A Novel Artificial Intelligence-Based Model for COVID-19 Diagnosis Using CT Scans

Abdulrahman Alhaidari [a], Mustafa ElNainay [a,b], Emad Nabil [a,c] *

[a] Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, Saudi Arabia
[b] Faculty of Computer Science and Engineering, AlAlamein International University, Matrouh, Egypt
[c] Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt
* Corresponding author's Email: e.nabil@fci-cu.edu.eg

Summary

By the end of the year 2019, a global pandemic novel coronavirus, known as COVID-19, hits the world. The most widely used test for COVID-19 is the Real-Time Polymerase Chain Reaction. However, Real-Time Polymerase Chain Reaction test is time-consuming. Moreover, it suffers from a high false-negative diagnosis rate (low sensitivity). Computed Tomography scans, compared to the Real-Time Polymerase Chain Reaction test, can produce a result in a short amount of time. In this paper, we propose a novel model that hybridizes deep learning and machine learning together. Deep learning is utilized to extract the important features from Computed Tomography images, then the selected features are passed to an ensemble model for the classification. We used RseNet50 for features selection, and the classification is performed by an ensemble model that combines Support-vector Machine, Logistic Regression, and Multilayer Perceptron. The proposed model is compared with eleven state-of-the-art techniques and surpassed them using accuracy, precision, recall, and F1-score. The contribution of this paper is introducing a novel model with high performance for the diagnosis of COVID-19. With the aid of this model, we could identify positive cases rapidly for early isolation. At the same time, we can use it in combination with Real-Time Polymerase Chain Reaction test to increase its sensitivity.

Keywords

COVID-19 diagnosis; Computed Tomography scans; Deep learning; Real-Time Polymerase Chain Reaction test; machine learning

1. Introduction

Artificial Intelligence (AI) proved to be effective in solving a lot of problems in many fields of our today's life [1]. Healthcare is one of these fields where AI has contributed to many success stories. Currently, AI techniques are widely used in the fight against one of the severest pandemics, COVID-19. It has infected over 219 million individuals, with over 4.55 million deaths and nearly 63 million recovered individuals, until the time of writing this paper, [2].

This means COVID-19 has about a 2.1% death rate, which is considered a high mortality rate. The severity of COVID-19 comes from its ease of spread, which results in an exponential growth rate. The number of suspected cases exceeds the capacity of doctors and hospitals in many countries. Thus, there is an urgent need for a fast diagnosis tool to classify new cases: suspected to have COVID-19 or healthy cases.

The most widely used and standard test for COVID-19 is the Real-Time Polymerase Chain Reaction (RT-PCR) [3], [4], which looks for the existence of the virus or not. However, RT-PCR is time-consuming [5], [6]; the results could take 1-2 days to be finalized. Moreover, it suffers from a high false-negative diagnosis rate (low sensitivity) due to several reasons, including the defective viral materials in the taken sample or an error in the procedure. The low sensitivity means a positive case may be diagnosed as negative, which means this case will move freely in public, and this will be a source of infection. Besides, RT-PCR devices are expensive and very limited in many developing countries. Since the results of the RT-PCR are typically received after 24-48 hours, it may delay the isolation of positive patients. These scenarios contribute to the infection spread due to the freely moving infected cases that are misdiagnosed. This enlightens the need for an alternative method to speed up the process of diagnosis. One of the good alternatives is to use Computed Tomography (CT). CT scan is not expensive, and it is a fast test relative to RT-PCR [7], [8]. Machine learning techniques are an excellent candidate that can help in the identification of infected cases by analyzing their radiological images. This is because of their high performance and accuracy. Thus, positive cases can be identified swiftly for isolation. After isolation, which will prevent the possible infection of others, RT-PCR can be examined for more confirmation of the infection.

In the study of [9], it was found that the time taken for automated positioning (AP) for CT images is 28% less than manual positioning (MP). Also, the results show a better ratio of positioning for AP over MP, 99% and 92% respectively [9]. These drive in the direction of automatic CT scans analysis.

