**ORIGINAL RESEARCH**

# Hybrid classification structures for automatic COVID-19 detection

Mohamed R. Shoaib[1] · Heba M. Emara[1] · Mohamed Elwekeil[1,4] · Walid El-Shafai[1,2] · Taha E. Taha[1] ·
Adel S. El-Fishawy[1] · El-Sayed M. El-Rabaie[1] · Fathi E. Abd El-Samie[1,3]

## Abstract

This paper explores the issue of COVID-19 detection from X-ray images. X-ray images, in general, suffer from low quality and low resolution. That is why the detection of different diseases from X-ray images requires sophisticated algorithms. First of all, machine learning (ML) is adopted on the features extracted manually from the X-ray images. Twelve classifiers are compared for this task. Simulation results reveal the superiority of Gaussian process (GP) and random forest (RF) classifiers. To extend the feasibility of this study, we have modified the feature extraction strategy to give deep features. Four pre-trained models, namely ResNet50, ResNet101, Inception-v3 and InceptionResnet-v2 are adopted in this study. Simulation results prove that InceptionResnet-v2 and ResNet101 with GP classifier achieve the best performance. Moreover, transfer learning (TL) is also introduced in this paper to enhance the COVID-19 detection process. The selected classification hierarchy is also compared with a convolutional neural network (CNN) model built from scratch to prove its quality of classification. Simulation results prove that deep features and TL methods provide the best performance that reached 100% for accuracy.

**Keywords** Coronavirus · Chest X-ray radiographs · Transfer learning · Deep feature extraction

## 1 Introduction

COVID-19 infection has appeared in Wuhan, China since December 2019. It is now considered a worldwide pandemic (Roosa et al. 2020; Yan et al. 2020). It may severely affect the human respiratory system. While COVID-19 causes mild symptoms in about 82% of the cases, the remaining cases suffer from fatal symptoms and some may need ventilators (National Geographic n.d.). Common COVID-19 infection signs include respiratory symptoms, fever, lowness of breath,

and respiratory difficulties. In severe cases, COVID-19 may cause pneumonia, severe acute respiratory syndrome and kidney failure leading to death (Stoecklin et al. 2020). Dealing with people who suffer from respiratory symptoms, such as coughing and sneezing, should be avoided. While in many countries of the world, due to the rising needs for intensive care units, the health system has been overburdened and may be on the verge of collapsing. Then, for COVID-19 detection, an automatic diagnosis system is required.

COVID-19 can be diagnosed with several methods, like reverse transcriptase-polymerase chain reaction (RT-PCR),

✉ Heba M. Emara
hebam9909@gmail.com

✉ Mohamed Elwekeil
mohamed.elwekeil@el-eng.menofia.edu.eg

Mohamed R. Shoaib
mohammed_reda1996@outlook.com

Walid El-Shafai
eng.waled.elshafai@gmail.com

Taha E. Taha
taha117@hotmail.com

Adel S. El-Fishawy
aelfishawy@hotmail.com

El-Sayed M. El-Rabaie
Srabie1@yahoo.com

Fathi E. Abd El-Samie
fathi_sayed@yahoo.com

1 Department of Electronics and Electrical Communications Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

2 Security Engineering Lab, Computer Science Department, Prince Sultan University, Riyadh 11586, Saudi Arabia

3 Department of Information Technology, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

4 Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, Cassino, Italy

blood testing, and medical image analysis (Huang et al. 2020). Although RT-PCR testing is very specific, it is a time-consuming, difficult, and complicated manual technique. Hence, different modalities of medical imaging can be used for the task of COVID-19 detection. Although CT scanning is the most accurate and effective tool for COVID-19 detection, X-ray imaging is still the best tool, because it is cheap and fast. COVID-19 infection is reflected in X-ray images with a glassy nature. Hence, there is a need for accurate inspection of X-ray images to detect COVID-19 cases.

Dependence on human operators with this high rate of infection may be infeasible due to the limited number of trained specialists and the need to force safety precautions. That is why artificial intelligence (AI) finds a major role in this task. Both machine learning (ML) and deep learning (DL) tools are required in the diagnosis process (Tahir et al. 2020; Chowdhury et al. 2019a, b; Kallianos et al. 2019).

Several attempts have been presented in the literature for COVID-19 detection from X-ray images. CNNs have found good applications for this task. The reason is that multiple convolutional layers can lead to different feature maps through the utilization of a variety of convolutional kernels. The most effective features can be kept through the utilization of a pooling strategy (Zhang et al. 2018; Sun et al. 2019; Gheisari et al. 2017).

The concept of transfer learning (TL) has been investigated deeply in this area. The rationale behind this trend is the small size of the available dataset and the large cost of the training process. The idea of TL depends on the utilization of a pre-trained network and the application of fine tuning for the task of interest. TL has led to good classification results in different applications (Emara et al. 2021a, b). Hence, it is recommended in this paper.

The presented topic of research is very important to speed up the diagnosis process of COVID-19, efficiently. We begin this study by investigating traditional ML algorithms and a CNN model built from scratch to allow classification of X-ray images acquired for suspected COVID-19 patients. Our numerical results reveal the low accuracy of traditional ML algorithms that depend on manual feature extraction. In addition, building a CNN from scratch is not sufficient with a large burden of training and optimization for achieving the required performance level. Hence, our main contribution in this paper is to make use of the TL strategy in the classification task. In this strategy, well-trained deep convolutional neural networks (DCNNs) are tuned to the task of interest. Another contribution in this paper is to perform the feature extraction through the large DCNNs such as ResNet50, ResNet101, Inception-v3 and InceptionResnet-v2, while the classification task is performed with random forest (RF) and Gaussian process (GP) classifiers. These classifiers are reported in previous studies as superior classifiers. This gives good classification performance. The main contributions of this work are summarized as follows:

- Developing different ML and DL algorithms for exploring the issue of COVID-19 detection from X-ray images.
- Summarizing the most recent related work concerned with classification challenges of medical images.
- Achieving higher classification and detection accuracies for identifying COVID-19 cases.
- Presenting a detailed comparative study between the proposed work and the most recent related studies.

## 2 Related works

In the literature, researchers developed several algorithms for COVID-19 detection. A large amount of data is a vital demand required to train various deep learning models. At the beginning of the pandemic, the available data was limited. In order to solve this problem, TL has been introduced.

Pham (2021) introduced a TL-based algorithm for COVID-19 detection. Its results revealed the importance of fine tuning. This can save time and cost by avoiding the development of more complicated models that produce the same or better results. Narin et al. (2020) proposed a TL-based algorithm for COVID-19 detection based on an inception model. This algorithm has been evaluated on 1065 CT images. An accuracy of 79.3% has been reported. Saiz and Barandiaran (2020) utilized the VGG16 pre-trained model for the detection process. Their algorithm was evaluated on 1500 X-ray images. An accuracy of 94.92% has been obtained. In Wang and Wong (2021), ResNet50 model was utilized for COVID-19 detection. Their algorithm was evaluated on 100 X-ray images. An accuracy of 98% has been achieved. Erdem and Aydın (2021) introduced a comparison between the pre-trained models, namely Inception-v3, MobileNet, SqueezeNet, Xception, and VGG16, to get the best performance. Their results revealed that Inception-v3 model has the highest accuracy that reaches 90%. Jain et al. (2021) utilized Xception, Inception-v3, and ResNet pre-trained models for COVID-19 diagnosis. These models were evaluated on 6432 X-ray images. Their results revealed that the Xception model presents the highest accuracy that reaches 97.97%.

Image pre-processing techniques have a vital role in the enhancement of the classification process. El-Shafai et al. (2021a) introduced an algorithm for segmentation and classification of COVID-19 images based on DL. Firstly, the classification process is employed to differentiate between COVID-19 and pneumonia images with the CNN model. Then, the segmentation process is applied on the COVID-19 and pneumonia images. Finally, the obtained segmented images are used to determine the infected regions in COVID-19 and pneumonia images. El-Shafai et al. (2021b)

introduced an automatic algorithm for image enhancement and classification. To get high-resolution versions of X-ray and CT images, their paper presented a hybrid SIGTra model. A generative adversarial network (GAN) has been used for the image super-resolution reconstruction purpose. In addition, for image classification, TL with CNN (TCNN) has been used. An accuracy of 99% for X-ray image classification has been achieved. Canayaz (2021) utilized meta-heuristic-based feature selection for COVID-19 detection. Firstly, an image contrast enhancement algorithm is used for pre-processing. Then, the features are extracted using different pre-trained models. Feature selection is implemented depending on metaheuristic algorithms. Finally, the obtained features are classified using a support vector machine (SVM). An accuracy of 99.83% was obtained.

