A Comparative Study of Stand-Alone and Hybrid CNN Models for COVID-19 Detection

Wedad Alawad[1], Banan Alburaidi[2], Asma Alzahrani[3], Fai Alflaj[4]

Department of Information Technology, College of Computer,
Qassim University, Buraydah, 51452, Saudi Arabia[1]
Department of Information Technology, College of Computer,
Qassim University, Buraydah, 51452, Saudi Arabia[2,3,4]

Abstract—The COVID-19 pandemic continues to impact both the international economy and individual lives. A fast and accurate diagnosis of COVID-19 is required to limit the spread of this disease and reduce the number of infections and deaths. However, a time consuming biological test, Real-Time Reverse Transcription–Polymerase Chain Reaction (RT-PCR), is used to diagnose COVID-19. Furthermore, sometimes the test produces ambiguous results, especially when samples are taken in the early stages of the disease. As a potential solution, machine learning algorithms could help enhance the process of detecting COVID-19 cases. In this paper, we have provided a study that compares the stand-alone CNN model and hybrid machine learning models in their ability to detect COVID-19 from chest X-Ray images. We presented four models to classify such kinds of images into COVID-19 and normal. Visual Geometry Group (VGG-16) is the architecture used to develop the stand-alone CNN model. This hybrid model consists of two parts: the VGG-16 as a features extractor, and a conventional machine learning algorithm, such as support-vector-machines (SVM), Random-Forests (RF), and Extreme-Gradient-Boosting (XGBoost), as a classifier. Even though several studies have investigated this topic, the dataset used in this study is considered one of the largest because we have combined five existing datasets. The results illustrate that there is no noticeable improvement in the performance when hybrid models are used as an alternative to the stand-alone CNN model. VGG-16 and (VGG16+SVM) models provide the best performance with a 99.82% model accuracy and 100% model sensitivity. In general, all the four presented models are reliable, and the lowest accuracy obtained among them is 98.73%.

Keywords—COVID-19; convolutional neural network; hybrid models; chest X-Ray; deep learning

I. INTRODUCTION

The continuous outbreak of the novel coronavirus was first reported in Wuhan, Hubei Province, China. In a preliminary report, it was revealed that the virus shares an 88% serial identity with two coronaviruses derived from bats, similar to SARS. The new coronavirus was preliminarily named nCov-2019. In February 2020, a study group from the International Committee on Taxonomy of Viruses classified the virus as SARS-CoV. Shortly thereafter, the World Health Organization (WHO) formally named the disease caused by the novel coronavirus "COVID-19" [1].

The most common symptoms of this disease are coughing, headaches, fatigue, shortness of breath, loss of smell, pain in the throat, and a high temperature. COVID-19 continues to have a destructive impact on global health and commerce, as

well as individuals' lives. The number of infections and deaths increases day by day. As of April 18, 2021, the total number of infected persons around the world has reached 141 million with a mortality rate of 3 million.

Therefore, there is a need to cooperate across disciplines and integrate resources to defend against COVID-19 and prevent it from further spread. Because this virus has the ability to spread fast between people, the first and most important step in COVID-19 defense is early detection. The most common method used to detect COVID-19 is the Transcription-Polymerase Chain Reaction (RT-PCR) test. The RT-PCR test usually takes up to 6 hours to produce results [1]. This waiting time is long when we consider the urgent need to test millions of samples and receive the results as fast as possible to prevent those who are infected from spreading the disease to others. Additionally, sometimes patients who have already been exposed to the virus and show severe symptoms could still get false negative results in the (RT-PCR) test [1].

Thus, the development of a fast and efficient alternative is necessary to improve the process of diagnosing COVID-19. One possibility is to exploit the fact that the disease can be diagnosed using chest X-Ray images. In contrast to other types of medical imaging, X-Ray images are available at most hospitals, and have been routinely used in COVID-19 diagnoses thus far. Moreover, they are cost-efficient and quickly produce results [2]. Machine learning and deep learning algorithms are promising in this case. Potential automated COVID-19 detection models can be provided by training such algorithms on chest X-Ray image datasets. These models may aid in producing test results within a few seconds as well as reduce inaccurate results.