This paper aims to design a machine learning model that is effective and efficient in detecting COIVD-19. We hypothesize a hybrid deep learning model and machine learning classifiers in one model can gain better results than one of them alone. Phase one of our proposal is selecting the most important features from a CT scan. Phase two is to send these features to a second model where we get the class of the input image. In phase two, we will utilize the ensemble approach, as we assume that combing more than one

Manuscript received March 5, 2022
Manuscript revised March 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.3.83

machine learning model will achieve better results than solo models. In this paper, we will investigate those hypotheses. Our proposal is compared to state-of-the-art models to prove its superiority. The system will help physicians in recommending immediate isolation of patients where it is necessary.

This paper is organized as follows. The related work section discusses the previous techniques used in COVID-19 diagnosis using CT scans. The materials and methods section presents the methodology, the data used in our study, the architecture of the proposed model, and the performed experiments. In the results section, the performance of our model is compared with another prior research. In the discussion section, we comment on our model and discuss the main findings and implementation issues. Finally, the conclusions and the future directions are presented.

2. Related work

The current standard test of COVID-19 is the RT-PCR. Still, it suffers from low sensitivity, e.g., many positive cases are misdiagnosed as negative cases as its sensitivity ranges from 59% to 71%. Using radiology images in COVID-19 diagnosis can help to solve this problem[10] [4], [11].

Chen et al. compared COVID-19 patients' results of both RT-PCR and CT-scan, who were initially diagnosed as negative cases. It was shown that patients whose cases were initially negative have a higher possibility than RT- PCR of depicting pulmonary consolidation in their CT-scan, which is vital for COVID-19 diagnosing [12].

As reported in two new studies, the diagnosis using CT scans is more sensitive than the initial RT-PCR with 98% vs. 71% and 88 vs. 59%. Moreover, many patients were identified as positive cases with CT findings reported as negative cases by RT-PCR ([4], [13]). Artificial Intelligence and Deep Learning proved to be very effective in the battle of fighting COVID-19 [14].

Based on the above reasons, many works were conducted using AI and ML techniques for the automatic diagnosis of COVID-19 by the analysis of CT images [15], [16], [17], [18], [19], [20]. A convolution neural network model based on ResNet50 has been used in [21] to detect COVID-19 from chest CT images collected from multiple hospitals, including more than 4000 CT images for 1296 COVID-19 cases, 1735 Community-Acquired Pneumonia, and 1325 non-pneumonia CT exams. The deep learning model was able to detect COVID-19 with high accuracy and distinguished them from community-acquired pneumonia and non-pneumonia lung diseases [21].

In [22], the authors use transfer learning to detect anomalies by using deep learning. Two datasets were utilized for patients with diverse diseases, ranging from confirmed viral pneumonia, bacterial pneumonia, healthy, and COVID-19. An ensemble was adopted on three dissimilar types of CNN for testing on unseen cases of 33 COVID-19 cases, 208 cases of pneumonia. To get the most accurate diagnosis based on the CT images, one crucial step is image preprocessing. It can affect how well the model can capture the variation in data; therefore, several preparation steps such as image enhancement, segmentation, and bone suppression were applied with the collected data to improve overall performance [23].

In [24], authors developed a model that can identify a patient's state, whether he has COVID-19, Influenza-A viral pneumonia patient, or healthy case. They used a 618 CT images dataset gathered from three hospitals from China. They applied the CNN ResNet-18 architecture with location attention and compared their results with the classical ResNet-18.

In [8], the authors compiled a CT image dataset from the Renmin Hospital of Wuhan University (Wuhan, Hubei province, China). The dataset for 106 patients, 51 COVID-19 cases, 55 non-COVID-19 cases. They used UNet++ CNN architecture to classify the data. Authors in [25] used Deep NN in their classification model; namely, they used ResNet152 CNN architecture. They collected 960 CT images from 496 positive COVID-19 cases and 262 negative COVID-19 cases. These data are collected from three hospitals in Wuhan, China. The authors also used 1125 negative cases from public datasets.

