two datasets, with 194 images in each class.	Dataset 1: F1-score 98.461%. Dataset 2: F1-score 95.633%.
[24]	X-ray	576 COVID-19, 4273 images of pneumonia, and 1583 images of normal images	Recall: 93 % with augmentation Recall: 94 % without augmentation
[25]	X-ray	1143 COVID-19, 1341 in Normal, and 1345 in classes for pneumonia.	Recall:98.70%
[26]	X-ray	10,192 healthy individuals, 3616 COVID-19 X-ray images, and 1345 images with pneumonia.	Recall:98.77%
[27]	X-ray	1143 COVID-19 positive, 1341 neutral images, and 1345 images of pneumonia	Recall:98.8%
[28]	X-ray	5,841 Other lung infections, 6,722 Covid-19, and 6,719 Normal.	Accuracy: 96.16%

In [17], the authors proposed a method for creating fake chest X-ray pictures by using a model called CovidGAN that is built on an Auxiliary Classifier Generative Adversarial Network. Only using CNN for

classification, the accuracy was 85%. The accuracy rises to 95% with the addition of synthetic images. There are still several shortcomings in this analysis. They can first enhance GAN architecture more. Second, they only employed a tiny dataset, and the synthetic samples created for this study may have been of higher quality. [18] Create a big and diversified dataset using a variety of image processing approaches to enhance COVID-19 X-ray images. This will help deep learning algorithms perform better when detecting the virus from chest X-rays. Based on DenseNet201, VGG16, and VGG19, they also suggest novel and reliable deep-learning models identify COVID-19 from a variety of X-ray pictures. According to a performance evaluation, the proposed models were 95.48% accurate for multi-class classification and 99.62% accurate for binary classification. [19] Compares several pre-trained models, including Inception ResNet V2, ResNet50, and VGGNet-19. Inception ResNet V2 surpasses VGG Net and ResNet models, with training and test accuracy of 99.26 and 94, respectively, according to experimental data. In [20] a model with the Support Vector Machine is built as the end layers and the pre-trained models VGG16, VGG19, InceptionV3, MobileNetV2, and Xception. The final layer of the proposed model was modified using SVM after the weights from the pre-trained model were utilized as initial values and modified during training. VGG-19 followed by the three CNN blocks employed as feature extraction at the classification step in [21], the suggested VGG19 + CNN outperformed with a recall of 93.75%, and it performed better.

AlexNet is used in [22] to extract features, and SVM is subsequently applied for classification. In [23], features are extracted using CNNs, and then they are classified using a variety of machine-learning techniques. The usage of only a few X-ray images is the study's main drawback. The model has gone through two different scenarios in [24]. The model was tested in the first scenario with image augmentation, and in the second scenario without augmentation of the data. To forecast the output class, the authors of this study suggested a generalized extreme value activation function. This study demonstrates that when one class predominates over another, this activation function performs better than the sigmoid activation function. An artificial bee colony technique was employed in [25] to improve the contrast of preprocessed chest X-rays. Deep network models have already been trained to extract features from X-ray images. Finally, the COVID-19 X-ray image was categorized using long-short-term memory. In [26], the two pre-trained CNN models, AlexNet and Xception, are employed to merge features retrieved from input X-rays utilizing the deep feature concatenation mechanism. To create an ensemble framework with three classifiers—GoogLeNet, ResNet-18, and DenseNet-121—a voting ensemble strategy is suggested in [27]. SVM trained on top of deep networks is used in [28] to identify Covid-19 in chest X-ray pictures. The maximum accuracy is 96.16%, which is attained by the suggested SVM on top of the deep network.

The primary contributions of this study are outlined as follows:

- Two models are suggested in this paper to contrast deep learning and conventional machine learning. They can be applied to multiclass and binary classification.
- The first model is the Deep Learning Decision Fusion Model (DLDF). It is based on three deep learning classifiers chosen according to the best performance metrics. The classifiers determine the category to which an image belongs. Results from the three classifiers then were combined to produce an estimate of better quality than could be obtained from either source alone. We use majority voting and Bayes optimal as two different decision fusion techniques to combine the result from the different classifiers.
- The second model is the Machine Learning Feature and Decision Fusion Model (MLFDF). It is based on VGG-16, DenseNet201, and ResNet50V2 models as feature extractors. Feature vectors obtained from these models use a fusion process to provide feature sets. After that, a voting classifier was used to optimize the classification metrics of three classifiers (SVM, k-nearest neighbors Classifier, and Logistic Regression classifier). Instead of selecting just one model or classifier, a voting classifier can estimate the mean of all of the classifier's outcomes.

- A comparison of the two models is done. Better results are achieved by the first model than that by the second model. The experimental findings demonstrate how well the suggested models perform in both binary and multiclass classification. In terms of the overall evaluation measures, the two models perform better than both base learners and previous techniques.

The rest of this paper is divided into the following sections. In section 2, the proposed models are presented with a thorough description. In section 3, we evaluate and discuss the classification performance analysis of the suggested models, and in section 4, we conclude our work.

2. Method

In this paper, two models are proposed to compare DL and traditional ML as shown in Fig.1. They can be applied to multiclass and binary classification. The first model is Deep Learning Decision Fusion Model (DLDF) and the second model is Machine Learning Feature and Decision Fusion Model (MLFDF). The MLFDF uses traditional machine learning classifiers like SVM, LR, and KNN. The other model uses pure deep learning and two different decision fusion techniques. One of these uses majority voting as a Decision fusion technique and the model is called (DLDF1). The other Decision technique is the Bayes optimal and the model is called (DLDF2).