This paper addresses the following research question: How would replacing the fully connected layer with a machine learning classifier affect the classification model performance? To answer this question, we examined the efficiency of two types of machine learning models used to detect COVID-19. Using one of the largest available chest X-Ray image datasets, we have built four different models to study and compare between the stand-alone CNN model and hybrid machine learning models with regard to their general effectiveness, giving particular attention to their COVID-19 classifying abilities.

The rest of this paper is structured as follows: Section II provides background and analysis of pre-existing relevant studies. Section III discusses the details of datasets and methodologies used in this study. Section IV presents and discusses

the results of our experiments. Section V concludes the paper by mentioning the most prominent points of this study. Finally, section VI provides ideas that may be implemented in the future.

II. RELATED WORK

Several studies have investigated the use of machine learning techniques to detect COVID-19. Among machine learning algorithms, most researchers used CNN techniques, e.g., Inception (GoogleNet), ResNet (Residual Networks), and DenseNet (Dense Networks), to build the detection models. From the dataset perspective, chest X-Ray images were used more frequently than Computed Tomography (CT) images to develop those models.

Most of the related studies have faced challenges due to a lack of available datasets. Researchers have applied methods such as augmentation and K-fold cross validation to overcome this setback. In [3], the dataset size reached 1,592 images after augmentation, and the authors used the ImageNet dataset and four CNN techniques to build detection models. The accuracy of VGG16 and VGG19 based models were the highest, having achieved a 99.38% accuracy. In contrast, K-fold cross validation was the method used to train the detection models on more data in [4]. In that study, VGG 16 and ResNet50 techniques were used to develop models that distinguish between COVID-19 and pneumonia. The dataset was comprised of 204 images, and their results showed that 89.2% and 80.39% of COVID-19 cases were identified correctly by these techniques, respectively.

In addition to the rarity of available datasets, the quality of obtainable images needs to be enhanced to improve the performance of detection models. To achieve this, the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm was applied in [5]. The authors conducted a comparison study to investigate the effect of using CLAHE to enhance covid diagnosis. They compared the detection accuracy of the model when the CLAHE was applied and when the original datasets were used to build the model without applying any image quality enhancement techniques. The accuracy of the developed model increased from 83.00% to 92.00% after the implementation of CLAHE.

Despite COVID-19 dataset challenges, many detection models with acceptable performance have been presented. The transfer learning concept was used to build detection models in previous studies. In [6], the researchers developed a model that detected COVID-19 from chest X-ray images. They used a modified version of VGG-19 by adding a MLP (multilayer perceptron) on top of the VGG-19 model. The accuracy of the model was 96.3%. Additionally, the authors in [7] used seven pre-trained models to develop a deep learning framework that identified COVID-19 cases. Their results showed that VGG19 and DenseNet201 achieved better performance when compared with other models. The accuracy was 90% and the sensitivity was 100% for COVID-19 and 80% for normal images.

Some papers have used hybrid models, meaning they combined multiple models to solve one problem. The authors in [8] used four different types of ensemble learning, feature ensembles, majority voting, feature classification, and class modification, to classify COVID-19 and pneumonia cases.

More specifically, SVM, Bagging Classifier, and AdaBoost were used as classifiers in the models they developed. The accuracy of the combination of Inception V3 and Bagging was 99.36

Table I shows a summary of previous studies that used CNN architectures to build COVID-19 detection models. Along with the performance of detection models, the table illustrates the size of datasets and the CNN technique used.

III. STAND-ALONE CNN MODEL VS. HYBRID MODELS

In this study we have conducted several experiments to investigate the abilities of stand-alone CNN and hybrid models to efficiently and accurately detect COVID-19 in patients. This section presents the datasets and methodologies used in our comparative study.

A. Dataset Collections

As aforementioned, one of the challenges that researchers have faced in previous studies is the limited repository of COVID-19 datasets. Moreover, the datasets that do exist are relatively small. Thus, we used images from four datasets to train and develop our models on COVID-19 cases. For normal cases, we used one dataset. Fig. 1 shows samples of COVID-19 and normal chest X-ray images and Table II illustrates a summary of these datasets.