In [26], authors used ResNet-50 CNN architecture. They compiled a CT images dataset of 1036 normal slices and 829 abnormal (positive COVID-19) slices for 157 patients from the USA and China. To increase the dataset size, they used image rotation, cropping, and flipping for data augmentation. Furthermore, transfer learning using various residual networks on 3-dimensional COVID-19 CT-scan to identify COVID-19 was used by Serte S. and Demirel H. and resulted in the area under the curve (AUC) of 96% [27]. Also, In [28], the authors used SVM to classify CT images of suspected COVID-19 patients into positive and negative cases. The authors used a dataset of 150 CT images. After that, they used four feature selection techniques as a preprocessing step before feeding the data to the SVM classifier. According to their experiment, the best feature selection for the SVM was Gray Level Run Length Matrix technique with 10-fold cross-validation. Our proposed model utilizes deep learning for feature extraction and classical machine learning models for classification. It yielded a superior classification accuracy compared to the existing models, even though our dataset is unbalanced.

3. Materials & Methods

Fig. 1 summarizes our methodology in this research. The used dataset contained grayscale images. The first step was data augmentation to increase the dataset size. Also, coloring the images and checking if the augmentation and coloring will increase the performance or not. Deep learning techniques proved to be powerful in feature selection. For so, we used deep learning models for selecting the most important features from input images. We inspected different deep learning models for that task and picked the best one. Ensemble models also proved to be a powerful candidate in solving classification problems. So, the selected features are passed to an ensemble model. The sub models of the ensemble model are selected based on their performance in classification using the selected features by the deep learning model. Finally, we checked if our final model is performing well as we predicted or not. The comparison will include state-of-the-art algorithms.

3.1. Data

The provided dataset by [29] collected from the Public Hospital (HSPM) in Sao Paulo, Brazil is used in this paper. It belongs to patients who were tested and/or confirmed for COVID-19 by RT-PCR, either positive or negative. There are 80 COVID-19 patients and 50 healthy subjects. The CT scans are in image format, pdf; Figures 2 and 3 show a sample of the CT slices for both.

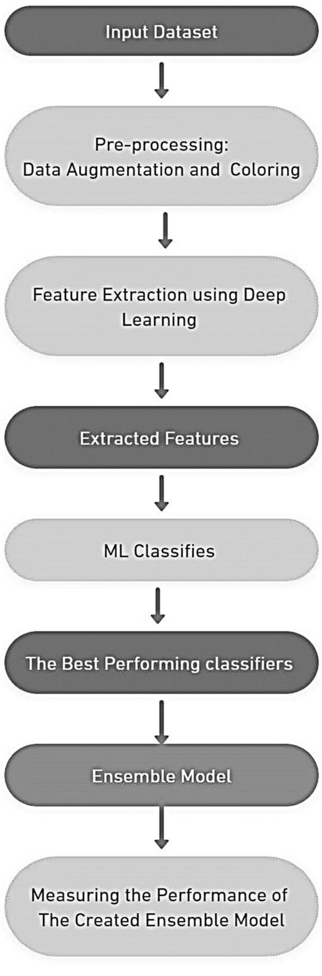

Fig. 1: Methodology of the research.

The dataset contains clean images, as there are no markings on the CT slices, unlike some other image datasets where notes, marking, or comments are printed on images; one example of such data is the dataset provided by [30]. Table 1 outlines the utilized data and the number of patients in addition to the associated CT images.

On average, each healthy patient has approximately 21 slices, while each COVID-19 infected patient has an average of 28 slices. The dataset also contains individuals who are not COVID-19 positive or negative, as they have other lung diseases, but it is not related to COVID-19. Therefore, they were excluded from our experiments.

Table 1: Dataset split: training: 80%, testing: 20% .

	# of (+) patients	(+) slices	# of (-) patients	# of (-) slices
Training	64	1702	40	620
Testing	16	465	10	137

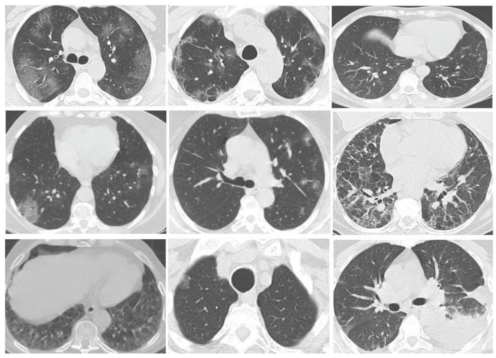

Fig. 2: Negative examples, not infected with COVID-19.