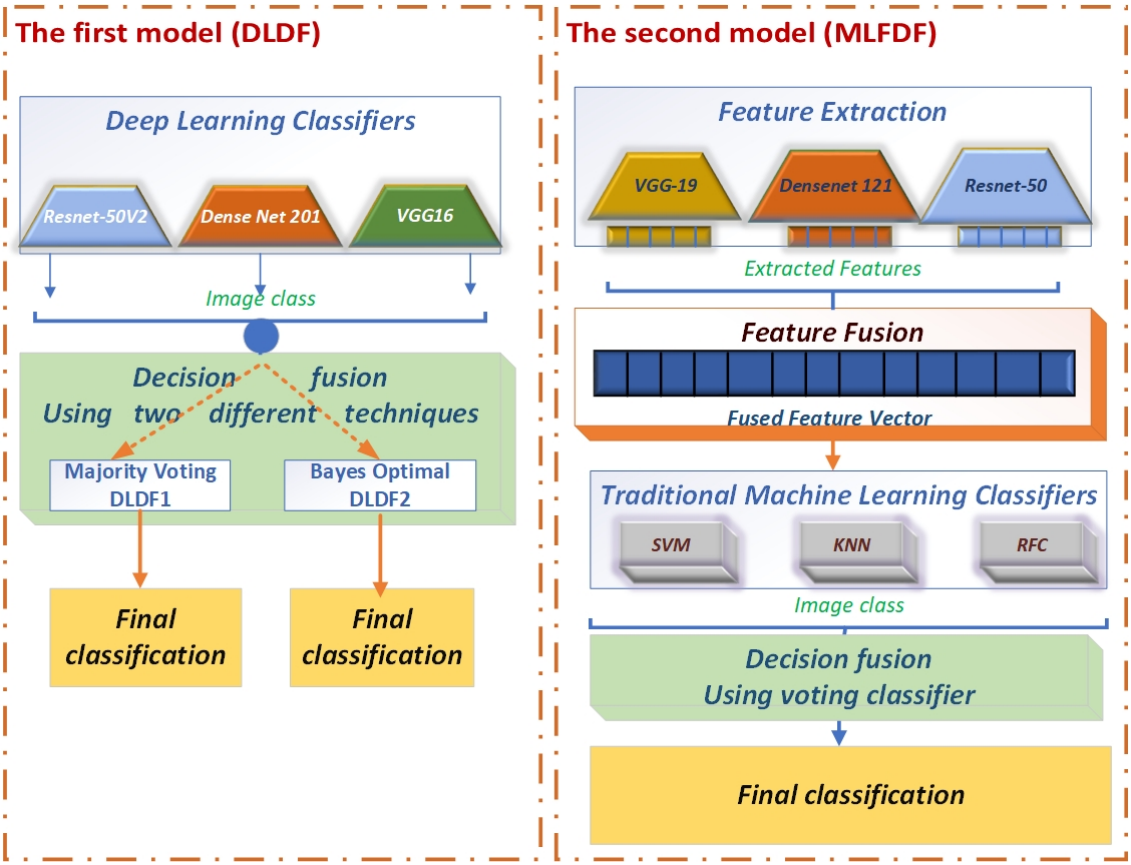


Fig. 1. Detection proposed models.

2.1. The proposed Deep Learning Decision Fusion (DLDF) Model

The data augmentation phase which generates new images from the original data set is the first of the DLDF model's four primary stages. To enhance the number of samples, we apply data augmentation techniques including flipping, slight rotation, and tiny distortions. The top 3 models are chosen based on the value of each deep learning classifier performance parameter in the second phase, which is the training and validation phase using deep CNN models. The classification step, the third phase, involves testing the pre-trained models and calculating the total test accuracy using the pre-trained CNN model's performance indicators. Finally, as shown in Fig. 2, the category to which the image belongs is chosen.

The decision fusion step, which comes last, enhances the classification task's performance. The process of integrating many classifiers' decisions into one final decision is known as decision fusion. It seeks to integrate the views of various classifiers to provide a consensus decision that is superior to the classifiers' opinions [29], [30]. This model does decision fusion utilizing two distinct decision fusion methods. The Bayes optimum method and majority voting are two of these methods. The proposed classification model can be applied to multiclass or binary classification.

2.2. The proposed Machine Learning Feature and Decision Fusion (MLFDF) Model

The first step of the proposed model, called Data Augmentation, is responsible for creating a new image from the original data set. The suggested model has four primary phases. The feature extraction phase, based on deep CNN models, is the second stage. As feature extractors, VGG-16, DenseNet201, and ResNet50V2 models are employed. The third phase is the feature fusion phase, where the previously extracted features are combined into a single vector known as the fused feature vector. Assume that the features from DenseNet201, ResNet50V2, and VGG-16 are each an m-dimensional vector, an n-dimensional vector, and a v-dimensional vector, respectively. Concatenating the retrieved vectors creates an (m + n + v) dimensional vector, which is then used to combine the features. In this method, the negative impacts of using subpar features from a single CNN network are reduced. The Support Vector Machine Classifier, the K-Nearest Neighbor Classifier, and the Logistic Regression Classifier are combined in the fourth phase, called Decision Fusion, to achieve the best classification scores. This model is illustrated in Fig 3. Finally, the model can be applied to multiclass or binary classification.