TABLE I. SUMMARY OF CNN-BASED MODELS FOR COVID-19 DETECTION

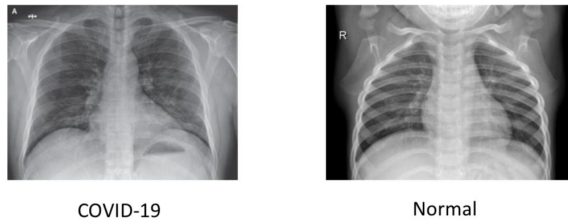| Ref.	Dataset size	CNN Technique	Accuracy	Sensitivity	Specificity
A Deep Learning Approach to Detect COVID-19 Patients from Chest X-ray Images [7]	COVID-19 : 295 Normal : 659	Sequential CNN	98.3%	-	-
A Deep-Learning-Based Framework for Automated Diagnosis of COVID-19 Using X-ray Images [3]	COVID-19 : 790 Normal : 802	VGG19	99.38%	100%	98.77%
Using X-ray images and deep learning for automated detection of coronavirus disease [9]	Pneumonia : 4,273 COVID-19 : 231 Normal : 1583	Incption-Resnet-V2	92.18%	92.11%	96.06%
COVID-19 Detection in Chest X-ray Images using a Deep Learning Approach [5]	Pneumonia : 647 COVID-19 : 204 Normal : 649	VGG16	92%	94.92%	92%
Deep learning COVID-19 detection bias- accuracy through artificial intelligence [6]	COVID-19 : 181 Normal : 364	VGG19	96.3%	97.1%	-
COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks [10]	Pneumonia : 700 COVID-19 : 224 Normal : 504	VGG19	98.75%	92.85%	98.75%
Finding COVID-19 from Chest X-Rays using Deep Learning on a Small Dataset [4]	COVID-19 : 102 Pneumonia : 102	ResNet50	89.2%	-	-
Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning [11]	COVID-19 : 520 Normal : 5000	Framework of(ResNet18, ResNet50, SqueezeNet, DenseNet-121)	-	98%	90%
COVIDX-Net: A Framework of Deep Learning ClassiFiers to Diagnose COVID-19 in X-Ray Images [12]	COVID-19 : 25 Normal : 25	VGG19, DenesNet121	90%	100%	-
Deep Learning Based Hybrid Models for Prediction of COVID-19 using Chest X-Ray [8]	Pneumonia : 4273 COVID-19 : 1182 Normal : 1583	VGG-16	90.19%	94.16%	-
		VGG-16 & SVM	91.19%	93.15%	-
		VGG-16 & Bagging	90.19%	92.16%	-
		VGG-16 & Ada Boot	90.19%	90.10%	-

Fig. 1. Normal and COVID-19 Chest X-Ray Images.

1) COVID-19 chest X-ray images datasets: The details of the four COVID-19 chest X-ray image datasets are as follows:

COVID-19_dataset-1: We retrieved this dataset from the Github repository, and it is more popular than any other currently available datasets. It was created and collected by Joseph Paul Cohen, a postdoctoral fellow at Montreal University [13]. The dataset contains 930 images of chest X-ray and CT images of patients with diverse diseases, including both bacterial and viral illnesses, as well as COVID-19 and pneumonia. COVID-19 images alone account for 584 of the dataset's images, with the remainder classified as other. The chest X-ray images are classified into four views, Posterior Anterior (PA), Anterior Posterior (AP), AP Supine, and Lateral.

COVID-19_dataset-2: We obtained the second dataset from the Github repository. It was created by Linda Wang and colleagues from the University of Waterloo in Canada [14]. The dataset contains 238 chest X-ray images, 58 of which are images of patients infected with COVID-19. The COVID-19 images are classified with two views: 32 images are PA and 26 images are AP.

COVID-19_dataset-3: The previous team, Linda Wang and colleagues, created the third dataset as well. This dataset contains 55 chest X-ray images, 35 of which are COVID-19 images and the rest are either Pneumonia or not classified [15].

COVID-19_dataset-4: Dataset 4 is also from the Github repository. It was created by the Institute for Diagnostic and Interventional Radiology at the Hannover Medical School in Hannover, Germany [16]. It contains 243 images of COVID-19 chest X-ray images. Those images include two views: 49 are PA images and 194 images are AP.

2) Normal Chest X-ray Images Datasets: The details of the normal chest X-ray image datasets are as follows:

Normal-dataset: We obtained this dataset from the Kaggle website. It was created by Paul Mooney, Developer Advocate at Kaggle [17]. The dataset contains 5,863 chest X-Ray images with two classes, pneumonia and normal. The number of normal images is 1,583 and the number of pneumonia images is 4,273. All images in this dataset were in AP view. To balance our data, we took only 690 normal images from this dataset.