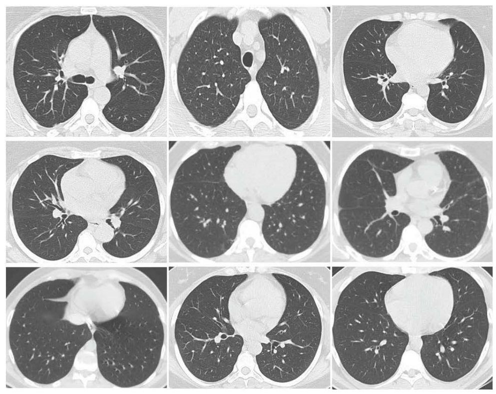

Fig. 3: Positive examples, infected lungs with COVID-19.

The used dataset is from [29], but one limitation of their work in the implementation is data splitting, according to their paper. The authors used mixed CT scan data in the designed model, which achieved a decent accuracy, 97.38%. However, the utilized CT slices were divided between training and testing as one set, meaning that no information supports which image belongs to which patient. This causes information leakage from training data to test data resulting in inaccurate testing results. As when new testing data is being classified, the model would already have information about the current image being classified. Even though the model can distinguish COVID-19 infection, it cannot be generalized. This means, if unseen data, out of the training and testing, is fed into the classifier, the probability of misclassification is high. It is due to the fact that the model is contingent upon having assistance from training data as the newly classified CT image is related to one or more existing images that the model already has information about. Therefore, we considered this limitation when we used the dataset and resolved this problem in our implementation—the patients' CT scan slices present in one set, either training or testing.

3.2. Preprocessing

To have a robust model, we carried out several experiments using the original data augmented with a modified version derived from the same CT images. To diagnose COVID-19 patients accurately based on lung CT scan, the built model robustness depends on data quality, processing, and the used method [31].

The size of the CT images varies; for example, the largest image is 200KB while the smallest is 20KB; the same applies to the dimensions. As the image size gets smaller, the amount of information extracted is expected to be diminished. Because the aim is to discriminate between images, we need to maximize the extracted information to get a better prediction. For example, the most diminutive dimensions are 185 by 121, while the largest dimensions are 535 by 425. Thus, we averaged the dimensions, and the result was that all images are 378 by 278 pixels. Thus, the best option based on what was observed is to resize the images above the dimensions while keeping the rest of the CT images settings unchanged.

As Table 1 shows, the classes are imbalanced; positive cases have more patients and CT scans. In order to verify the validity of increasing the number of images that might improve the model's performance, we used data augmentation, such as adding more data using the same dataset by shearing, flipping, rotating, and sharping images' pixels. However, the results show that when images are adjusted, the prediction deteriorates. In CT scans, a white-flecked pattern indicates whether the subject's lung has a disease or not [32]. For this reason, we inverted the colors of the images, and the goal is to give a lower weight for black and a higher to white since RGB for white color is (255, 255,255) and for the black color is (0, 0, 0). Nevertheless, the original coloring was better in prediction, and instead of improving the performance, it deteriorated. Thus, we kept images in the original colorant.

3.3. Proposed Model

We have several hypotheses when it comes to solving our problem. By using deep learning for feature extraction, the extracted data will be used for10 classification. The final model will be selected based on different evaluation metrics, and the superior method will be our final solution.

Table 1 shows that the dataset was split into 80% of the subjects in the training data while 20% were used for model evaluation. Data partitioning depends on the number of patients rather than on how many CT scan each patienthas since each has a slightly varying number of CT slices, as COVID-19 patients have a higher CT scan number of slices than healthy ones. Each patient's CT scan presents in either training or testing through our experiments.

Deep learning is used for feature extraction, and then the features are passed to a classification model. One approach is ensembling models that proved to be a powerful classification technique. Merging both deep learning and ensembling would enable us to check whether the result is better than other approaches or not.

