Efficient COVID-19 Prediction by Merging Various Deep Learning Architectures

Zakariya A. Oraibi[1,*], Safaa Albasri[2]

[1]Department of Computer Science, College of Education for Pure Sciences, University of Basrah, Basrah, Iraq
[2]Electrical Engineering Department, College of Engineering, Mustansiriyah University, Baghdad, Iraq
E-mail: zakaria_au@uobasrah.edu.iq, asafaa@uomustansiriyah.edu.iq
*Corresponding author

Keywords: COVID-19, deep learning, coronavirus, hybrid models, feature maps

Received: November 15, 2023

In late 2019, COVID-19 virus emerged as a dangerous disease that led to millions of fatalities and changed how human beings interact with each other and forced people to wear masks with mandatory lockdown. The ability to diagnose and detect this novel disease can help in isolating the infected patients and curb the spread of the virus. Artificial intelligence techniques including machine learning and deep learning showed huge potential in accurately classifying COVID-19 chest X-ray images. In this paper, we propose to combine the feature maps of multiple powerful CNN models (Xception, VGG-16, VGG-19) using the rule of sum. Each of these models is trained from scratch and tested on the given test images. The dataset was collected from a large public repository of COVID images with three classes: COVID, Normal, and Pneumonia. During experiments, data augmentation is also applied to provide more training samples. Experimental results show that combining multiple models improve the classification accuracy and achieve better performance than standalone models. An accuracy of 97.91% was achieved using a combination of three models which outperforms state-of-the-art techniques.

Povzetek:

1 Introduction

The pandemic that started more than three years ago caused by coronavirus (SARS-COV-2) led to changing the world drastically from forcing people to wear masks in public areas to driving nations to impose a complete lockdown to prevent its spread [1, 2]. Infected people can easily spread the virus through sneeze, cough, or even by speaking. The major concern related to COVID-19 is that it affects human lung by damaging the tissues. Symptoms at early stages of infection could be similar to the usual flu infection and may include fever, throat pain, and headache [3]. It is estimated that the virus infected hundreds of millions of people and caused millions of fatalities until a vaccine was developed in 2021 which limited the danger of the infection and restored life to normality [4].

Since COVID-19 virus is very contagious, it should be detected early to help isolating and curing the patient. The traditional technique of diagnosing coronavirus involves using Polymerase Chain Reaction (PCR) [11]. However, using PCR to perform such test is also challenging due its low sensitivity leading to difficulty in detecting positive COVID-19 cases. In addition, PCR needs more time to obtain the results with limited availability of kits in clinics and hospitals [12]. As a result, chest X-ray images can be used for screening by employing machine learning techniques to automate the process of detecting COVID-19 cases. These techniques range from traditional image classifiers like SVM, kNN, and RFs to deep learning models like VGG-16, VGG-19, Xception, Inception, MobileNet, and ResNet [13, 14].

In literature, Kaur et al. [5] presented a hybrid deep learning model to classify COVID-19 images into three and four classes. InceptionV4 model is applied to extract image features then, SVM is applied to classify COVID-19 samples. Their method achieved a high accuracy of 95.51% on three classes COVID-19 dataset. Hossain et al. [6] applied transfer learning with fine-tuning on different state-of-the-art deep learning models including ResNet50, VGG-16, InceptionV3, and MobileNet-V2. A high accuracy of 99.17% was achieved on a public COVID-19 dataset with two classes. Monshi et al. [7] proposed to optimize data augmentation and training hyperparameters using EfficientNetB0 model. The new model is named CovidXrayNet which achieved a high classification accuracy of 95.82%. Kaya et al. [8] proposed using MobileNet model with new fine-tuning technique to predict COVID-19 images. They applied their approach on a large dataset with three classes and achieved high accuracy. Kausar et al. [9] designed an end-to-end model called Style Distribution transfer Generative Adversarial Network (SD-GAN) and managed to achieve high classification accuracy on X-ray datasets. Ensemble learning techniques were used by

Table 1: Summary of related work.