3) Data Preprocessing: We converted all images to JPEG format to provide ease by handling only one format type, and to reduce the dataset size to accelerate the training process. Furthermore, we applied normalization to improve image clarity and overall quality. Furthermore, we resized them to 224*224. We tried several rations to split data into train and

TABLE II. DETAILS OF DATASETS USED

Dataset	Medical Images Type	Classes	View	Count of COVID-19 Images	Dataset Size
COVID-chestxray-dataset [13]	X-ray , CT	COVID-19, viral, bacterial, pneumonias	AP, PA, AP Supine, Lateral	504	930
Actualmed-COVID-chest xray-dataset [14]	X-ray	COVID-19 , no finding	PA, AP	58	238
Figure1-COVID -chestxray-dataset [15]	X-ray	COVID-19, pneumonias , no finding	-	35	55
COVID-19-image-repository [16]	X-ray	COVID-19	AP ,PA	243	243
Chest X-Ray Images (Pneumonia) [17]	X-ray	Pneumonia, Normal	AP	0	5863

test sets. We found that using 60% of data for training the model and 40% for testing it achieved the best performance.

B. Experimental Environment

We used Keras to implement our COVID-19 detection models, which is an open-source library for deep learning applications written in Python [18]. The code was implemented by the Colab environment, a service hosted by Jupyter. The Colab environment provides free access to computing resources, including GPU, which is the most widely used computing technology in artificial intelligence [19]. Additionally, we used TensorFlow as the backend for the machine learning platform [20].

C. Stand-alone CNN Model

As part of our study, we developed a stand-alone CNN model to distinguish between COVID-19 and normal cases. Convolutional neural networks, also known as (CNN), are a specialized kind of neural network used in the computer vision field that contributes to automatic feature extraction and data processing with a known grid-like topology. The CNN architecture has three main layers: convolutional layer, pooling layer, and fully connected layer [21]. The main components of CNN architecture are illustrated in Fig. 2.

Convolutional layer : This layer applies several filters to the input to generate feature maps [21].

Pooling layer : This layer reduces the size of feature maps in terms of reducing the internal dimensions. Max pooling and average pooling are the two operations available in this layer [21].

Fully connected layer : This layer is also called the dense layer. In the fully connected layer, the inputs are connected to the output with a learnable weight and are assigned to the final outputs [21].

Activation function : This is a function used to facilitate knowledge of difficult and complex patterns. It includes sigmoid, tanh, and Rectified Linear Unit (ReLU). The most common is the corrected linear unit (ReLU) [21].

From 1989 to present, improvements have been made in the CNN architecture in terms of number of layers, parameters, and functions. These architectures vary from lightweight to heavyweight structures [23]. From among those architectures, we chose the Visual Geometry Group (VGG-16) architecture to implement our detection models based on information we
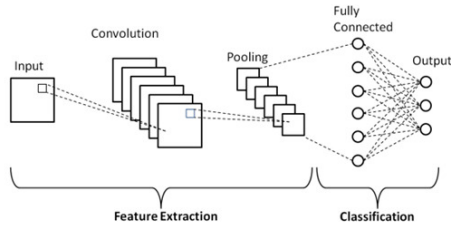
Fig. 2. Key Components of CNN Architecture [22].

TABLE III. PARAMETERS OF OUR STAND-ALONE CNN MODEL.

Number of training epochs	50
Batch size	26
Optimization	Adam
Learning rate	0.001
input shape	(224,224,3)
Number of Fully connected layers	2
Number of convolution layers	13
Number of pooling layers	5
Activation function in output layer	Sigmoid

garnered from the literature review. VGG architecture was proposed by Simonyan and Zisserman in 2014. It includes VGG-11, VGG-11-LRN, VGG-13, VGG-16, and VGG-19. The depth of the layers vary [24].

To develop our detection model, we exploited the principle of transfer learning. We used weights of a pre-trained VGG-16 that were trained on a large dataset called ImageNet. It learned a good representation of low level features like spatial, edges, rotation, lighting, shapes. These features can be integrated to enable the knowledge transfer and act as a feature extractor for new images in different computer vision problems [25].