A combination of different CNN models was presented to get higher detection accuracy. Xu et al. (2020) utilized a location-attention network and a ResNet18 model for COVID-19 detection. Their algorithm has been evaluated on 618 CT images for COVID-19, viral pneumonia, and normal cases. Their algorithm reported an accuracy of 86.7%. Karar et al. (2021) introduced cascaded DL models to increase the efficiency of COVID-19 detection. Eleven pre-trained models were exploited and compared for classification purposes. According to their results, the VGG16, ResNet50, and DenseNet169 models achieve the best detection accuracy. Emara et al. (2021b) used CNN models with various learning procedures for COVID-19 diagnosis. Firstly, a CNN-based TL algorithm was used to automatically diagnose COVID-19 from X-ray images with various training and testing ratios. The second task was to train the CNN model from scratch. Their results indicate that training of the TL-based CNN models produces high performance. Wang et al. (2021) presented a modified inception model, followed by internal and external validation for COVID-19 detection. Their model has been evaluated on 1065 CT images for COVID-19 and viral pneumonia cases. The accuracy of the internal validation was 89.5%, and the accuracy of the external validation reaches 79.3%. Song et al. (2021) proved that identifying possible lesions from CT images may be useful for COVID-19 detection. The feature pyramid network (FPN) was combined with the ResNet50 model. That ResNet50 was used to extract local and relational features. The global features extracted from the original image are concatenated with these features. The classification process is carried out using a multi-layer perceptron (MLP). The sensitivity of this model was 96%.

Several researchers implemented AI and heuristic optimization algorithms such as genetic algorithms (GA) in this topic. Mimetic genetic algorithms (MGAs) were exploited to solve several problems such as network optimization, vehicle routing, several graph theory and electronic manufacturing units. Roy et al. (2019) introduced MGAs to solve

the traveling salesman problem (TSP). Boltzmann probabilistic selection and parents crossover were combined with the ergodic mutation. The cost and distance are compared for the adjacent nodes of the involved parents. Standard benchmarks were obtained from TSP versus classical genetic algorithms (GAs). Biswas and Pal (2019) presented a fuzzy goal programming (FGP) method based on GA. In order to solve the congestion management (CM) problem, membership functions are converted into membership goals. The GA computational scheme achieves the required goals according to their priorities. Li et al. (2020) proposed a DCNN model for COVID-19 detection that is called COVNet. Their model was tested on CT images collected from six hospitals. It achieved an accuracy of 96%. Using the COVID-19 chest X-ray dataset, Ghoshal and Tucker (2020) used drop-weights-based Bayesian convolutional neural networks (BCNNs) to compute uncertainty in DL models to increase the diagnostic performance. An accuracy of 89.92% has been reported. Wang et al. (2021) proposed a DCNN model for COVID-19 detection called COVID-Net. Their algorithm was tested on a collection of 16,756 X-ray images from 13,645 cases obtained from two open access data sources. An accuracy of 92.4% has been reported.

The main advantage of DCNN models is the automatic feature extraction. DCNN models can be used as feature extractors followed by traditional ML models for the classification process. Loey et al. (2021) presented an algorithm for COVID-19 detection. Their algorithm includes two stages. The first stage depends on ResNet50 for feature extraction. The second stage depends on decision tree (DT), SVM, and ensemble algorithm for classification. Their results revealed that the SVM classifier outperforms the other ML algorithms, and achieves an accuracy of 99.64%. Wu et al. (2020) presented an ML algorithm for COVID-19 detection from blood tests. Random forest (RF) classifier allows discrimination from 49 blood tests. An accuracy of 95.95% has been obtained. Rahman et al. (2020) built a data-driven dynamic clustering method to mitigate the COVID-19 negative impact on the economy. Their method have mainly three components: data analysis, dynamic clustering, and data security. A clustering technique has been presented, and it has been simulated in four scenarios to reveal its benefits and drawbacks. In the lock-down coverage experiment, the presented clustering method improved the performance indicators by 60–80%.
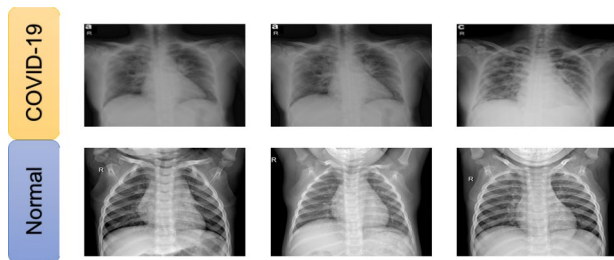
**Fig. 1** X-ray images for COVID-19 and normal cases (Alqudah and Qazan 2020)

# 3 Materials and methods

The used chest X-ray image dataset includes 912 X-ray images for regular people, and 912 X-ray images for COVID-19 infected people. It was presented in (Alqudah and Qazan 2020). Figure 1 shows different samples from this dataset that were used to test the proposed models for COVID-19 detection (Fig. 2).

The proposed approach is presented in Fig. 2. We have investigated traditional ML algorithms for COVID-19 detection. Firstly, traditional ML algorithms with manual feature extraction have been studied. After that, a 15-layer CNN model built from scratch has been investigated for efficient classification of COVID-19 cases. In addition, a TL strategy has been exploited. Different DCNNs such as ResNet50, ResNet101, Inception-v3 and InceptionResnet-v2 have been tuned to the task of interest. Finally, feature extraction is performed through DCNNs, while the classification task is implemented with traditional ML classifiers.

## 3.1 Convolution neural network trained from scratch

The principal structure of a CNN network is a combination of convolution, batch normalization (BN), and pooling layers. The BN layers are used to normalize the local features once the convolution layers have retrieved them. Pooling layers are used to minimize the number of features extracted. To reflect the variations in local activity levels, max-pooling is used. It reveals the details of the edges. The largest values observed primarily correspond to edges. X-ray images contain many details. The representation of the output feature map is illustrated as follows (Bhandary et al. 2020; Bosch et al. 2007; Cheng and Bao 2014):

$$Y_j^l = f\left(\sum_{i \in N_j} Y_i^{l-1} * X_{ij}^{\ l} + b_j^{\ l}\right) \tag{1}$$

where $Y_j^l$ indicates the local features collected from the preceding layers, and $X_{ij}$ refers to the adjustable kernels. The bias is employed to prevent over-fitting, and it is represented by the symbol $b_j^l$. The pooling process is represented in Eq. (2)

$$Y_j^l = down(Y_j^{l-1}) \tag{2}$$

The down-sampling function is represented by $down(.)$. All activations in the previous layer are connected directly to the fully-connected (FC) layer. The FC layer adds discriminative features to the input image, allowing it to be classified into different classes.

**Fig. 2** Block diagram of the proposed approaches for COVID-19 detection
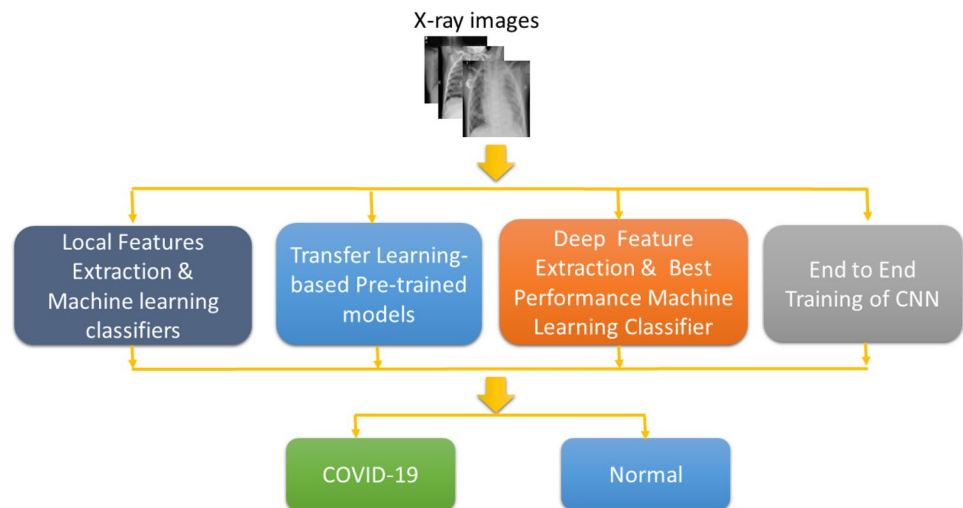
## 3.2 Local feature extraction and machine learning classifiers

First of all, X-ray images are resized and converted from gray-scale into HVS images. Then, mean, standard deviation (std), skewness, kurtosis, histogram minimum and histogram maximum are estimated for H, V and S channels. These features are used as inputs for the ML models. To the best of our knowledge, this set of features has never been considered in the literature for ML-based COVID-19 detection. Different ML models have been evaluated for classification. Specifically, we assessed logistic regression (LR), K-nearest neighbours (KNN), SVM, naive Bayes (NB), DT, RF, gradient boosting (GB), stochastic gradient descent (SGD), GP, MLP, adaptive boosting (AdaBoost) and extreme gradient boosting (XGBoost) classifiers.