Technique	Year	Dataset	Performance
Features extraction using InceptionV4 + SVM Classifier [5]	2021	1900 Chest XR	Detection accuracy = 95.51%
Transfer learning + Fine-tuning CNN architectures [6]	2022	COVID-19 Radiography dataset	Classification accuracy = 99.17 (two classes)%
Ensemble of pre-trained models [18]	2021	2326 X-ray images	Sensitivity = 90.5% , Specificity = 90.0%
Covidxraynet [7]	2021	COVIDx dataset	Classification accuracy = 95.82%
MobileNet + Novel Fine-tuning [8]	2023	9457 COVID XR	Classification accuracy = 97.61%
SD-GAN [9]	2023	[A-D] datasets	Accuracy = [98.7%, 99.3%, 99.1%, 98.6%]
New CNN [10]	2023	COVID-Xray-5k	Sensitivity = 95% , Specificity = 99.32%

Aldhahi et al. [15] by introducing a new technique called Uncertain-CAM, which provides better classification accuracy for COVID-19 images. In Our recent work [16, 10], we proposed to train a new CNN architecture from scratch to classify two COVID-19 classes. We achieved 95.0% and 99.32% in both sensitivity and specificity rates on a standard COVID-19 dataset. In general, many methods have been proposed in literature and the majority of them rely on deep learning models and how to exploit multiple architectures to improve the prediction performance. Table 1 summarizes the work in literature.

In this paper, we exploit the power of pre-trained stat-of-the-art deep learning architectures by merging them to improve the COVID-19 classification accuracy. These models include VGG-16, VGG-19, and Xception and were combined using the rule of sum. In addition, COVID-19 images were collected from two public image repositories where two classes (COVID and Normal) where collected from a huge set of images available online for research purposes. The third class, Pneumonia, was collected from the work of [17]. In total we collected 964 images. The proposed technique is simple yet it improves the classification accuracy and provides a good baseline for future developments.

The structure of the paper is organized as follows: Section 2 describes in details the methodology of hybrid CNN models. Section 3 describes in details the dataset used in the experiments. Section 4 lists the results of the proposed approach. Section 5 discusses the results and finally, we present conclusions and future work in section 6.

2 Hybrid CNN models

The methodology used in our work relies on merging various state-of-the-art CNN architectures. These models include Xception, VGG-16, and VGG-19 [18, 19]. The intuition behind this technique is that each model proved to be very powerful and capable of extracting the required patterns from the image and perform very well in computer vision tasks. As a result, combining the feature maps of these models will result in an efficient hybrid model that can be used to obtain high classification accuracy. Xception model produces $7 \times 7 \times 2048$ feature maps on its last layer of feature extractor. Both VGG-16 and VGG-19 produce $7 \times 7 \times 512$ features maps. The concatenation of these models will result in a feature map of $7 \times 7 \times 3072$ in size. Finally, a flatten layer is added with three dense layers. Figure (1) shows the pipeline of the proposed approach.

In order to apply our approach, the COVID-19 dataset of three classes is divided into two subsets, training and testing. During the training phase, images are further divided into training and validation. In order to train the model efficiently, data augmentation has been used by modifying the training set of images using techniques like shearing, zooming, image flipping, and shifting. Then, each model is imported and the trainable layers are turned off. After that, the concatenated model is used during training. When the model is finished training, the accuracy is computed using the subset of test images to report the classification results. More details about each model used in our approach are given in the following subsections.

2.1 Xception model

This powerful model was introduced by Google in 2017 which is inspired by Inception model [18]. In Xception, authors proposed decoupling the correlations of cross-channels and spatial correlations in the feature maps of CNN. The model has 36 convolutional layers organized into 14 modules. Linear residual connections are used in every module except for the first and last modules. The structure of this model makes it very easy to modify since it is implemented as a linear stack of depth-wise separable convolution layers. In our work, we used the standard Xception model and trained it from scratch. Figure (2) shows the layers of the model.

2.2 VGG models

Both VGG-16 and VGG-19 CNN models were introduced in 2014 by simonyan et al. [19]. The architecture of these