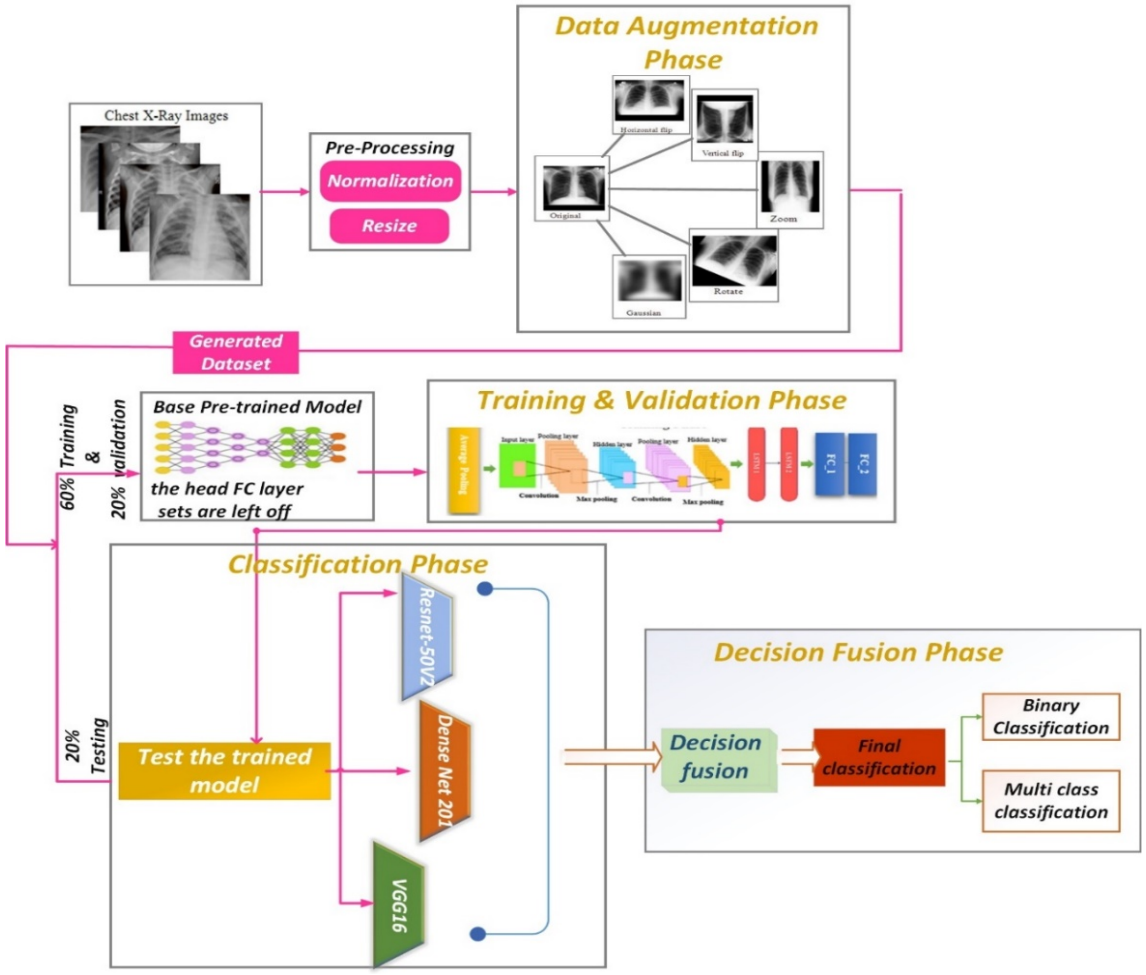


Fig. 2. Deep Learning Decision Fusion (DLDF) Model.

3. Results and Discussion

3.1. Data Description

Chest X-ray radiographs have been utilized in this investigation to predict COVID-19 patients. On a dataset comprised of COVID-19 X-ray images and normal, pneumonia, and other X-ray images, we normally evaluate the models. The imbalanced dataset and the lack of COVID-19 images in earlier research like [1], [15], [17], [19], [20], [22], [24], [25], and [27] are also key issues that have been resolved by this dataset. The COVID-ChestXray-15k dataset was compiled from 11 sources and contained 4420 COVID-19 pictures before data augmentation, 5000 images of normal chest X-rays, and 5000 images with pneumonia. After augmentation, there were 15,000 more COVID-19 images, up from 4420 COVID-19 samples before. This dataset was obtained from the open-source GitHub repository shared by Abeer Badawi et al. [18].

60% of the dataset's X-ray images, comprising both healthy and unhealthy instances, are selected for training, 20% are used for validation, and 20% are used for testing deep learning classifiers. All of the Deep CNNs architectures in this study's training parameters are: The batch size is 32, and the learning rate is set to 0.001. Due to its high convergence and short training times, stochastic gradient descent is used to train all deep network classifiers. All experimental simulations were carried out on TPUs with 35.5 GB of RAM using Google Colab Pro.

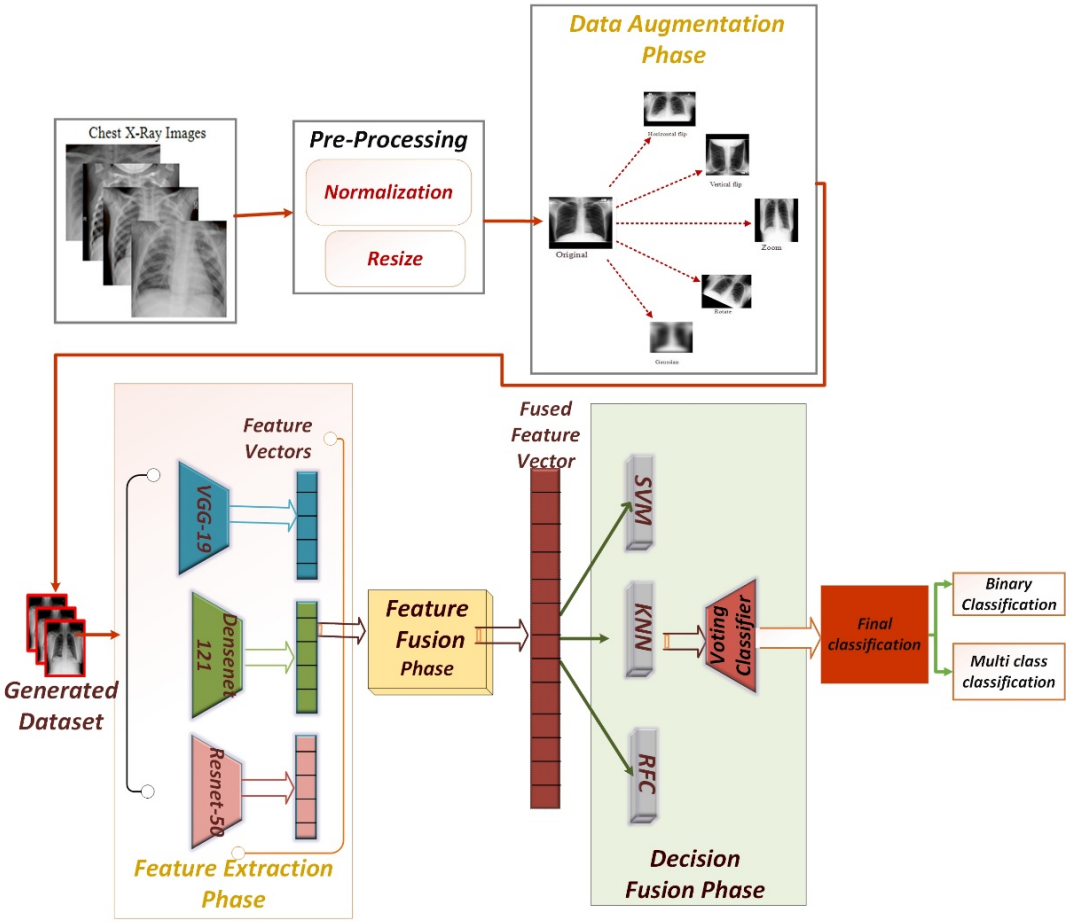


Fig. 3. Machine Learning Feature and Decision Fusion (MLFDF).

3.2. Performance of Binary Classification

For all models, a training phase lasting up to 50 epochs was implemented. In Fig. 4, the model's confusion matrix and typical test outcomes are displayed. First, 952 of the COVID-19 were categorized as True Positive by the pre-trained VGG19 model, whereas 939 of the normal were labelled as True Negative. Second, 830 instances of normal were labelledd as True Negatives by the ResNet50 model,

while 969 instances of COVID-19 were categorized as True Positives. Finally, 968 of the COVID-19 were categorized by DenseNet121 as True Positives, while 930 of the normal were classified as True Negatives. For each model, receiver operating characteristic curve (ROC) plots and regions are shown in addition to the confusion matrix.