To verify the effectiveness of the proposed approach, extraction of the adequate features from the input images is required. The mean, std, skewness, kurtosis, histogram minimum and histogram maximum are considered. Higher-order statistics such as skewness, and kurtosis (Groeneveld and Meeden 1984) are utilized for classifying X-ray images. The use of these statistics is inspired by the fact that distribution of the samples of a dataset is often characterized by their level of dispersion and asymmetry. For an $N$-point data sample sequence, $X = x_1, x_2, \ldots, x_N$, the corresponding skewness $\beta_1$, and kurtosis $\beta_2$ are calculated as:

$$\beta_1 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{x_i - \mu}{\sigma} \right)^3 \qquad (3)$$

$$\beta_2 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{x_i - \mu}{\sigma} \right)^4 \qquad (4)$$

where $\mu$ is the mean of the data and $\sigma$ is the standard deviation. The second-, third-, and fourth-order moments are used to calculate the skewness and kurtosis.

In Fig. 3, the corresponding values of histogram minima, histogram maxima, mean, and std are also displayed for H, V, and S channels. It is important to note that the histograms
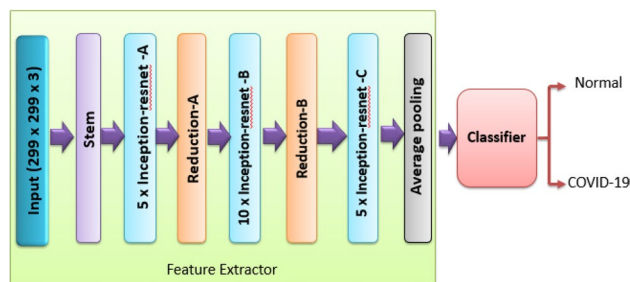


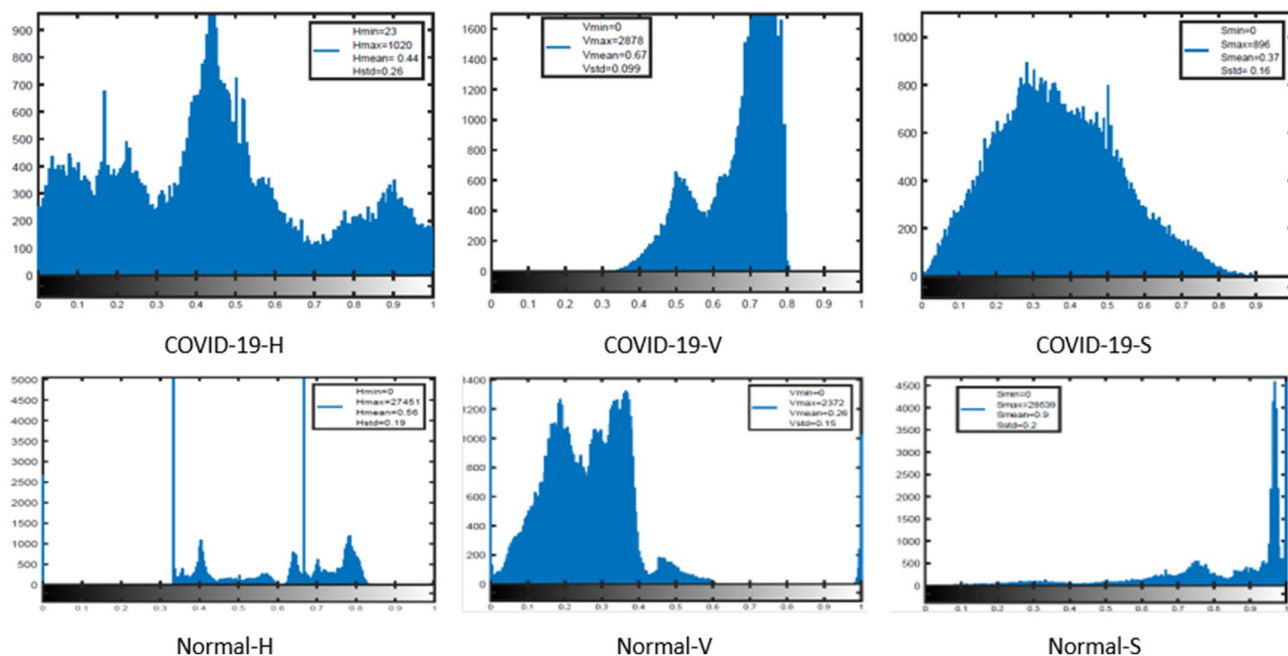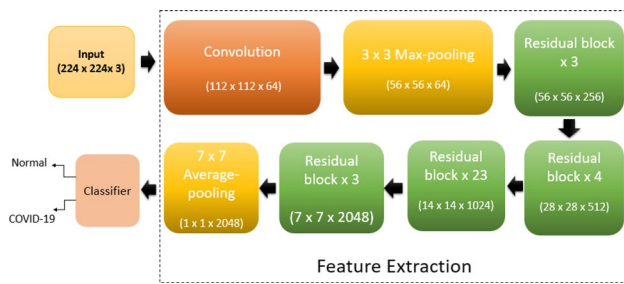**Fig. 4** Deep feature extraction architecture with modified Inception-Resnet-v2



**Fig. 3** Histograms of H, V and S channels for COVID-19 and normal cases

**Fig. 5** Deep feature extraction architecture with modified ResNet101 model

of the X-ray images are different for each case. The associated histogram minima, histogram maxima, mean, std, skewness, and kurtosis values are all different, and these variables are representatives of dataset dispersion, asymmetry, and peakedness. As a result, it is logical to assume that these statistical metrics are more effective for classifying X-ray images.

## 3.3 Deep feature extraction with machine learning (ML) classifiers

An ML classifier is used instead of the DL classifier, because DL classifiers require a large dataset for training and validation. The deep features of the pooling layer are retrieved and fed into the ML classifier. ResNet50, ResNet101, Inception-v3, and InceptionResnet-v2 are employed as pre-trained CNN models. Furthermore, the GP classifier is employed and compared with the RF classifier.

Figures 4 and 5 present the architecture for the deep feature extraction with the modified InceptionResnet-v2 and ResNet101 models, respectively.