In our model, we removed the top layer of the pre-trained model to train the model on a new chest X-ray images dataset. For optimization, we used Adam's algorithm, an effective stochastic optimization method for training deep learning models with a 0.001 learning rate. The number of training epochs was 50 and the batch size for each epoch was 26. The activation function in the hidden layer was ReLU, and sigmoid was used in the last layer because our classification problem is binary. The parameters of our stand-alone CNN model are illustrated in Table III and the model implementation steps are shown in Fig. 3.

D. Hybrid Models

In addition to the stand-alone CNN model, we have developed three hybrid models. Each model consists of a CNN architecture (VGG-16) for feature extraction and one of the following classification algorithms for classification: Support Vector Machine (SVM), Random Forests (RF), and Extreme Gradient Boosting (XGBoost). Fig. 4 shows an illustrative diagram of the proposed hybrid models.

A brief description of the selected classification algorithms follows:

Support Vector Machines (SVM): Essentially, SVM is a classification algorithm which tries to find the plane that separates the classes with the widest margin in the sample space in the most convenient way [26].

Random Forest (RF): RF is a collective classification and regression algorithm that uses decision trees as a classifier. Each decision tree is trained using a random data set derived from the original data set. The majority voting is used for the final classification [26].

Extreme Gradient Boosting (XGBoost): This is a boosting algorithm for classification and regression tree models, which is derived from the gradient lifting decision tree [27].
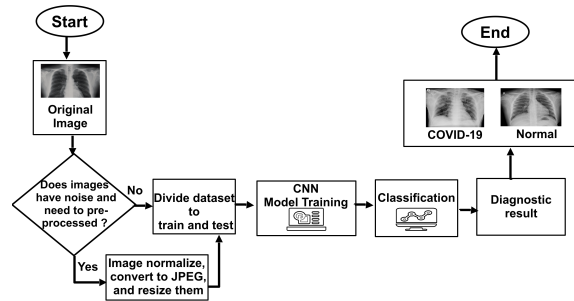

Fig. 3. Implementation Steps of the Stand-alone CNN Model.

We followed several steps to develop our hybrid models, which you can find illustrated in Fig. 5. First, we trained the VGG16 on our dataset. Then, we selected the last max pooling layer, which is the layer that comes after all convolution layers, to extract features. We added a flatten layer after the max pooling layer to handle the dimensionality issues, see Fig. 6. Because we use CNN 16 for the feature extraction part and not for classification, we discarded the fully connected and softmax layers, which are the dense layers after the flatten layer. After that, we used the extracted features to train the classification algorithms and develop models that can distinguish between COVID-19 and normal cases accurately. The number of extracted features is 25,088 features for every single image in the training dataset.

IV. RESULTS AND DISCUSSION

This section outlines the performance metrics we use to evaluate the developed models. Furthermore, it summarizes and discusses the main results of our experiments and presents a discussion related to previous studies.

A. Performance Metrics

We used various evaluation metrics to evaluate the proposed models. These metrics are as follows:

Confusion matrix: This is a technique that summarizes the performance of the classifier used. It presents true positive (TP) and true negative (TN) values, which means the number of correctly rated positive and negative instances. It also shows false positive (FP) and false negative (FN) values, which means the number of misclassified negative and positive instances [28].

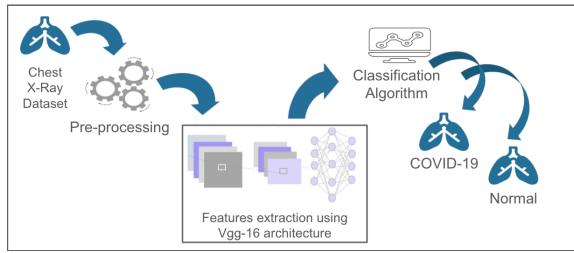**Accuracy:** Accuracy is the percentage of the test set

Fig. 4. An illustrative Diagram of the Proposed Hybrid Models.

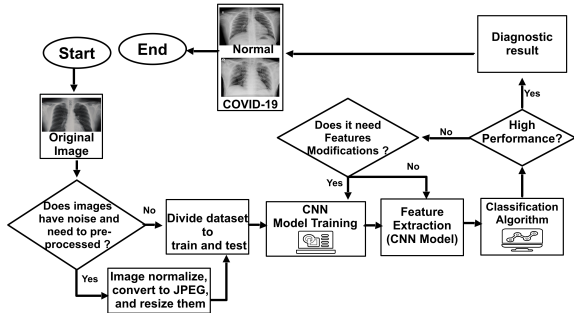Fig. 6. The Added Flatten Layer to the VGG-16 Architecture.