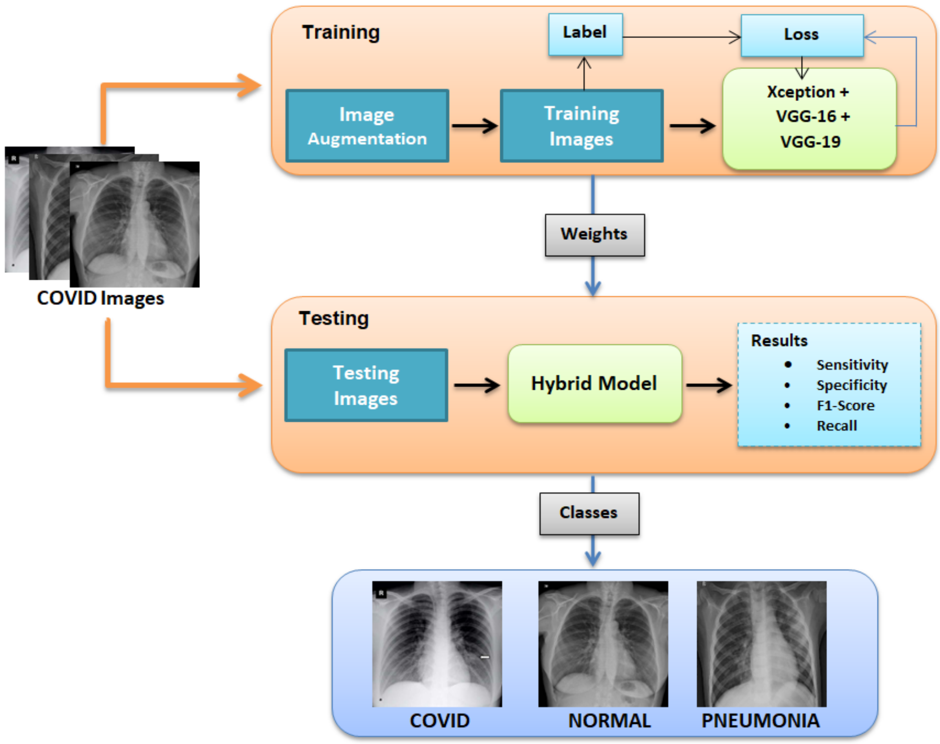

Figure 1: COVID-19 classification pipeline. Training is performed after the feature maps of the three models (VGG-16, VGG-19, and Xception) are concatenated to form a bigger input layer. Then, weights are used to predict the testing images into three classes.

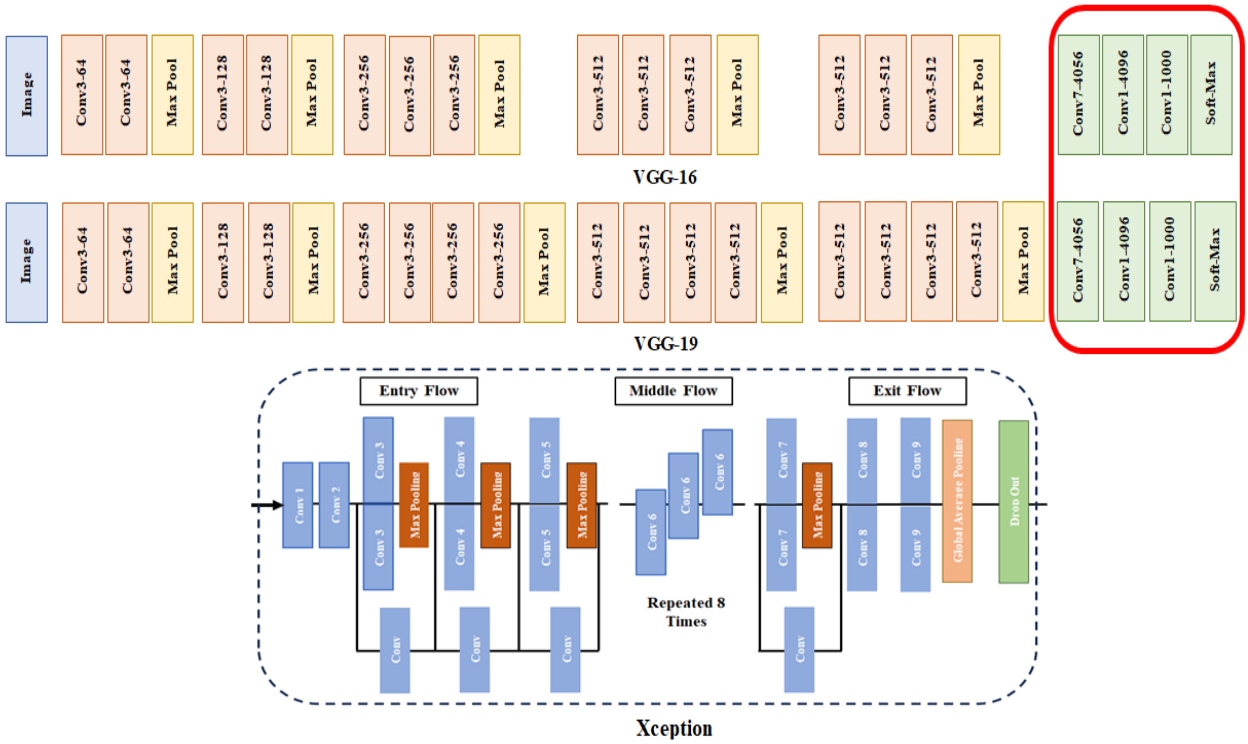

Figure 2: Architectures of VGG-16, VGG-19, and Xception Models.