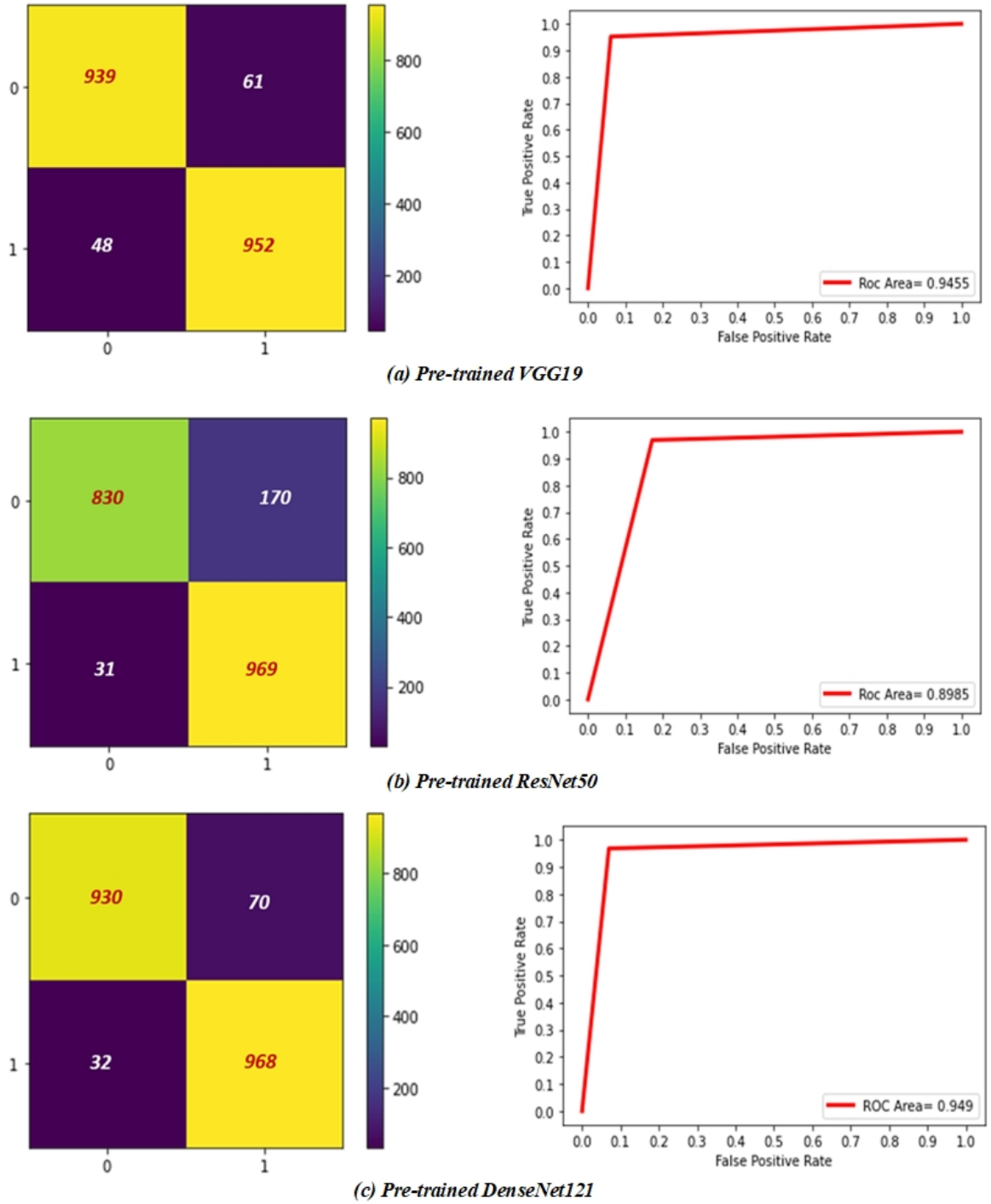


Fig. 4. The ROC curves and confusion matrices that obtained using pre-trained models.

The confusion matrices of COVID-19 and the outcomes of the three proposed models' normal tests are shown in Fig. 5. Using conventional machine learning, the first model is MLFDF. The other two models use pure deep learning with two distinct methods of decision fusion. One of these two models, known as the majority voting model, employs decision fusion as a strategy (DLDF1). Bayes optimum is a different decision-making technique, and the model is known as (DLDF2).