Fig. 5. Implementation Steps of the Hybrid CNN Models.

instances that are correctly classified by the classifier.

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

Sensitivity (Recall): Recall is also referred to as the true positive rate, which means the proportion of positive cases that are correctly identified.

$$Sensitivity = \frac{TP}{TP+FN} \qquad (2)$$

Specificity: This is the true negative rate. It refers to the proportion of negative instances that are correctly identified.

$$Specificity = \frac{TN}{TN+FP} \qquad (3)$$

Precision: Precision is defined as the proportion of the true positives against all the positive results (both true positives and false positives).

$$Precision = \frac{TP}{TP+FP} \qquad (4)$$

F1 score: It represents the harmonic mean between recall and precision values.

$$F1score = \frac{2*Precision*Recall}{Precision+Recall} \qquad (5)$$

B. Tests Results

In general, the results showed that all of our developed detection models are sufficient, especially the CNN and hybrid (CNN+SVM) models. This finding indicates that both stand-alone and hybrid models could achieve high performances when used to detect COVID-19 cases. Even though all our models are adequate, the CNN+XGBoost model is considered

the least effective. The values of its accuracy, sensitivity, specificity, precision, and F1-score were 0.9873%, 0.9928%, 0.9874%, 0.9821%, 0.9819%, respectively.

Furthermore, the sensitivity of three models (CNN, CNN+RF, and CNN+SVM) was 100%; however, the CNN+RF model achieved less specificity compared to the two other models. The specificity of the CNN and CNN+SVM models was 0.9964% and it was 0.9855% for the CNN+RF model. Just like the specificity, the accuracy, precision, and F1-score of the CNN and CNN+SVM models were better than those of the two other models. The accuracy, precision, and F1-score of the CNN and CNN+SVM models were 0.9982%, 0.9964%, and 0.9982%, respectively. Also, the accuracy, precision, and F1-score of the CNN+RF model were 0.9928%, 0.9857%, and 0.9929%, respectively. Fig. 7 shows the confusion matrices of the presented models, and Table IV-A summarizes their performances.

Overall, there is not a substantial difference in the performance of our models; all of them are excellent in terms of their ability to detect COVID-19. Besides the stand-alone CNN model, two of the hybrid models, the CNN+SVM and CNN+RF, classified 100% of COVID-19 cases correctly. Nonetheless, there are two main limitations to the generalization of our findings. The first one is the lack of available Covid-19 datasets. The second limitation is that this study focused only on one of the CNN techniques and three machine learning classifiers because of the following constraints: Authors used the Google Colab platform to perform experiments due to the unavailability of powerful computing resources. However, the restrictions in the provided resources that the platform applies prevented them from conducting further investigations. Additionally, this research was not financially supported, and thus researchers only had limited Internet data, and they used personal computers with bounded processing power and memory capacity.

TABLE IV. PERFORMANCE METRICS FOR THE PRESENTED MODELS

	CNN	CNN-RF	CNN-XGBoost	CNN-SVM
Accuracy	0.9982	0.9928	0.9873	0.9982
Sensitivity(Recall)	1.0000	1.0000	0.9928	1.0000
Specificity	0.9964	0.9855	0.9874	0.9964
Precision	0.9964	0.9857	0.9821	0.9964
F1-score	0.9982	0.9929	0.9819	0.9982

C. Discussion Related to Previous Studies

This subsection compares our models with pre-existing models presented in previous studies. These prior models
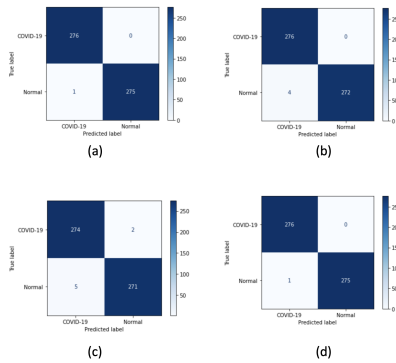
Fig. 7. Generated Confusion Matrices for a) Stand-alone CNN Model, b) CNN-RF model c) CNN-XGBoost Model, d) CNN-SVM Model

were developed using chest X-ray images and the VGG-16 model as a feature extractor or a classifier. Table V shows the performance for our models as compared to previous models.