models uses a fixed size of input image ($224 \times 224 \times 3$). A very small receptive field filter is used (3×3) with a convolutional stride fixed to 1. Five max-pooling layers are employed which are performed over a (2×2) pixel window, having a stride of 2. Three fully connected layers are added at the end of the stack of convoluitonal layers. Finally softmax layer is used for classification. The difference between VGG-16 and VGG-19 is that the first one has 16 convolutional layers while the latter has 19 convolutional layers. In the experiments, VGG-19 always generates better results than VGG-16 especially in the classification tasks. In our work, we used both models and trained them from scratch. We don not rely on the weights trained on ImageNet dataset or transfer learning. Figure (2) shows the structure of both models.

3 Materials

3.1 Dataset description

The hybrid CNN model proposed in this paper is applied on a dataset collected from a public repository in the link below [1]. The dataset consists of three classes: COVID, Normal, and Pneumonia. Both COVID and Normal classes were collected from the aforementioned repository. The third class, Pneumonia, was collected from the work of Shastri et al. [17].

In total we have 964 images across the three classes. Images of the dataset vary in size. As a result we had to resize all images to ($224 \times 224 \times 3$) to meet the VGG-16, VGG-19, and Xception models needs. The number of images per class are listed in Table 2. Sample images of each class are shown in Figure (3).

Table 2: Number of images per class of the COVID dataset.

Class	No. of Images
COVID	357
Normal	365
Pneumonia	242

Table 3: Total number of COVID, Normal, and Pneumonia classes training images after augmentation.

Class	Original	Augmented
COVID	285	1460
Normal	292	1425
Pneumonia	193	965

[1] https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset

Table 4: Hyperparameters used in the experiments.

Hyperparameter	Value
Epochs	5
Learning Rate	0.001
Optimizer	Adam
Batch Size	64
Loss Function	Categorical Crossentropy
Pooling Size	2×2

4 Experimental results

The results of applying the hybrid model proposed in this paper are presented in this section. Five-fold cross validation was used in the experiments. This means, 80% of the original data were used for training and 20% were used for testing. The final accuracy was generated by finding the average of the five-fold results. As we mentioned earlier, augmentation was used to increase the number of training images. Table 3 shows the number of images used for training in fold 1 and the corresponding number of images resulted from augmenting the training set samples.

To evaluate the performance of our approach, five metrics were used: accuracy, sensitivity, specificity, precision, and F1 score. The corresponding equations to compute each metric are given in the equations below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

$$Sensitivity = \frac{\sum TN}{\sum FP + \sum TN} \tag{2}$$

$$Specificity = \frac{\sum TP}{\sum TP + \sum FN} \tag{3}$$

$$F1\ Score = \frac{2\sum TP}{2\sum TP + FP + FN} \tag{4}$$

$$Precision = \frac{\sum TP}{\sum TP + \sum FP} \tag{5}$$

The combined model of Xception, VGG-16, and VGG-19 proposed in this paper is implemented using Keras software [20]. Google Colaboratory was used to benefit from GPU. In the training stage, Adam optimizer is used. The batch size used during training is fixed to 64. This is to prevent any session crash. The number of epochs applied in the training stage was fixed to 5 for all experiments. The reason to choose such a small number of epochs is because the network converges very fast. Learning rate metric was fixed to 0.001. For the loss function, we employed cross-entropy in the implementation. Table 4 summarizes the hyperparameters used in the implementation.