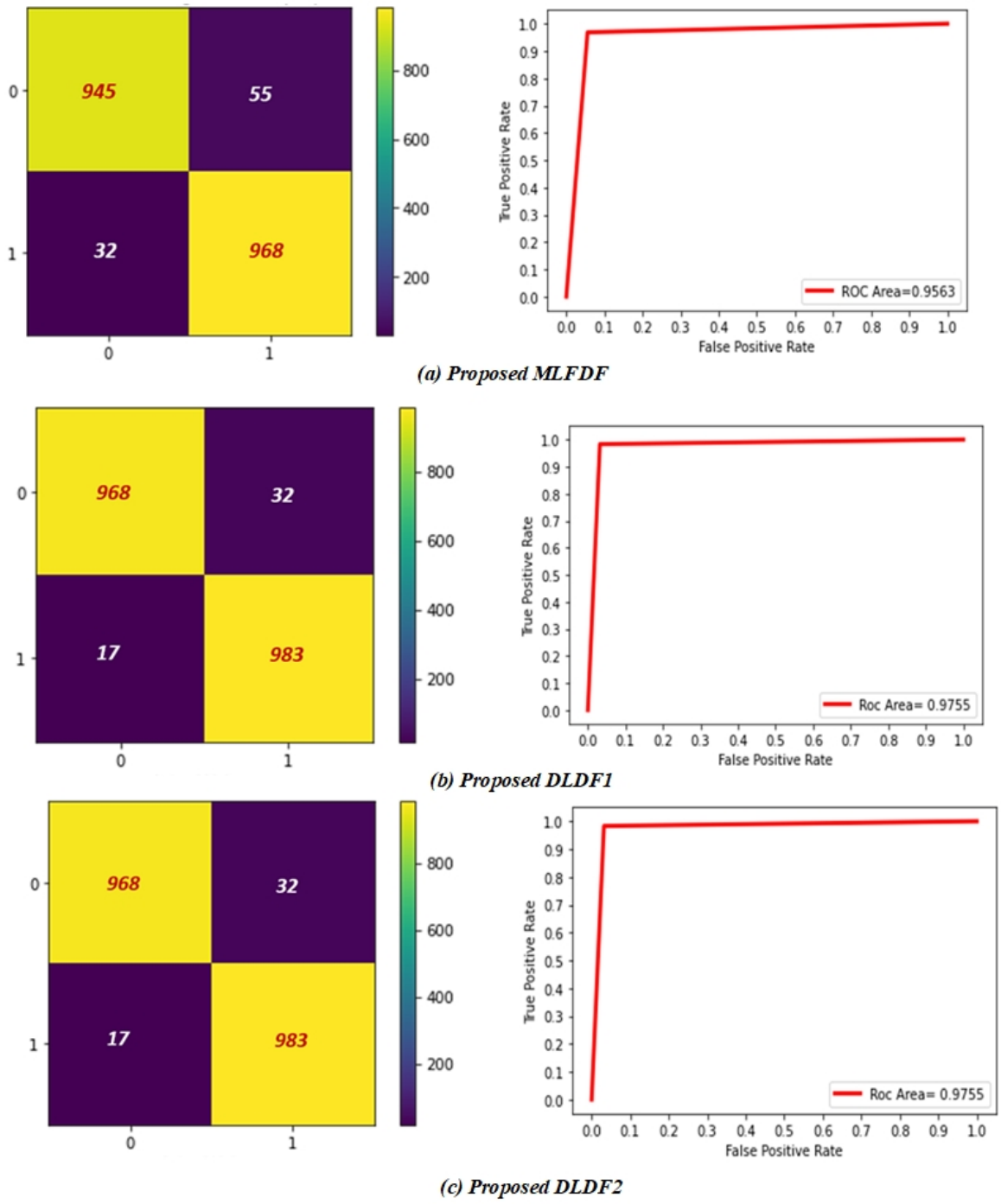


Fig. 5. The ROC curves and confusion matrices that obtained using the proposed models.

First, the MLFDF model identified 945 COVID-19 positives as True Positives and 968 normal individuals as True Negatives. Second, 983 of the COVID-19 were categorized as True Positive and 968 as True Negative by the DLDF1 model, respectively. Finally, identical to the DLDF1 model, the DLDF2 model identified 968 of the normal as True Negative and 983 patients with COVID-19 as True Positive. For each model, ROC curve plots and regions are included in addition to the confusion matrix. Comparing the DLDF1 and DLDF2 models to the trained models, they appear to be quite high.

Pre-trained models' area under the ROC curve (AUC) scores range from 0.89 to 0.949, with 0.5 denoting a subpar classifier and 1 denoting a superior one. With AUCs of 0.956 and 0.975, the three suggested models are excellent classifiers. Models DLDF1 and DLDF2 seem to be very high.

Table 2 displays the COVID-19 and normal binary classification findings. Table 2 displays thorough performance comparisons of all models utilizing the test data. For each model's accuracy, precision, recall, specificity, and F1 score have been provided. In Table 1, the terms Accuracy (Acc), Recall (Rec), Specificity (Spe), Precision (Pre), F1-Score (F1), and Zero One Loss (ZOL) are abbreviated (zol).

Table 2 examines the efficiency of MLFDF, DLDF1, and DLDF2 models compared to other methods. The Pre-trained VGG 19 has the lowest performance with a minimum Recall of 95.2%. Furthermore, the ResNet50 and DenseNet121 models showed slightly improved results, with a nearer sensitivity of 96.4% and 96.8% respectively. Finally, the DLDF1 and DLDF2 models achieved exactly equivalent classifier results with a sensitivity of 98.3%.

Table 2. Performance analysis of the various models for Binary classification.

Models	Acc	Rec	Spe	Pre	F1	zol
VGG19	94.55	95.20	93.90	93.98	94.59	0.054
ResNet50	89.95	96.40	85.07	83	89.20	0.101
DensNet121	94.90	96.80	93.00	93.26	95	0.051
Proposed MLFDF	**95.65**	**96.84**	**94.43**	**94.69**	**95.75**	**0.043**
Proposed DLDF1	**97.55**	**98.30**	**96.80**	**96.85**	**97.57**	**0.024**
Proposed DLDF2	**97.55**	**98.30**	**96.80**	**96.85**	**97.57**	**0.024**
Enhanced VGG16 [18]	99.62	99.64	99.67	99.57	99.60	-
Enhanced VGG19 [18]	99.00	98.94	98.66	98.94	98.94	-
Enhanced DenseNet201 [18]	91.75	89.34	78.00	94.24	91.72	-

The best performance was attained by the DLDF1 and DLDF2 models, which had accuracy, recall, and specificity values of 97.5%, 98.3%, and 96.8% respectively. In addition, the MLFDF model outperformed pre-trained models concerning the accuracy, recall, and specificity, achieving 95.6%, 96.8%, and 94.3%, respectively. When comparing the proposed models to the enhanced models in [18], the proposed models perform better than Enhanced DenseNet201, but Enhanced VGG19 and VGG16 perform better than the proposed models. This is because the proposed models were trained and tested with images that were 128128 in size, whereas the images in [18] were trained and tested with images that were 224224 in size. The best image resolution selection could therefore enhance neural network performance for a variety of radiology-based machine learning and deep learning activities. Our suggested models employ an image size of 128128 due to limited resources, however, they run faster.

Zero-one loss (ZOL) or misclassification rate is a convenient measure of performance in classification tasks and is the most commonly used. It simply assigns a loss of 1 to the failure to guess the correct class. As shown in Table 1 DLDF1 and DLDF2 models have the lowest ZOL value.

3.3. Performance of Multi-Class Classification

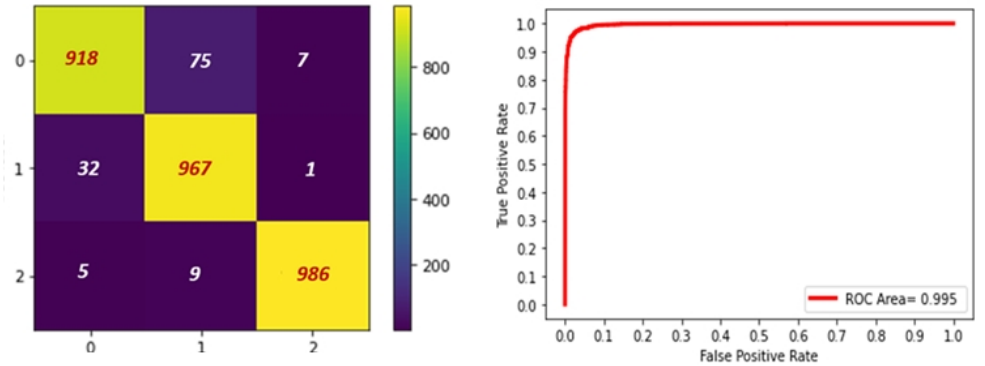
50 epochs of the training phase were run. Using multi-class classification with class 0 for normal, class 1 for COVID-19, and class 2 for pneumonia, we train the models using training and validation data. Table 3 displays the results of the multi-class classification after we evaluated the models on the test set.