In [8], the authors presented a modified CNN based model to detect COVID-19. The model accuracy and sensitivity before modification were 90.19% and 94.16%, respectively. After modifying the last layer of the original model, its performance improved. Its accuracy and sensitivity became 99.52% and 97.93%, respectively. Still, the accuracy and sensitivity of our models are higher. In [5], even though the researchers have used the CLAHE method to enhance their detection model, its performance is still low. Its accuracy was 83%, and then became 92%.

Compared to previous models, a better model has been presented in [3]. Its accuracy was 99.3% and its sensitivity was 99.28%. Our stand-alone model outperformed it. Some factors that lead to that improvement in ours is that the number of COVID-19 images that we used to develop our model is larger compared to these studies, and we have used a balanced dataset as well.

Just like the stand-alone CNN model, our hybrid models outperform those that have been proposed in [8]. Two out of the three classification algorithms that have been used in that paper are different from ours. Authors have selected the following algorithms: SVM, Bagging, and AdaBoost, while we have selected these classification algorithms: SVM, random forest, and XGBoost. As shown in Table V, the SVM is the classifier that leads to developing models with high accuracy in both studies. Our two other algorithms, however, surpassed Bagging and AdaBoost.

V. CONCLUSION

This study addressed some of the challenges of the traditional COVID-19 test method. It exploited the power of machine learning to accelerate the process of detecting COVID-19 and to enhance its efficiency. We investigated the effectiveness of stand-alone CNN models and hybrid machine learning models in detecting the disease. We combined five chest X-Ray images datasets to develop four COVID-19 detection models: a stand-alone CNN model and three hybrid machine learning models. As a comparison to some of the previous studies that have been published in the early few months of the pandemic,

the count of chest X-Ray images that we used to develop our models is considered one of the biggest.

TABLE V. COMPARISON WITH STATE-OF-THE-ART MODELS

Ref.	Model	Accuracy	Sensitivity	Specificity	Precision	F1-Score
[5]	VGG-16	92%	94.92%	92%	-	-
[3]	VGG-16	99.33%	99.28%	99.38%	-	99.28%
[8]	VGG-16	90.19%	94.16%	-	94.12%	94.14%
	VGG-16 & SVM	91.19%	93.15%	-	94.12%	93.63%
	VGG-16 & Baggimg	90.19%	92.16%	-	92.22%	92.19%
	VGG-16 & Ada Boot	90.19%	90.10%	-	89.16%	89.63%
Our study	VGG-16	99.82%	100%	99.64%	99.64%	99.82%
	VGG-16 & RF	99.28%	100%	98.55%	98.57%	99.29%
	VGG-16 & XGBoost	98.73%	99.28%	98.74%	98.21%	98.19%
	VGG-16 & SVM	99.82%	100%	99.64%	99.64%	99.82%

Our findings illustrate that all of the four proposed models are effective in detecting COVID-19. The lowest detection accuracy obtained was 98.73% which is the accuracy of the VGG16+XGBoost model. The highest accuracy was 99.82% which is the accuracy of both VGG-16 and(VGG16+SVM) models. Furthermore, one of the most promising findings is that the sensitivity of the VGG-16, (VGG16+SVM), and (VGG16+RF) models is 100%, meaning they have a zero false negative case rate. That means that from all the examined cases, 100% of the COVID-19 positive images were detected. This finding plays an important role in reducing the possibility of spreading the virus to more people.

VI. FUTURE WORK

Several experiments can be conducted to study the effect of different CNN architectures and optimizers on the model's performance. Furthermore, one of the areas that warrants additional study are COVID-19 mutations. It is necessary to develop machine learning and deep learning-based models capable of detecting new versions of COVID-19, such as B.1.1.7, B.1.1.207, P.1 and B.1.525 automatically. Additionally, it is important to build robust models that have the ability to distinguish between SARS, MERS, and COVID-19 accurately.

REFERENCES

[1] Z. Y. Zu, M. D. Jiang, P. P. Xu, W. Chen, Q. Q. Ni, G. M. Lu, and L. J. Zhang, "Coronavirus disease 2019 (COVID-19): A perspective from china," *Radiology*, vol. 296, no. 2, pp. E15–E25, 2020.