Training accuracy and validation accuracy converged very fast after only 5 epochs as shown in Figure (4). This is because we are using a combination of three powerful

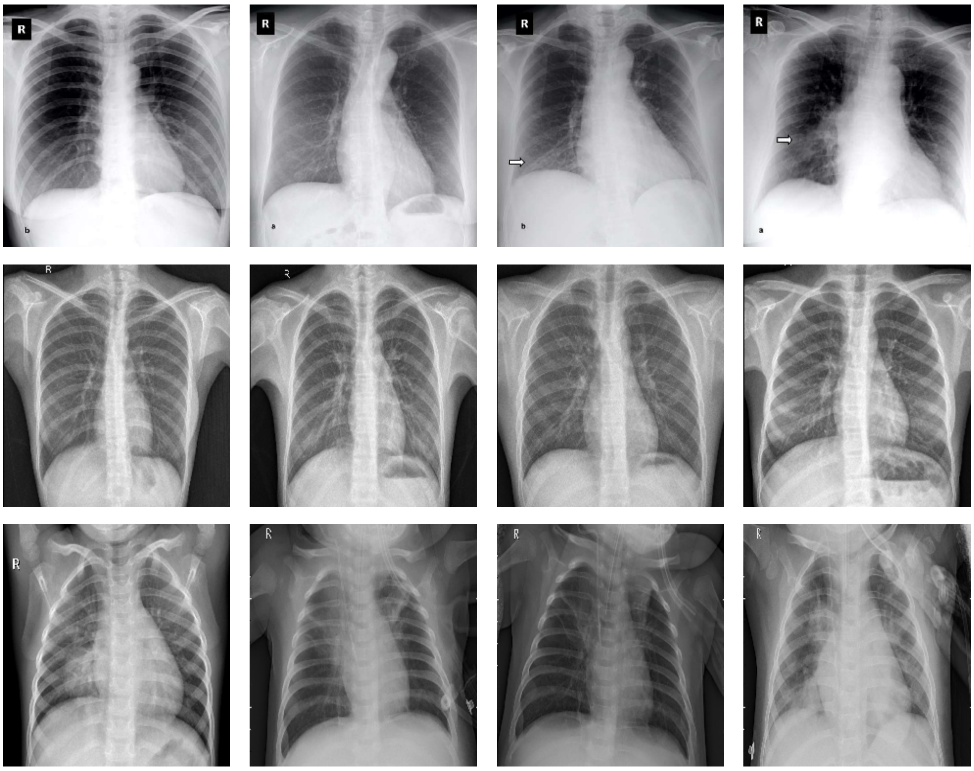

Figure 3: Sample images of the COVID dataset. images in row 1 represent COVID class. Images in row 2 represent Normal class. Finally, images in row represent Pneumonia class.

models during training: Xception, VGG-16, and VGG-19. In addition, the graph on the right of Figure (4) shows that training loss and validation loss reach to 0 after only 5 epochs.

Table 5 shows the results of the experiments conducted using our approach of combining multiple CNN models. Xception and VGG-16 models achieved decent results of 94.31% and 93.32%. On the other hand, VGG-19 outperformed them by achieving a high accuracy of 96.91%. However, combining Xception and VGG-16 together outperformed each model alone by scoring 95.85%. Finally, combining the three models outperformed all the previous results with 97.91% accuracy. Other metrics including sensitivity, specificity, F1 score and precision are also very high for the triple combined model. Figure (5) shows the confusion matrix generated from applying our hybrid approach using Adam optimizer with 0.001 learning rate. For the COVID and Normal classes, only 2 samples were misclassified as Pneumonia for each. For the Pneumonia class, all test samples were classified correctly.

In the previous work related to COVID classification, authors proposed several approaches for COVID-19 prediction and applied them on various datasets including standard ones and ones that were collected from public repositories. Table 6 provides a comparison in accuracy between our proposed methodology and other state-of-the-art approaches applied on three-classes COVID datasets. Ioannis et al. [21] proposed using VGG-19 with transfer learning and achieved 93.48% accuracy. Ali et al. [22] suggested a new deep learning architecture called DRE-Net for COVID-19 automatic detection and achieved 95.51% accuracy. COVID-Net was proposed by Wang et al. [23] and scored 92.4%. Tulin et al. [24] proposed DarkCovid-Net and achieved a low accuracy of 87.02%. A work that is similar to our proposed technique is Xu et al. [25] which used ResNet architecture with location attention achieved an accuracy of 86.7%. Our approach of merging multiple CNN models proved to outperform these models. In addition, the standalone models used in our work were trained from scratch without any need for transfer learning.