The DLDF2 model performed the best, with accuracy, recall, and precision values totaling 97.5%, 97.5%, and 97.49%, respectively. The DLDF1 proposed model also performed well, with accuracy, recall, and precision values totaling 97.4%, 97.39%, and 94.5 percent, respectively. Pre-trained ResNet50 produced the lowest performance with an accuracy of 86%, recall of 86.2%, and precision value of 87.4%; Enhanced DenseNet201 produced an accuracy of 91.97%, recall of 88.3%, and precision value of 94.07%. It is evident from a comparison of the DLDF1 and DLDF2 proposed models with the improved models in [18] that our models outperform the improved DenseNet201, VGG19, and VGG16. The proposed models used an image size of 128×128, increasing model speed.

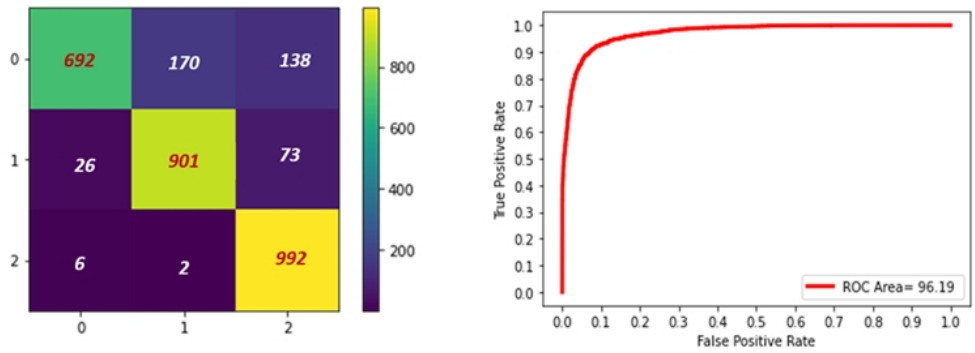
Table 3. Multi-class classification Performance for the various models.

Models	Weighted				
	Acc	*Rec*	*Pre*	*F1*	*ZOL*
VGG19	95.7	95.7	95.77	95.70	0.043
ResNet50	86	86.2	87.4	85.8	0.137
DensNet121	96.4	96.4	96.47	96.4	0.035
Proposed MLFDF	94.5	95.6	96.12	95.8	0.065
Proposed DLDF1	97.4	97.4	97.39	97.39	0.026
Proposed DLDF2	97.5	97.5	97.49	97.49	0.025
Enhanced VGG16 [18]	95.48	95.41	95.48	95.41	-
Enhanced VGG19[18]	95.03	94.95	95.01	94.96	-
Enhanced DenseNet201[18]	91.97	88.30	94.07	89.44	-

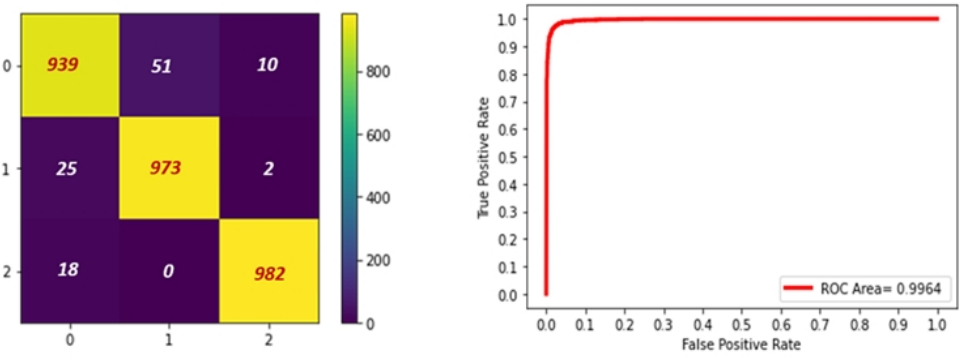
The pre-trained model's ROC curve plot and confusion matrix are displayed in Fig. 6. Fig. 7 displays the confusion matrix for COVID-19 and the outcomes of the normal test for the suggested model. The AUCs for the suggested model are 0.968 and 0.998, indicating that they are excellent classifiers.



(a) Pre-trained VGG19



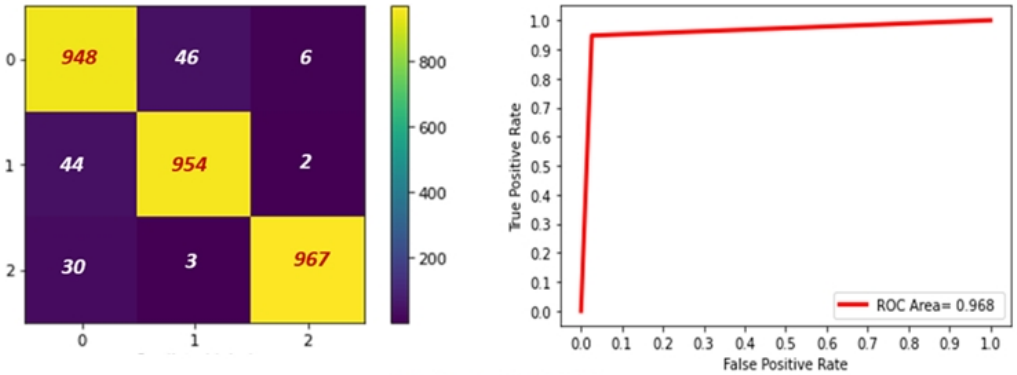
(b) Pre-trained ResNet50



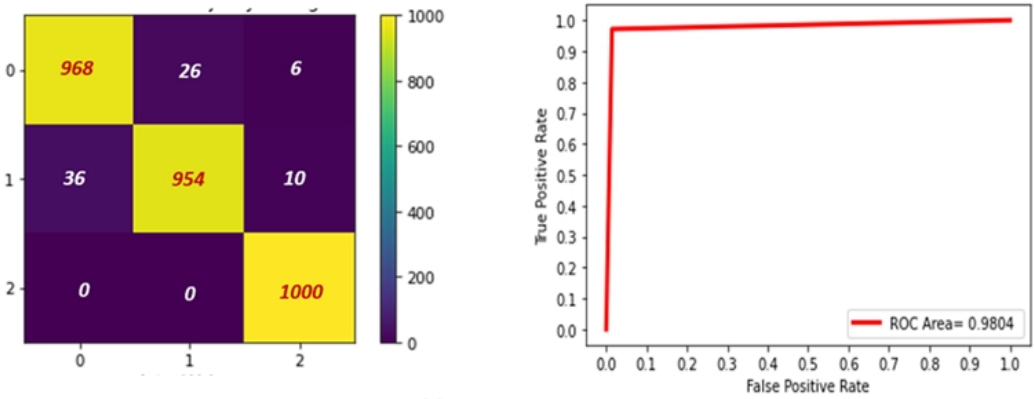
(c) Pre-trained DenseNet121

Fig. 6..ROC curves and the confusion matrices obtained using pre-trained models.