## 3.4 Transfer-learning-based pre-trained models

Deep learning from scratch is a time-consuming process that requires data labeling and splitting. TL is appropriate for removing the huge strain of this task. In TL, small changes in deep pre-trained networks are induced in response to input data. The pre-trained models are loaded, and the BN, ReLU, and softmax layers are used in place of the last three FC layers. The models are trained with a learning rate of 0.00001 with 6 epochs. The final aim of the proposed approach is to use a tuned pre-trained model to classify image batches into COVID-19 or normal cases. The block diagram of the TL-based model for the COVID-19 detection is presented in Fig. 6.
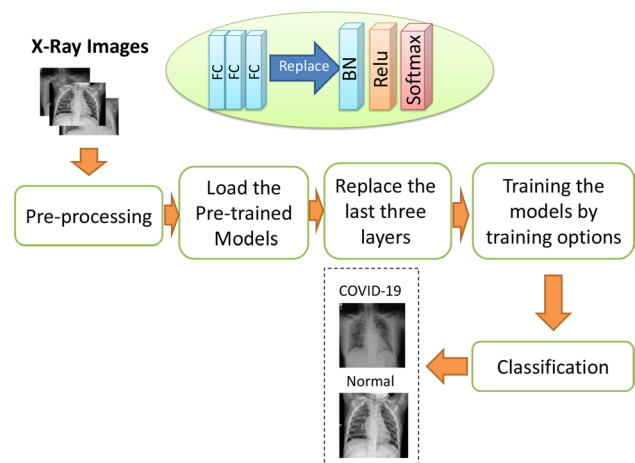


**Fig. 6** Block diagram of the proposed pre-training-based TL model

## 3.5 Performance metrics

Standard metrics like accuracy (*ACC*), sensitivity (*SEN*), specificity (*SPEC*), precision (*Preci*), mis-classification rate ($M_r$), and false positive rate ($F_{pr}$) are used to evaluate the proposed model (Sokolova and Lapalme 2009). The number of correctly identified abnormal cases is known as true positive ($T_p$). The number of accurately identified normal cases ($T_n$) gives the true negative. A set of normal cases categorized as anomaly diagnoses represents false positive ($F_p$). The false negative ($F_n$) represents the collection of abnormalities seem to be normal. Sensitivity is given as:

$$SEN = \frac{T_p}{T_p + F_n} \times 100 \tag{5}$$

Specificity is given as:

$$SPEC = \frac{T_n}{T_n + F_p} \times 100 \tag{6}$$

Accuracy is given as:

$$ACC = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \tag{7}$$

Precision is given as:

$$Preci = \frac{T_p}{T_p + F_p} \tag{8}$$

The misclassification rate is given as:

$$M_r = \frac{F_p + F_n}{T_p + T_n + F_p + F_n} \tag{9}$$

**Table 1** Architecture of the proposed CNN model for COVID-19 detection

| | | |
|---|---|---|
| 1 | "input" | $224 \times 224 \times 3$ images |
| 2 | "conv-1" | 16 $3 \times 3$ convolutions with stride 1 and padding "same" |
| 3 | "BN-1" | Batch normalization |
| 4 | "relu-1" | ReLU |
| 5 | "pool-1" | $2 \times 2$ max pooling with stride 2 |
| 6 | "conv-2" | 32 $3 \times 3$ convolutions with stride 1 and padding "same" |
| 7 | "BN-2" | Batch normalization |
| 8 | "relu-2" | ReLU |
| 9 | "pool-2" | $2 \times 2$ max pooling with stride 2 |
| 10 | "conv-3" | 64 $3 \times 3$ convolutions with stride 1 and padding "same" |
| 11 | "BN-3" | Batch normalization |
| 12 | "relu-3" | ReLU |
| 13 | "pool-3" | $2 \times 2$ max pooling with stride 2 |
| 14 | "fc" | Three fully-connected layers |
| 15 | "softmax" | Softmax |
| 16 | "Classification" | Crossentropyex |

**Fig. 7** Training progress of the CNN model

False positive rate is given as:

$$F_{pr} = \frac{F_p}{T_n + F_p} \tag{10}$$

## 4 Experimental results

The investigated ML models, TL-based pre-trained models, deep feature models, and the model trained from scratch, are tested using (Alqudah and Qazan 2020) dataset.

## 4.1 Results for the first approach

For the first approach, a simple CNN model was built and trained from scratch. As demonstrated in Table 1, the proposed CNN model is discussed. The conv-1, conv-2, and conv-3 layers have 16, 32, and 64 filters, respectively, that have $3 \times 3$ pixels in size. The max-pooling function

is employed for dimensionality reduction. The network is trained from scratch using the Adam optimizer, with a learning rate of 0.00001. Figure 7 displays the performance of the CNN trained from scratch in terms of both accuracy and loss. There is a coincidence in performance between validation and training accuracy as well as validation and training loss. The minimal square error (MSE) has been chosen as



**Fig. 8** Confusion matrix and ROC curve for the CNN model trained from scratch
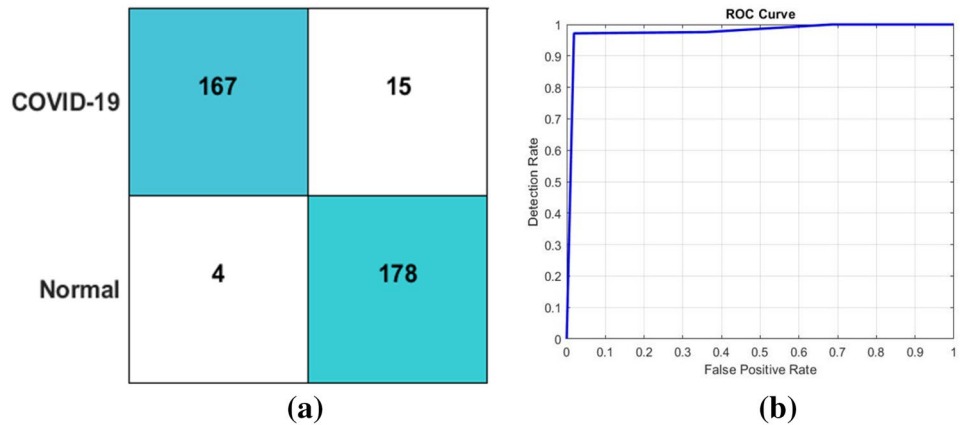
**Table 2** Detection performance results obtained from different ML models

| Models | Evaluation metrics | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| | ACC% | SEN% | SPEC% | Preci% | Fpr | Mr |
| LR | 84.11 | 84.62 | 83.61 | 83.70 | 0.159 | 0.164 |
| KNN | 95.34 | 96.65 | 94.09 | 94.02 | 0.0466 | 0.0591 |
| SVM | 84.11 | 84.24 | 83.98 | 84.24 | 0.1589 | 0.1602 |
| NB | 67.67 | 63.98 | 74.42 | 82.07 | 0.3233 | 0.2558 |
| DT | 89.86 | 92 | 87.9 | 87.5 | 0.1014 | 0.1211 |
| RF | 97.53 | 96.30 | 98.86 | 98.91 | 0.0247 | 0.0114 |
| GB | 66.30 | 62.66 | 73.39 | 82.07 | 0.3370 | 0.2661 |
| SGD | 81.64 | 77.47 | 87.5 | 89.67 | 0.1836 | 0.125 |
| GP | 97.53 | 97.81 | 97.25 | 97.28 | 0.0247 | 0.0275 |
| MLP | 96.16 | 95.21 | 97.18 | 97.28 | 0.0384 | 0.0283 |
| AdaBoost | 65.48 | 62.39 | 70.99 | 79.35 | 0.3452 | 0.2901 |
| XGBoost | 87.12 | 90.53 | 84.18 | 83.15 | 0.1288 | 0.1582 |
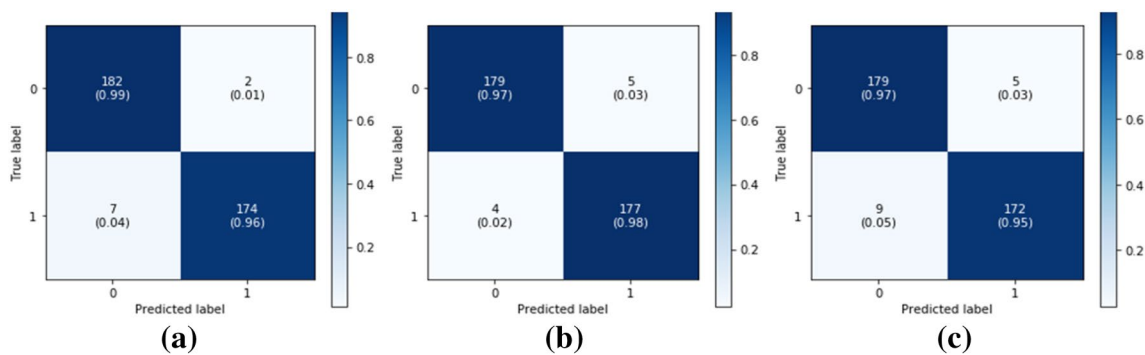