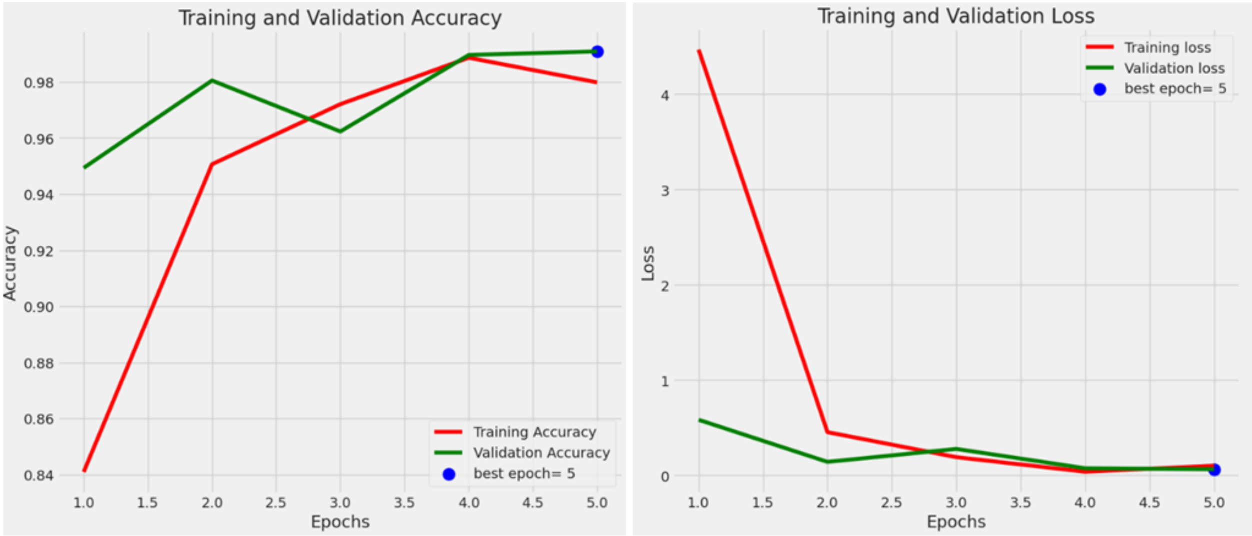

Figure 4: Training accuracy vs validation accuracy legends are on the left. Training loss vs validation loss legends are on the right. These graphs are the results of training our hybrid CNN model of Xception, VGG-16, and VGG-19 using 5 epochs.

Table 5: Performance of our hybrid model using five accuracy metrics.

Model	Accuracy	Sensitivity	Specificity	F1 score	Precision
Xception	94.31%	94.0%	93.31%	93.31 %	94.31%
VGG-16	93.32%	94.0%	93.67%	93.67%	93.32%
VGG-19	96.91%	% 96.32	97.29%	97.63%	96.88%
Xception + VGG-16	95.85%	95.63%	95.31%	95.32%	95.81%
Xception + VGG-16 + VGG-19	**97.91%**	**97.33%**	**98.0%**	**97.96%**	**97.9%**

5 Discussion

COVID prediction approach employed in this paper relies on pre-trained state-of-the-art CNN architectures. We trained three well-known models from scratch: Xception, VGG-16, and VGG-19 and combined them using the rule of sum applied on the feature maps. Every model accepts images of size $224 \times 224 \times 3$ and images were collected from various resources and grouped into three classes: COVID, Normal, and Pneumonia. Results of applying this approach showed huge potential resulted from merging these models. The combination of the first two models, Xception + VGG-16, improved the accuracy by almost 1% from the standalone Xception model. Furthermore, the combination of the three models, Xception + VGG-16 + VGG-19, further improved the best accuracy by almost 1% achieving 97.91%. It is worthy to mention that training parameters for all models were selected to achieve best training accuracy. We used fewer number of epochs, 5, and showed that the best hybrid model converges quickly with that number. The learning rate was selected to be 0.001 and Adam optimizer was used across all experiments. Data augmentation was also employed to increase the number of training images for the three classes.

The techniques summarized in Table 1 relied on transfer learning, ensemble learning, and deep features merging. Some of these methods were applied on two classes COVID-19 datasets while others were applied on three and four classes. We showed that using a combination of deep learning models with data augmentation and batch size of 64 can produce superb results and can outperform the standalone models performance. The merge of Xception, VGG-16, and VGG-16 improved the accuracy by 1% than the best standalone model (VGG-19). The potential of combining various CNN architectures is huge and we intend to apply more models and merge them to further improve the accuracy.