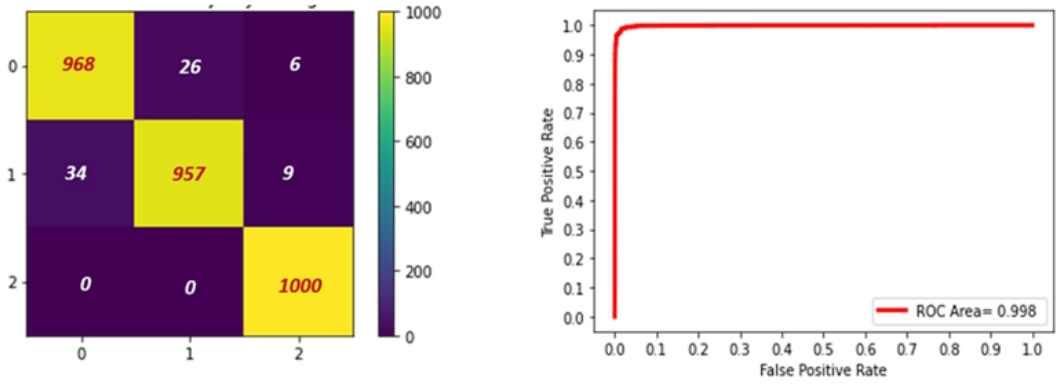
As shown in Table 3, DLDF1 and DLDF2 models have the lowest ZOL value. And that improves that we have obtained the best performance for the proposed models. A detailed comparative study of the proposed models against other methods is presented in Table 3. ResNet50 Pre-trained approach showed the lowest performance with a minimum sensitivity of 86.2%. Furthermore, the VGG 19 and DenseNet121 models showed slightly improved results, with sensitivities approaching 95.7% and 96.4%, respectively. Finally, DLDF1 and DLDF2 models have obtained results with a sensitivity of 97.5% and 97.4%. The proposed models show optimal classification results.



(a) Proposed MLFDF



(b) Proposed DLDF1



(c) Proposed DLDF2

Fig. 7. ROC curves and the confusion matrices obtained using the proposed models.

The DLDF1 and DLDF2 perform better when compared to the MLFDF model, as can be seen from the preceding findings. This is so that DLDF1 and DLDF2 can classify objects into binary and multiple classes. The MLFDF, in contrast, relies on standard machine learning classifiers like SVM, LR, and KNN. Deep learning and conventional machine learning differ mostly in how well they perform as data volume grows. Deep learning algorithms struggle to process little data. This is because deep learning

algorithms need a lot of data to work effectively. Therefore, it appears that traditional machine learning techniques are not appropriate for handling big data sets. It performs effectively with small amounts of data, and its performance reaches an optimal value and does not increase as the data samples increase.

As shown from the result Ensemble learning has better generalization performance. The ensemble learning model combines the advantages of both base learner models and the final model gives better performance. Combining models in parallel can effectively reduce the training time depending on the hardware availability for running the parallel models. DLDF1 uses majority voting as a decision fusion technique. A majority voting combines the results of the base learners. Instead of averaging the probability scores, the majority voting counts the base learners' votes and predicts the final labels as the label with the majority of votes, this improves performance as shown in Binary and Multiclass Classification Performance. The DLDF2 uses the Bayes optimal method and also yields better results than a single base learner.

4. Conclusion

To detect Covid-19 from chest X-rays, multi-class classification models were presented in this study. We use the COVID-Chest Xray-15k dataset, which contains a total of 15,000 radiographs after data augmentation. Our system can complete tasks requiring binary and multiple-class classification. To assess how each of the suggested models fared on the test set, a thorough experimental analysis was done. The final result, with 97.55% accuracy, 96.85% precision, 98.30% recall, and 97.57% F1 score, is an improvement above the other models. For binary classification, it offers 0.7955 AUC; for multi-class classification, it yields 0.998 AUC and greater classification performance of 97.5% accuracy, 97.49% precision, 97.5% recall, and 97.49% F1-score. Future research could concentrate on expanding the provided data to include other chest radiographs from different illnesses kinds and possibly CT on this crucial area of study. It is possible to include suggested models for imaging modalities. Additionally, by addressing the issue of limited resources with various activation functions, the performance of the suggested model can be further enhanced.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.
Funding statement. None of the authors have received any funding or grants from any institution or funding body for the research.
Conflict of interest. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," *Comput. Biol. Med.*, vol. 121, p. 103792, Jun. 2020, doi: 10.1016/J.COMPBIOMED.2020.103792.

[2] M. E. H. Chowdhury *et al.*, "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, 2020, doi: 10.1109/ACCESS.2020.3010287.

[3] S. Lalmuanawma, J. Hussain, and L. Chhakchhuak, "Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review," *Chaos, Solitons & Fractals*, vol. 139, p. 110059, Oct. 2020, doi: 10.1016/J.CHAOS.2020.110059.

[4] A. Alimadadi, S. Aryal, I. Manandhar, P. B. Munroe, B. Joe, and X. Cheng, "Artificial intelligence and machine learning to fight covid-19," *Physiol. Genomics*, vol. 52, no. 4, pp. 200–202, Apr. 2020, doi: 10.1152/PHYSIOLGENOMICS.00029.2020/ASSET/IMAGES/LARGE/ZH70042044120001.

[5] Q. Sen Sun, S. G. Zeng, Y. Liu, P. A. Heng, and D. S. Xia, "A new method of feature fusion and its application in image recognition," *Pattern Recognit.*, vol. 38, no. 12, pp. 2437–2448, Dec. 2005, doi: 10.1016/J.PATCOG.2004.12.013.

[6] M. A. Ganaie, M. Hu, A. K. Malik, M. Tanveer, and P. N. Suganthan, "Ensemble deep learning: A review," *Eng. Appl. Artif. Intell.*, vol. 115, p. 105151, Oct. 2022, doi: 10.1016/J.ENGAPPAI.2022.105151.