**Fig. 9** Confusion matrices for the best ML models used for COVID-19 detection **a** RF classifier, **b** GP classifier and **c** MLP classifier
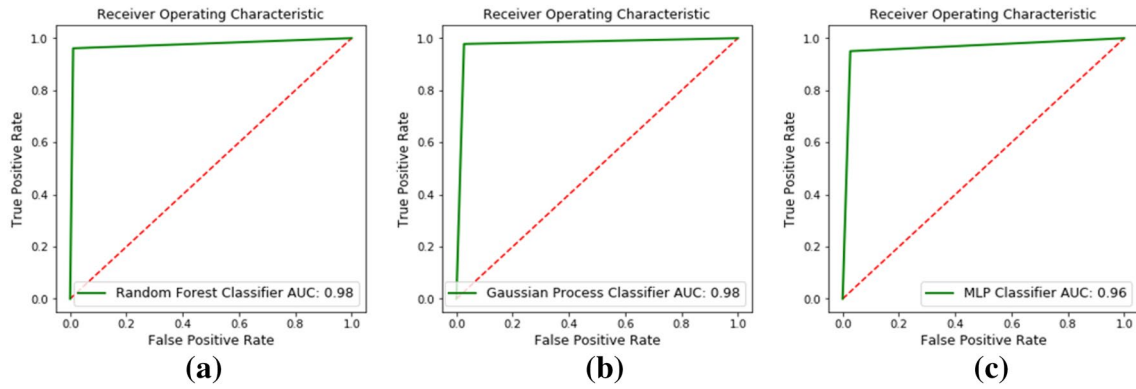
**Fig. 10** ROC curves for the best performance ML models used for COVID-19 detection **a** RF classifier, **b** GP classifier and **c** MLP classifier

**Table 3** Detection performance results obtained from different pre-trained models with GP classifier

| Model | Evaluation metrics | | | | | |
|---|---|---|---|---|---|---|
| | ACC% | SEN% | SPEC% | Preci% | Fpr | Mr |
| InceptionResnet-v2 | 100 | 100 | 100 | 100 | 0.0 | 0.0 |
| Inception-v3 | 99.21 | 99.17 | 99.25 | 99.25 | 0.0082 | 0.0078 |
| ResNet50 | 99.60 | 100 | 100 | 99.25 | 0.0 | 0.0039 |
| ResNet101 | 100 | 100 | 100 | 100 | 0.0 | 0.0 |

**Table 4** Detection performance results obtained from different pre-trained models with RF classifier

| Model | Evaluation metrics | | | | | |
|---|---|---|---|---|---|---|
| | ACC% | SEN% | SPEC% | Preci% | Fpr | Mr |
| InceptionResnet-v2 | 97.65 | 100 | 95.55 | 100 | 0.0 | 0.0234 |
| Inception-v3 | 96.875 | 97.17 | 97.52 | 97.74 | 0.0247 | 0.03125 |
| ResNet50 | 97.65 | 97.52 | 97.52 | 97.77 | 0.024 | 0.0234 |
| ResNet101 | 98.04 | 97.52 | 98.51 | 97.79 | 0.0247 | 0.0195 |

**Table 5** Detection performance results obtained for TL-based pre-trained models

| Model | Evaluation metrics | | | | | |
|---|---|---|---|---|---|---|
| | ACC% | SEN% | SPEC% | Preci% | Mr | Fpr |
| InceptionResnet-v2 | 100 | 100 | 100 | 100 | 0.0 | 0.0 |
| Inception-v3 | 99.45 | 98.90 | 100 | 100 | 0.055 | 0.0 |
| ResNet50 | 98.13 | 97.80 | 100 | 98.78 | 0.0110 | 0.0 |
| ResNet101 | 100 | 100 | 100 | 100 | 0.0 | 0.0 |

the loss function. Based on the MSE, a distance minimization approach is used. The confusion matrix and ROC curve for the trained model from scratch are shown in Fig. 8. The CNN model reports an accuracy of 94.78% and an $F_{pr}$ of 0.0522.

## 4.2 Results for the second approach

The results are presented in Table 2. It is clear that the best models for detection are RF, GP, MLP and KNN with accuracy levels up to 97.53%, 97.53%, 96.16% and 95.34%,

respectively, with a 80/20 training/testing ratio. Moreover, the NB, GB and AdaBoost classifiers achieve low accuracies of about 67.67%, 67.88% and 65.48%, respectively.

Figures 9 and 10 present confusion matrices and ROC curves for MLP, RF and GP classifiers, which give higher performance than other models. AUCs of 98%, 98%, 96% and 95% are obtained from RF, GP, MLP and KNN classifiers, respectively.

### 4.3 Results for the third approach

Four pre-trained models, namely InceptionResnet-v2, Inception-v3, ResNet50, and ResNet101 were used in this study. Moreover, GP and RF classifiers were used for the purpose of classification. Tables 3 and 4 present the detection

performance results obtained from different pre-trained models with GP and RF classifiers, respectively. It is clear that InceptionResnet-v2 and ResNet101 models with GP classifier outperform other models.

### 4.4 Results for the fourth approach

Table 5 presents detection performance results obtained for TL-based pre-trained models. It is clear that ResNet101 outperforms the other models with an accuracy that reaches 99.18%.

Confusion matrices and ROC curves for the proposed pre-trained models are presented in Figs. 19, and 20. Inception-Resnet-v2 and ResNet101 models have the same performance with an accuracy that reaches 100%.
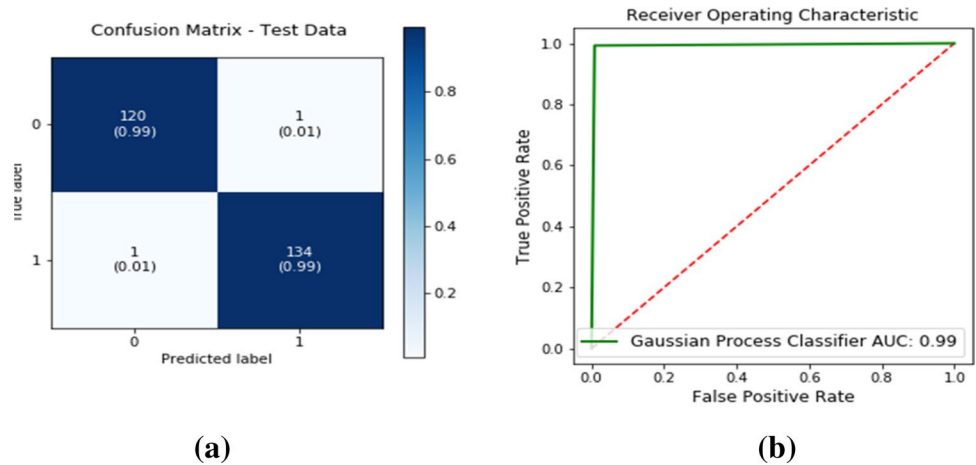


**Fig. 11** Confusion matrix and ROC curve for inceptionResnet-v2 model with GP classifier
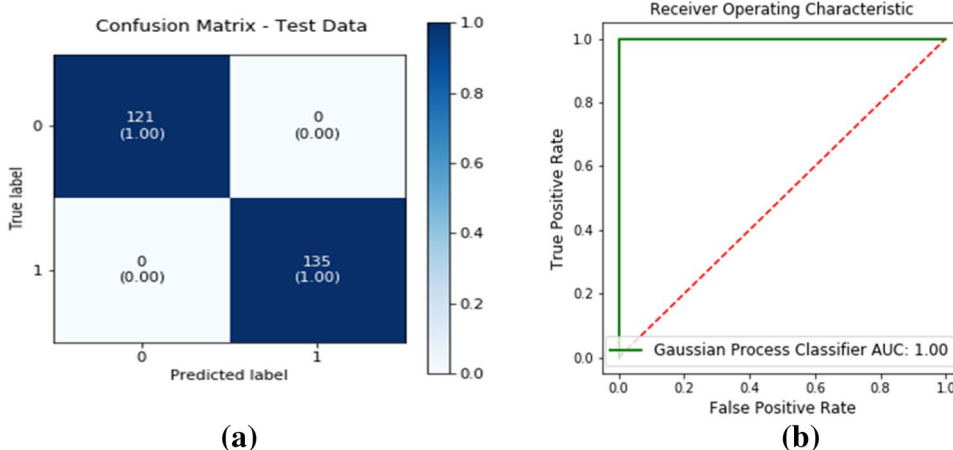