6 Conclusion

The methodology proposed in this paper is simple yet it is very effective in achieving high accuracy multi-class prediction. State-of-the-art deep learning models were used and trained from scratch using the same hyperparameters in each experiment. These models include Xception, VGG-16, and VGG-19. Then, we combined Xception and VGG-

Table 6: Evaluating the proposed hybrid model with the previous work applied on COVID datasets of three classes.

Methodology	**Accuracy (%)**
VGG-19 [21]	93.49
DRE-Net [22]	95.51
Covid-Net [23]	92.4
DarkCovidNet [24]	87.02
ResNet + Location Attention [25]	86.7
Proposed (Xception + VGG-16)	**95.85**
Proposed (Xception + VGG-16 + VGG-19)	**97.91**

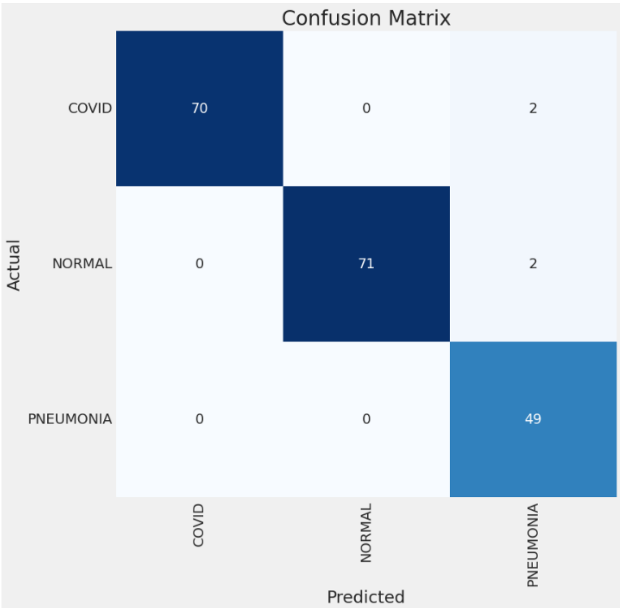

Figure 5: Confusion matrix generated for the hybrid model.

16 and showed that this binary hybrid model is better than each model alone. After that, we combined the three models and showed the final merged model is better that the three models alone and better than the combination of Xception and VGG-16. We set the number of epochs to 5 and learning rate to 0.001 with a batch size of 64. Adam optimizer was employed and all experiments were performed on Google Colaboratory with GPU. The dataset used in the experiments was collected from multiple resources with three classes: COVID, Normal, and Pneumonia. Five-fold cross validation was used in the experiments with 80% of samples in each class were used for training and the remaining 20% were used for testing. Augmentation techniques were applied to the subset of training images to enlarge it to perform well during training. Five accuracy metrics were reported: Accuracy, precision, F1-score, sensitivity, and specificity. The best accuracy achieved from merging the three models was 97.91%.

For the future work, since physicians need a robust COVID classification system, we will focus on improving the accuracy by applying ensemble methods with various voting mechanisms. In addition, we will explore training more state-of-the-art models to check their performances on binary and multi-class COVID datasets. Furthermore, we will apply our method on bigger COVID datasets with various hyperparameters settings. Our main goal will remain to achieve the highest classification accuracy since this system is related to human lives.

Ethical considerations and practical applicability

The confidentiality of patient's COVID-19 data is very crucial and must adhere to specific protocols in regard to privacy and security of these information. In addition, COVID-19 detection using machine learning models can easily be biased because of the nature of datasets used during training the models. Hence, careful evaluation and mitigation of model bias is very important. Moreover, explaining to patients how the decision was made by a machine learning model is very necessary since these models are complex. In terms of practical applicability, COVID-19 prediction systems that depend on AI must adhere to regulatory scrutiny to make sure they are efficient, safe, and amenable with healthcare system regulations.

References

[1] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, *et al.*, "A novel coronavirus from patients with pneumonia in china, 2019," *New England journal of medicine*, vol. 382, no. 8, pp. 727–733, 2020.

[2] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, *et al.*, "A new coronavirus associated with human respiratory disease in china," *Nature*, vol. 579, no. 7798, pp. 265–269, 2020.

[3] C. Wang, P. W. Horby, F. G. Hayden, and G. F. Gao, "A novel coronavirus outbreak of global health con-

cern," *The lancet*, vol. 395, no. 10223, pp. 470–473, 2020.

[4] T. T. Le, Z. Andreadakis, A. Kumar, R. G. Román, S. Tollefsen, M. Saville, S. Mayhew, *et al.*, "The covid-19 vaccine development landscape," *Nat Rev Drug Discov*, vol. 19, no. 5, pp. 305–306, 2020.