[2] L. Duran-Lopez, J. P. Dominguez-Morales, J. Corral-Jaime, S. Vicente-Diaz, and A. Linares-Barranco, "Covid-xnet: a custom deep learning system to diagnose and locate covid-19 in chest x-ray images," *Applied Sciences*, vol. 10, no. 16, p. 5683, 2020.

[3] I. U. Khan and N. Aslam, "A deep-learning-based framework for automated diagnosis of covid-19 using x-ray images," *Information*, vol. 11, no. 9, p. 419, 2020.

[4] L. O. Hall, R. Paul, D. B. Goldgof, and G. M. Goldgof, "Finding covid-19 from chest x-rays using deep learning on a small dataset," *arXiv preprint arXiv:2004.02060*, 2020.

[5] F. A. Saiz and I. Barandiaran, "Covid-19 detection in chest x-ray images using a deep learning approach," *International Journal of Interactive Multimedia and Artificial Intelligence, InPress (InPress)*, vol. 1, 2020.

[6] S. Vaid, R. Kalantar, and M. Bhandari, "Deep learning covid-19 detection bias: accuracy through artificial intelligence," *International Orthopaedics*, vol. 44, pp. 1539–1542, 2020.

[7] K. F. Haque and A. Abdelgawad, "A deep learning approach to detect covid-19 patients from chest x-ray images," *AI*, vol. 1, no. 3, pp. 418–435, 2020.

[8] R. K. Dubey *et al.*, "Deep learning based hybrid models for prediction of covid-19 using chest x-ray," *TechRxiv*, 2020.

[9] K. Elasnaoui and Y. Chawki, "Using x-ray images and deep learning for automated detection of coronavirus disease," *Journal of Biomolecular Structure and Dynamics*, no. just-accepted, pp. 1–22, 2020.

[10] 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*, p. 1, 2020.

[11] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," *arXiv preprint arXiv:2004.09363*, 2020.

[12] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[13] J. P. Cohen, P. Morrison, and L. Dao, "COVID-19 Image Data Collection," *arXiv e-prints*, p. arXiv:2003.11597, Mar. 2020.

[14] "Actualmed covid-19 chest x-ray dataset," 2020. Accessed on: Feb. 13, 2021. [Online]. Available: https://github.com/agchung/Actualmed-COVID-chestxray-dataset.

[15] "Figure1-covid-chestxray-dataset," 2020. Accessed on: Feb. 13, 2021. [Online]. Available: https://github.com/agchung/Figure1-COVID-chestxray-dataset.

[16] "Covid-19-image-repository," 2020. Accessed on: Feb. 13, 2021. [Online]. Available: https://github.com/ml-workgroup/covid-19-image-repository.

[17] P. Mooney, "Chest x-ray images (pneumonia)," 2020. Accessed on: Feb. 13, 2021. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia.

[18] "Keras documentation," accessed on: February. 20, 2021. [Online]. Available: https://keras.io/about/.

[19] "Google colaboratory," accessed on: February. 20, 2021. [Online]. Available: https://colab.research.google.com/notebooks/intro.ipynb.

[20] "Tensorflow," accessed on: February. 20, 2021. [Online]. Available: https://github.com/tensorflow/tensorflow.

[21] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[22] V, Phung and E, Rhee , "A high-accuracy model average ensemble of convolutional neural networks for classification of cloud image patches on small datasets," *Appl. Sci. (Basel)*, vol. 9, no. 21, p. 4500, 2019.

[23] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, 2020.

[24] Z. Li, W. Yang, S. Peng, and F. Liu, "A survey of convolutional neural networks: analysis, applications, and prospects," *arXiv preprint arXiv:2004.02806*, 2020.

[25] S. Tammina, "Transfer learning using vgg-16 with deep convolutional neural network for classifying images," *International Journal of Scientific and Research Publications*, vol. 9, no. 10, pp. 143–150, 2019.

[26] P. Mooney, "Classification algorithm - an overview — sciencedirect topics," 2021. Accessed on: April. 28, 2021. [Online]. Available: http://www.sciencedirect.com/topics/engineering/classification-algorithm.

[27] O. D. Gülgün and E. Hamza, "Medical image classification with hybrid convolutional neural network models," *Bilgisayar Bilimleri ve Teknolojileri Dergisi*, vol. 1, no. 1, pp. 28–41.

[28] M. Hossin and M. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, p. 1, 2015.