[7] U. G. Mangai, S. Samanta, S. Das, and P. R. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Tech. Rev. (Institution Electron. Telecommun. Eng. India)*, vol. 27, no. 4, pp. 293–307, Jul. 2010, doi: 10.4103/0256-4602.64604.

[8] T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," *Inf. Fusion*, vol. 57, pp. 115–129, May 2020, doi: 10.1016/J.INFFUS.2019.12.001.

[9] Y. Liu, X. Chen, Z. Wang, Z. J. Wang, R. K. Ward, and X. Wang, "Deep learning for pixel-level image fusion: Recent advances and future prospects," *Inf. Fusion*, vol. 42, pp. 158–173, Jul. 2018, doi: 10.1016/J.INFFUS.2017.10.007.

[10] A. A. Ardakani, A. R. Kanafi, U. R. Acharya, N. Khadem, and A. Mohammadi, "Application of deep learning technique to manage COVID-19 in routine clinical practice using CT images: Results of 10 convolutional neural networks," *Comput. Biol. Med.*, vol. 121, p. 103795, Jun. 2020, doi: 10.1016/J.COMPBIOMED.2020.103795.

[11] U. Özkaya, Ş. Öztürk, and M. Barstugan, "Coronavirus (COVID-19) Classification Using Deep Features Fusion and Ranking Technique," *Stud. Big Data*, vol. 78, pp. 281–295, 2020, doi: 10.1007/978-3-030-55258-9_17.

[12] Z. Huang *et al.*, "Multi-center sparse learning and decision fusion for automatic COVID-19 diagnosis," *Appl. Soft Comput.*, vol. 115, p. 108088, Jan. 2022, doi: 10.1016/J.ASOC.2021.108088.

[13] K. H. Abdulkareem *et al.*, "Automated System for Identifying COVID-19 Infections in Computed Tomography Images Using Deep Learning Models," *J. Healthc. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/5329014.

[14] P. gifani, A. Shalbaf, and M. Vafaeezadeh, "Automated detection of COVID-19 using ensemble of transfer learning with deep convolutional neural network based on CT scans," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 16, no. 1, pp. 115–123, Jan. 2021, doi: 10.1007/S11548-020-02286-W.

[15] A. Narin, C. Kaya, and Z. Pamuk, "Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks," *Pattern Anal. Appl. 2021 243*, vol. 24, no. 3, pp. 1207–1220, Mar. 2020, doi: 10.1007/s10044-021-00984-y.

[16] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. Jamalipour Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, p. 101794, Oct. 2020, doi: 10.1016/J.MEDIA.2020.101794.

[17] A. Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection," *IEEE Access*, vol. 8, pp. 91916–91923, 2020, doi: 10.1109/ACCESS.2020.2994762.

[18] A. Badawi and K. Elgazzar, "Detecting Coronavirus from Chest X-rays Using Transfer Learning," *COVID 2021, Vol. 1, Pages 403-415*, vol. 1, no. 1, pp. 403–415, Sep. 2021, doi: 10.3390/COVID1010034.

[19] R. K. Gupta, N. Kunhare, R. K. Pateriya, and N. Pathik, "A Deep Neural Network for Detecting Coronavirus Disease Using Chest X-Ray Images," *https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJHISI.20220401.oa1*, vol. 17, no. 2, pp. 1–27, Jan. 1AD, doi: 10.4018/IJHISI.20220401.OA1.

[20] G. H. G. S. A. D. Dhanapala and S. Sotheeswaran, "Transfer Learning Techniques with SVM For Covid-19 Disease Prediction Based On Chest X-Ray Images," *ICARC 2022 - 2nd Int. Conf. Adv. Res. Comput. Towar. a Digit. Empower. Soc.*, pp. 72–77, 2022, doi: 10.1109/ICARC54489.2022.9754029.

[21] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, "A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images," *Alexandria Eng. J.*, vol. 64, pp. 923–935, Feb. 2023, doi: 10.1016/J.AEJ.2022.10.053.

[22] M. F. Aslan, "A robust semantic lung segmentation study for CNN-based COVID-19 diagnosis," *Chemom. Intell. Lab. Syst.*, vol. 231, p. 104695, Dec. 2022, doi: 10.1016/J.CHEMOLAB.2022.104695.

[23] E. F. Ohata *et al.*, "Automatic detection of COVID-19 infection using chest X-ray images through transfer learning," *IEEE/CAA J. Autom. Sin.*, vol. 8, no. 1, pp. 239–248, Jan. 2021, doi: 10.1109/JAS.2020.1003393.

[24] K. A. Mohamed, E. Elsamahy, and A. Salem, "COVID-19 Disease Detection based on X-Ray Image Classification using CNN with GEV Activation Function," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 9, pp. 890–898, Dec. 2022, doi: 10.14569/IJACSA.2022.01309103.

[25] M. B. Er, "COVID-19 detection based on pre-trained deep networks and LSTM model using X-ray images enhanced contrast with artificial bee colony algorithm," *Expert Syst.*, p. e13185, 2022, doi: 10.1111/EXSY.13185.

[26] M. ; Ksibi *et al.*, "COVID-AleXception: A Deep Learning Model Based on a Deep Feature Concatenation Approach for the Detection of COVID-19 from Chest X-ray Images," *Healthc. 2022, Vol. 10, Page 2072*, vol. 10, no. 10, p. 2072, Oct. 2022, doi: 10.3390/HEALTHCARE10102072.

[27] K. Agrawal, R. Kumar, and S. Jain, "An Efficient Ensemble Model for Diagnosing Covid-19 and Pneumonia Using Chest X-Ray Images," *Indian J. Sci. Technol.*, vol. 15, no. 38, pp. 1900–1906, Oct. 2022, doi: 10.17485/IJST/V15I38.1269.

[28] T. N. Do, V. T. Le, and T. H. Doan, "SVM on Top of Deep Networks for Covid-19 Detection from Chest X-ray Images," *J. Inf. Commun. Converg. Eng.*, vol. 20, no. 3, pp. 219–225, Sep. 2022, doi: 10.56977/JICCE.2022.20.3.219.

[29] Y. Chandola, J. Virmani, H. S. Bhadauria, and P. Kumar, "End-to-end pre-trained CNN-based computer-aided classification system design for chest radiographs," *Deep Learn. Chest Radiogr.*, pp. 117–140, Jan. 2021, doi: 10.1016/B978-0-323-90184-0.00011-4.

[30] S. A. Singh and S. Majumder, "Short and noisy electrocardiogram classification based on deep learning," *Deep Learn. Data Anal.*, pp. 1–19, Jan. 2020, doi: 10.1016/B978-0-12-819764-6.00002-8.