[5] P. Kaur, S. Harnal, R. Tiwari, F. S. Alharithi, A. H. Almulihi, I. D. Noya, and N. Goyal, "A hybrid convolutional neural network model for diagnosis of covid-19 using chest x-ray images," *International Journal of Environmental Research and Public Health*, vol. 18, no. 22, p. 12191, 2021.

[6] M. B. Hossain, S. H. S. Iqbal, M. M. Islam, M. N. Akhtar, and I. H. Sarker, "Transfer learning with fine-tuned deep cnn resnet50 model for classifying covid-19 from chest x-ray images," *Informatics in Medicine Unlocked*, vol. 30, p. 100916, 2022.

[7] M. M. A. Monshi, J. Poon, V. Chung, and F. M. Monshi, "Covidxraynet: Optimizing data augmentation and cnn hyperparameters for improved covid-19 detection from cxr," *Computers in biology and medicine*, vol. 133, p. 104375, 2021.

[8] Y. Kaya and E. Gürsoy, "A mobilenet-based cnn model with a novel fine-tuning mechanism for covid-19 infection detection," *Soft Computing*, vol. 27, no. 9, pp. 5521–5535, 2023.

[9] T. Kausar, Y. Lu, A. Kausar, M. Ali, and A. Yousaf, "Sd-gan: A style distribution transfer generative adversarial network for covid-19 detection through x-ray images," *IEEE Access*, vol. 11, pp. 24545–24560, 2023.

[10] Z. A. Oraibi and S. Albasri, "A robust end-to-end cnn architecture for efficient covid-19 prediction form x-ray images with imbalanced data," *Informatica*, vol. 47, no. 7, 2023.

[11] E. E. Walsh, A. R. Falsey, I. A. Swinburne, and M. A. Formica, "Reverse transcription polymerase chain reaction (rt-pcr) for diagnosis of respiratory syncytial virus infection in adults: Use of a single-tube "hanging droplet" nested pcr," *Journal of medical virology*, vol. 63, no. 3, pp. 259–263, 2001.

[12] A. R. Khan, T. Saba, T. Sadad, *et al.*, "Computer vision in co-clinical medical imaging for precision medicine," 2022.

[13] H. S. Basavegowda and G. Dagnew, "Deep learning approach for microarray cancer data classification," *CAAI Transactions on Intelligence Technology*, vol. 5, no. 1, pp. 22–33, 2020.

[14] A. Ahmad, D. Saraswat, V. Aggarwal, A. Etienne, and B. Hancock, "Performance of deep learning models for classifying and detecting common weeds in corn and soybean production systems," *Computers and Electronics in Agriculture*, vol. 184, p. 106081, 2021.

[15] W. Aldhahi and S. Sull, "Uncertain-cam: Uncertainty-based ensemble machine voting for improved covid-19 cxr classification and explainability," *Diagnostics*, vol. 13, no. 3, p. 441, 2023.

[16] Z. A. Oraibi and S. Albasri, "Predicting covid-19 from chest x-ray images using a new deep learning architecture," in *2022 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1–6, IEEE, 2022.

[17] S. Shastri, I. Kansal, S. Kumar, K. Singh, R. Popli, and V. Mansotra, "Cheximagenet: a novel architecture for accurate classification of covid-19 with chest x-ray digital images using deep convolutional neural networks," *Health and Technology*, vol. 12, no. 1, pp. 193–204, 2022.

[18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[19] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[20] F. Chollet *et al.*, "Keras." https://github.com/fchollet/keras, 2015.

[21] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and engineering sciences in medicine*, vol. 43, pp. 635–640, 2020.

[22] A. Al-Bawi, K. Al-Kaabi, M. Jeryo, and A. Al-Fatlawi, "Ccblock: an effective use of deep learning for automatic diagnosis of covid-19 using x-ray images," *Research on Biomedical Engineering*, pp. 1–10, 2020.

[23] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific reports*, vol. 10, no. 1, p. 19549, 2020.

[24] S. Asif, Y. Wenhui, H. Jin, Y. Tao, and S. Jinhai, "Automatic detection of covid-19 using x-ray images with deep convolutional neural networks and machine learning," 2020.

[25] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, *et al.*, "A deep learning system to screen novel coronavirus disease 2019 pneumonia," *Engineering*, vol. 6, no. 10, pp. 1122–1129, 2020.