**Fig. 12** Confusion matrix and ROC curve for inceptionResnet-v2 with RF classifier
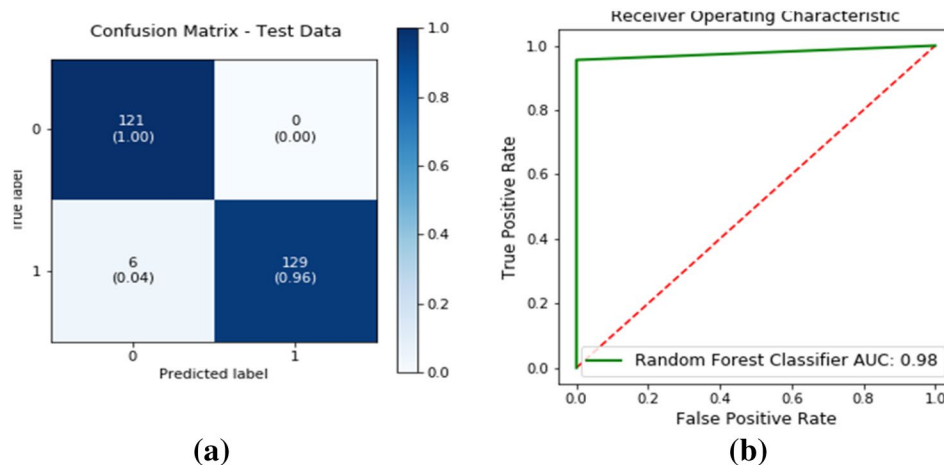
**Fig. 13** Confusion matrix and ROC curve for Inceptionv-3 model with GP classifier



**(a)**



**(b)**

**Table 6** Computational time of the examined approaches

| Method | Computational time (s) |
|---|---|
| ResNet101Features + GP | 30.9 |
| InceptionResnet-v2 features + GP | 38.7 |
| Fine-tuning of ResNet101 | 98.4 |
| Fine-tuning of InceptionResnet-v2 | 116.2 |
| CNN trained from scratch | 1198.1 |
| Local features + GP | 51.76 |

## 5 Discussion and comparison with the-state-of-the-art methods

As can be seen from the obtained results, the TL-based and deep feature extraction approaches present better accuracy levels than those of the CNN model and local feature extraction approaches. The obtained results reveal that pre-trained models work on both TL and deep feature extraction. These models were trained with a huge number of images that reach 25 million images. The convolution layer filters were selected to be efficient for new applications such as COVID-19 detection. Furthermore, the depth of these CNN models has a considerable impact on the application accuracy. The approach that comprises deep feature extraction with GP and RF classifiers gives the highest accuracy levels. This is attributed to the ability of the deep features to represent image activities, efficiently, in addition to the inherent characteristics of the used RF classifier to reduce the overfitting and increase accuracy. In addition, the GP is a powerful algorithm for classification problems (Figs. 11, 12, 13).

The computation time is the ultimate comparison metric between the proposed approaches. It is shown in Table 6, which clearly reveals that the deep feature extraction with Resnet101 and GP classifier gives a run time of 30.9 s, which is the least time. The deep feature extraction with InceptionResnet-v2 model and GP classifier is used to achieve the second best runtime of 38.7 s. The CNN model
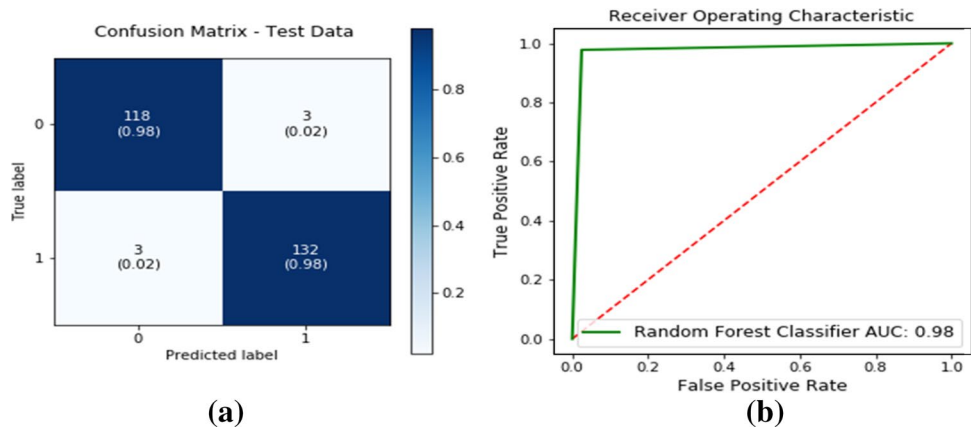
**Fig. 14** Confusion matrix and ROC curve for Inception-v3 model with RF classifier



**(a)**



**(b)**
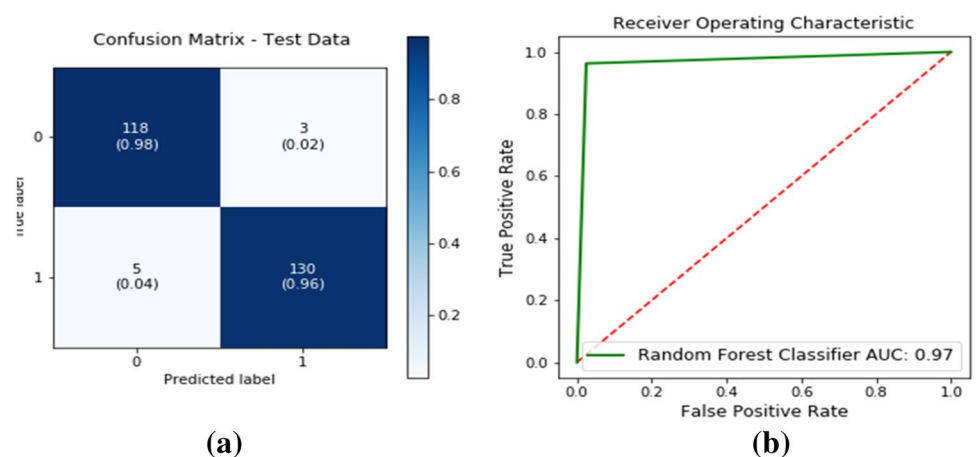
**Fig. 15** Confusion matrix and ROC curve for ResNet50 model with GP classifier



(a)

(b)
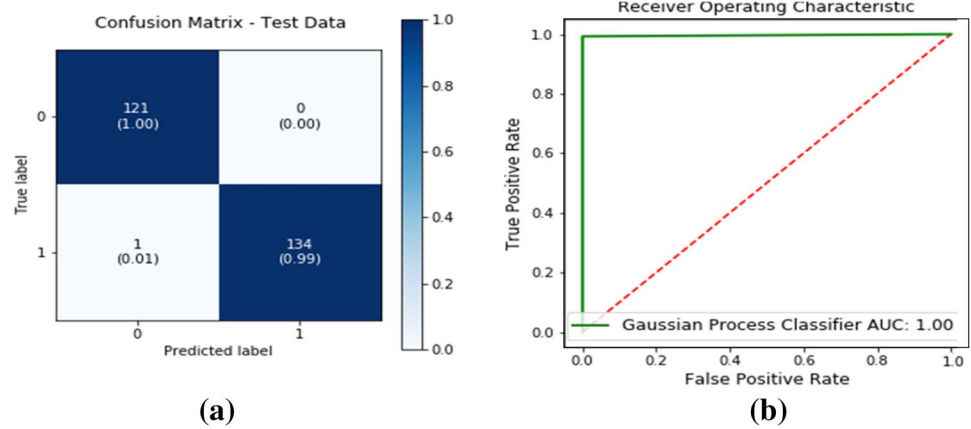
**Fig. 16** Confusion matrix and ROC curve for ResNet50 model with RF classifier



(a)

(b)

trained from scratch has the longest runtime of 1198.1 s (Figs. 14, 15, 16).

The accuracy level with the proposed approach reaches 100%, which is higher than the levels of traditional methods given in Table 7. These findings confirm the efficiency of the deep feature extraction process with efficient classifiers to perform the required classification task (Figs. 17, 18, 19, 20).

## 6 Conclusions

This paper presented four approaches for the detection of COVID-19 cases from X-ray images. A model has been built from scratch for this purpose. In addition, machine learning has been investigated for COVID-19 detection using histogram-based and statistical features. The task of deep feature extraction has also been investigated with the best machine learning classifiers for COVID-19 detection. Finally, transfer learning has been utilized to enhance the performance of the detection process. The obtained results proved that the transfer-learning-based and the deep features-based approaches outperform the local feature extraction approaches and the CNN model built from scratch. Deep features with Gaussian process (GP) and random forest (RF) classifiers perform better than the other approaches. This is attributed to the ability of the deep features to represent image activities, efficiently, in addition to the inherent characteristics of the used RF classifier to reduce the overfitting and increase the accuracy. In addition, the GP is a powerful algorithm for classification problems. Moreover, the deep feature extraction with Resnet101 and GP classifier gives a run time of 30.9 s, which is the least time compared to those of other proposed approaches. While chest X-ray images have been used to diagnose many lung diseases such as tuberculosis, pneumonia, and lung carcinomas, the proposed approaches are limited to the recognition of COVID-19 versus normal cases. In future work, feature fusion can be exploited for features extracted from different networks to enhance the detection accuracy. In addition, the proposed approaches can be extended to different diagnosis tasks.

**Table 7** Comparison of the proposed work with state-of-the-art models

| Model name | Method | Modality | Accuracy (%) |
|---|---|---|---|
| COVNet (Li et al. 2020) | COVID-19 CT | CT | 96 |
| BCNN (Ghoshal and Tucker 2020) | Dropweights-based Bayesian convolutional neural network | X-ray | 89.92 |
| Location-attention network (Xu et al. 2020) | Location-attention network + ResNet18 | CT | 86.7 |
| Transfer learning (Narin et al. 2020) | A modified inception TL model | CT | 79.3 |
| COVID-Net (Wang et al. 2021) | DCNN | X-ray | 92.4 |
| Pre-trained model (Wang and Wong 2021) | ResNet50 | X-ray | 98 |
| Transfer learning (Saiz and Barandiaran 2020) | VGG16 | X-ray | 94.92 |
| Acomprehensive study (Erdem and Aydın 2021) | Inception-v3 | X-ray | 90 |
| Metaheuristic algorithm (Canayaz 2021) | Binary particle swarm optimization + SVM | X-ray | 99.83 |
| Transfer learning-based (Jain et al. 2021) | Xception model | X-ray | 97.97 |
| CONVCOVID (Emara et al. 2021b) | CNN ResNet101 model | X-ray | 93 |
| | | | 99 |
| Transfer learning-based (Wang et al. 2021) | A modified inception TL model | CT | 89.5 |
| FPN (Song et al. 2021) | ResNet50+MLP | CT | 96 |
| Deep features (Loey et al. 2021) | ResNet50+SVM | X-ray | 99.64 |
| Segmentation (El-Shafai et al. 2021a) | DCNN | CT | 99.56 |
| SIGTra (El-Shafai et al. 2021b) | SR-GAN+TCNN | X-ray | 99 |
| Proposed work | Histogram minimum, histogram maximum, kurtosis, skwness and std+GP | X-ray | 97.53 |
| | Deep feature with InceptionResnet-v2 + GP | | 100 |
| | TL-based with InceptionResnet-v2 | | 100 |
| | CNN model built from scratch | | 94.78 |



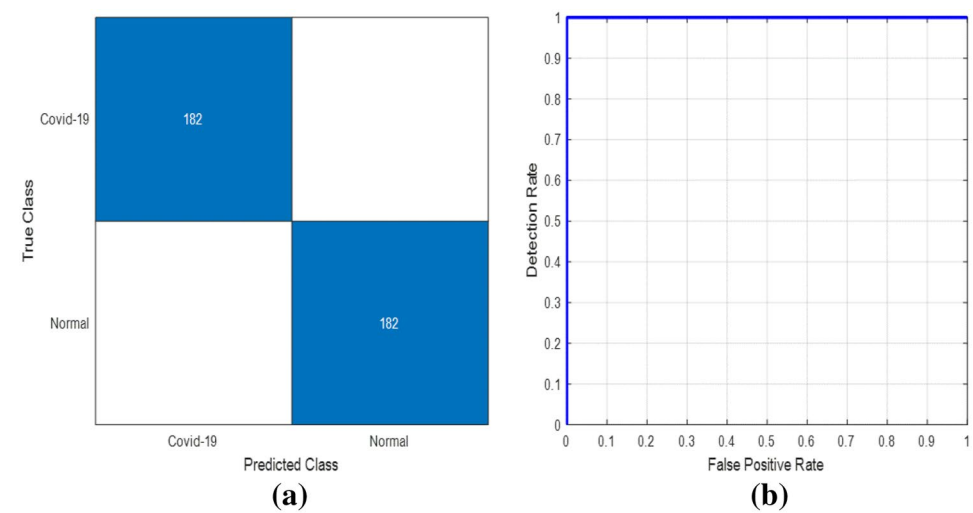**Fig. 17** Confusion matrix and ROC curve for ResNet101 model with GP classifier
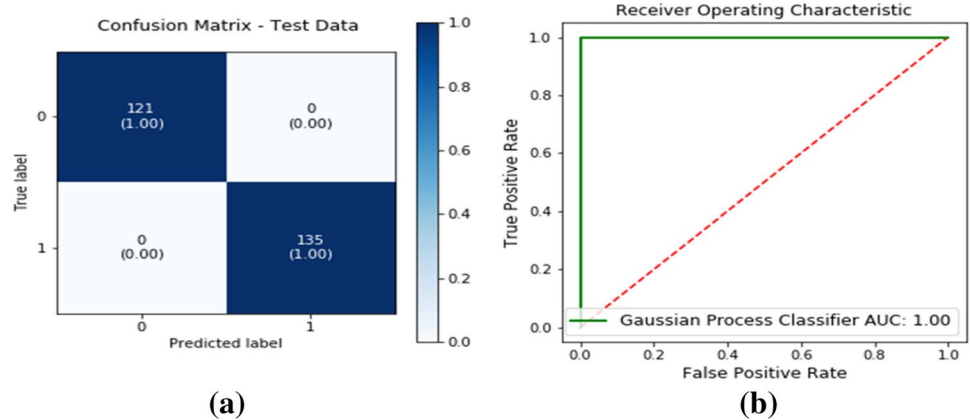
**Fig. 18** Confusion matrix and ROC curve for ResNet101 model with RF classifier



(a)

(b)
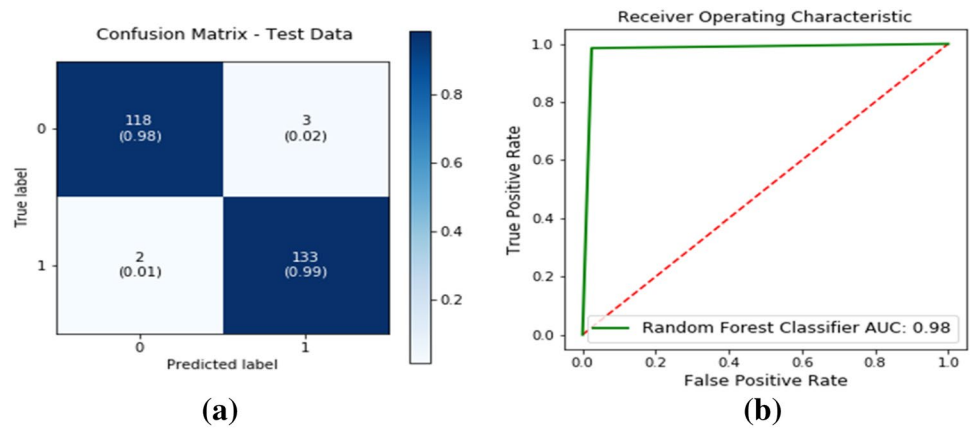
**Fig. 19** Confusion matrix and ROC curve for ResNet101
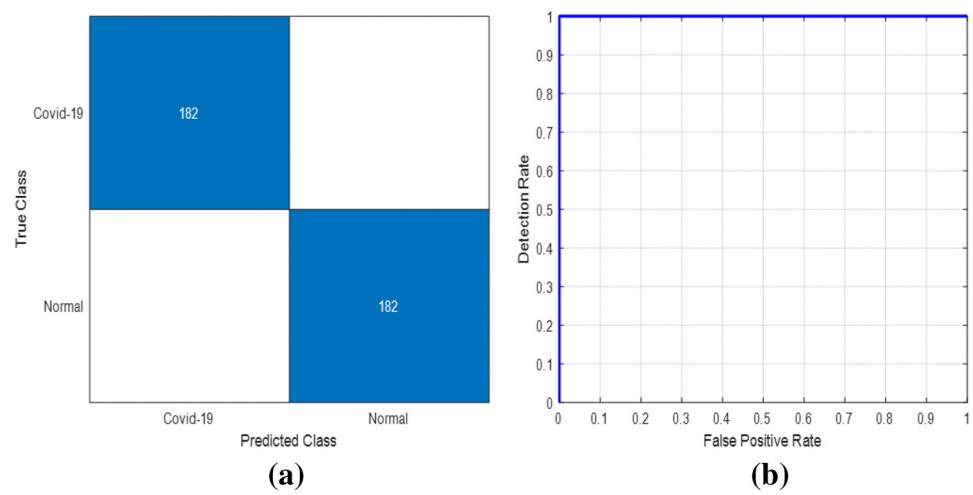


(a)

(b)

**Fig. 20** Confusion matrix and ROC curve for inceptionResnet-v2



(a)

(b)

# References

Alqudah AM, Qazan S (2020) Augmented COVID-19 X-ray images dataset. Mendeley Data, V4. https://doi.org/10.17632/2fxz4px6d8.4

Bhandary A, Prabhu GA, Rajinikanth V, Thanaraj KP, Satapathy SC, Robbins DE, Shasky C, Zhang Y-D, Tavares JMR, Raja NSM (2020) Deep-learning framework to detect lung abnormality—a study with chest X-ray and lung CT scan images. Pattern Recognit Lett. 129:271–278

Biswas P, Pal BB (2019) A fuzzy goal programming method to solve congestion management problem using genetic algorithm. Decision Making Appl Manag Eng 2(2):36–53

Bosch A, Zisserman A, Munoz X (2007) Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval, pp 401–408

Canayaz M (2021) Mh-covidnet: diagnosis of COVID-19 using deep neural networks and meta-heuristic-based feature selection on X-ray images. Biomed Signal Process Control 64:102257

Cheng L, Bao W (2014) Remote sensing image classification based on optimized support vector machine. TELKOMNIKA Indonesian J Electr Eng 12(2):1037–1045

Chowdhury ME, Alzoubi K, Khandakar A, Khallifa R, Abouhasera R, Koubaa S, Ahmed R, Hasan A (2019a) Wearable real-time heart attack detection and warning system to reduce road accidents. Sensors 19(12):2780

Chowdhury ME, Khandakar A, Alzoubi K, Mansoor S, Tahir MA, Reaz MBI, Al-Emadi N (2019b) Real-time smart-digital stethoscope system for heart diseases monitoring. Sensors 19(12):2781

El-Shafai W, El-Hag NA, El-Banby GM, Khalaf AA, Soliman NF, Algrani AD, El-Samie A (2021a) An efficient CNN-based automated diagnosis framework from COVID-19 CT images. Comput Mater Continua 69(1):1323–1341

El-Shafai W, Ali AM, El-Rabaie ESM, Soliman NF, Algrani AD, El-Samie A (2021b) Automated COVID-19 detection based on single-image super-resolution and CNN models. Comput Mater Continua 70(1):1141–1157

Emara HM, Elwekeil M, Taha TE, El-Fishawy AS, El-Rabaie E-SM, El-Shafai W et al (2021a) Efficient frameworks for EEG epileptic seizure detection and prediction. Ann Data Sci 1–36

Emara HM, Shoaib MR, Elwekeil M, El-Shafai W, Taha TE, El-Fishawy AS et al (2021b) Deep convolutional neural networks for COVID-19 automatic diagnosis. Microsc Res Tech. https://doi.org/10.1002/jemt.23713

Erdem E, Aydın T (2021) Detection of Pneumonia with a Novel CNN-based Approach. Sakarya Univ J Comput Inform Sci 4(1):26–34

Gheisari M, Wang G, Bhuiyan MZA (2017) A survey on deep learning in big data. 2017 IEEE Int Conf Comput Sci Eng CSE IEEE Int Conf Embedded Ubiquitous Comput EUC 2:173–180

Ghoshal B, Tucker A (2020) Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection. *arXiv preprint* arXiv:2003.10769

Groeneveld RA, Meeden G (1984) Measuring skewness and kurtosis. J R Stat Soc Series D (The Statistician) 33(4):391–399

Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395(10223):497–506

Jain R, Gupta M, Taneja S, Hemanth DJ (2021) Deep learning based detection and analysis of COVID-19 on chest X-ray images. Appl Intell 51(3):1690–1700

Kallianos K, Mongan J, Antani S, Henry T, Taylor A, Abuya J, Kohli M (2019) How far have we come? Artificial intelligence for chest radiograph interpretation. Clin Radiol 74(5):338–345

Karar ME, Hemdan EE-D, Shouman MA (2021) Cascaded deep learning classifiers for computer-aided diagnosis of COVID-19 and pneumonia diseases in X-ray scans. Complex Intell Syst 7(1):235–247

Li L, Qin L, Xu Z, Yin Y, Wang X, Kong B et al (2020) Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. Radiology. https://doi.org/10.1148/radiol.2020200905

Loey M, Manogaran G, Taha MHN, Khalifa NEM (2021) A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement 167:108288

Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. *arXiv preprint* arXiv:2003.10849

National Geographic (n.d.) Coronavirus. https://www.nationalgeographic.com/science/2020/02/here-is-what-coronavirus-does-tothe-body/20.03.2020. Accessed 7 June 2020

Pham TD (2021) Classification of COVID-19 chest X-rays with deep learning: new models or fine tuning? Health Inf Sci Syst 9(1):1–11

Rahman MA, Zaman N, Asyhari AT, Al-Turjman F, Bhuiyan MZA, Zolkipli M (2020) Data-driven dynamic clustering framework for mitigating the adverse economic impact of COVID-19 lockdown practices. Sustain Cities Society 62:102372

Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, Yan P, Chowell G (2020) Real-time forecasts of the COVID-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Model 5:256–263

Roy A, Manna A, Maity S (2019) A novel memetic genetic algorithm for solving traveling salesman problem based on multi-parent crossover technique. Decision Making Appl Manag Eng 2(2):100–111

Saiz FA, Barandiaran I (2020) COVID-19 detection in chest X-ray images using a deep learning approach. Int J Interact Multim Artif Intell 6(2):1–4

Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manag 45(4):427–437

Song Y, Zheng S, Li L, Zhang X, Zhang X, Huang Z et al (2021) Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images. IEEE/ACM Trans Comput Biol Bioinform 18(6):2775–2780. https://doi.org/10.1109/TCBB.2021.3065361

Stoecklin SB, Rolland P, Silue Y, Mailles A, Campese C, Simondon A, Mechain M, Meurice L, Nguyen M, Bassi C et al (2020) First cases of coronavirus disease 2019 (COVID-19) in France: Surveillance, investigations and control measures, January 2020. Eurosurveillance 25(6):2000094

Sun P, Li J, Bhuiyan MZA, Wang L, Li B (2019) Modeling and clustering attacker activities in IoT through machine learning techniques. Inf Sci 479:456–471

Tahir AM, Chowdhury ME, Khandakar A, Al-Hamouz S, Abdalla M, Awadallah S, Reaz MBI, Al-Emadi N (2020) A systematic approach to the design and characterization of a smart insole for detecting vertical ground reaction force (VGRF) in gait analysis. Sensors 20(4):957

Wang L, Wong A (2020) COVID-net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *arXiv preprint* arXiv:2003.09871

Wang S, Kang B, Ma J, Zeng X, Xiao M, Guo J et al (2021) A deep learning algorithm using CT images to screen for Corona virus disease (COVID-19). Eur Radiol 31:6096–6104. https://doi.org/10.1007/s00330-021-07715-1

Wu J, Zhang P, Zhang L, Meng W, Li J, Tong C et al (2020) Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. MedRxiv. https://doi.org/10.1101/2020.04.02.20051136

Xu X, Jiang X, Ma C, Du P, Li X, Lv S et al (2020) A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering 6(10):1122–1129

Yan L, Zhang HT, Xiao Y, Wang M, Guo Y, Sun C et al (2020) Prediction of criticality in patients with severe Covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. https://doi.org/10.1101/2020.02.27.20028027

Zhang Y, Li J, Sun C, Bhuiyan MZA, Yu W, Zhang R (2018) Hotml: a DSM-based machine learning system for social networks. J Comput Sci 26:478–487

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.