3.4. Deep learning for feature extraction

Initially, we performed several experiments to determine which deep learning model is better and more suitable for our problem. We compared the results of three powerful models, ResNet50, Xception, and DenseNet201. All of which are in the original setting and no image preprocessing, except that a Sigmoid function replaced the original prediction layer to predict positive and negative cases. Among the three models, the best-performed model was selected to be used for CT scan feature extraction. The obtained results are outlined in Table 2

Table 2: Original ResNet-50, Xception, and DenseNet201 comparison that was trained on our dataset

Model	Accuracy	Precision	Recall	F1-score
ResNet-50	86.21%	92.26%	89.68%	90.95%
Xception	80.73%	86.74%	88.60%	87.66%
DenseNet201	65.78%	89.12%	63.44%	74.11%

Overall, these experiments were carried out to determine how the final model's performance would be maximized and the decision to use a pre-trained model because it shows its potential, especially ResNet50 when used as feature extraction in integration with the classical machine and deep learning algorithms.

As the goal is to customize the nominated model to extract or shrink the number of features, the main focus was on features extracted rather than the pre-trained models themselves. Table 2 compares the performance of the three models. The best performed is the convolutional deep learning ResNet50, which shows its superiority among the other models.

3.5. Resnet50 for Feature Extraction

The original ResNet50 model has 50 layers. The first layer is the image input, and the final layer of the ResNet50 is the predictor with 1000 predictors. As the final layer is not needed because we are using this model as a feature extractor, it was removed. Then, two additional layers were added. The first added layer is to normalize the features to twodimensional average pooling generating, as a vector of 2048 for each CT image. The second is a dense layer of 30 nodes and uses Rectified Linear Unit (ReLU) as an activation function.

The initial aim was to diminish the number of features from 2048 to 30, and, surprisingly, the final layer produces only two features out of 30. The rest of the values are zeros. These two features were used to predict infected cases, and the result of this approach is reported in the Results section. However, when we removed the 30 nodes in this ReLU layer, the model's performance increased, and the 2048 features achieved superior performance. Even though the classification with two features resulted in a decent accuracy, the 2048 features had better classification results.

3.6. Feature Extraction for The Proposed Architecture

In the proposed architecture, CT scan images were resized. Then, each image was assigned either 0 or 1 in an external file and linked to each record to a CT image. Zero means a case is not infected with COVID-19 and one is vice-versa.

As we removed the top of ResNet50, the input dimension was made flexible instead of having the default fixed input size (224,224,3) of the original model. A global average pooling layer is added to normalize and reduce the dimension of the previous layer since we need the feature vectors to be in a 2-D format. Following this, a ReLU bottleneck was added to reduce the number of features. At the end of each model, a Sigmoid function was joined as an output, which gives a probability that the input picture belongs closely to either class 0 or 1. A Sigmoid probability of less than 0.5 is considered a COVID-19 negative, and on the other hand, an output greater than or equal to 0.5 is considered a positive case since we kept the class separation threshold at0.5.

$$g(x) = \frac{1}{1 + e^{-x}} \qquad (1)$$

The function g(x) in (1) will not reach the value one unless x is a significantly large number. Therefore, it is rarely possible to see a prediction of precisely 1. Moreover, to prevent the model from overfitting, when the model learns the training data and fails on the validation and cannot be generalized, activity regularization with a value of L1=0.01 is applied on the Sigmoid function. It is a penalty value executed as a pre-layer bias for the output of the global averaging pooling, the preceding layer.

After having the above adjustments, we loaded ImageNet weights into the ResNet-50 and set the original layers to non-trainable, meaning that their weights cannot be adjusted throughout the training process. Then, the weights of the added layers were set as trainable. This is because when we sat all layers trainable, the training accuracy was mediocre, which is a sign of underfitting. The loss value declines up to some point and raises again. one reason for this is that as the data set is not balanced and the training is moving towards fitting training data, it tends to move closer to positive cases as they are the majority. Thereupon, early stopping was very effective in preventing the model from being biased to the infected cases.

The training set, 70 %of the data, was used to train each model separately. When the model achieved a low loss between 0.1700 and 0.1400, we stopped training. The determination of the loss range was based on different experiments, as 10% out of 80% of the training data was used for evaluation. Our main aim here is to extract features as near as possible to the training later to extract features for all CT scans, so we are not interested in the model's accuracy at this stage rather in extracting features. However, it is crucial to minimize the training image loss but avoid overfitting when the model memorizes instead of learning the pattern of data, with early stopping. Training the model is to initiate the weight and make the extracted features near to the training data, noting that the testing set is kept separate and has no interference with the training process.

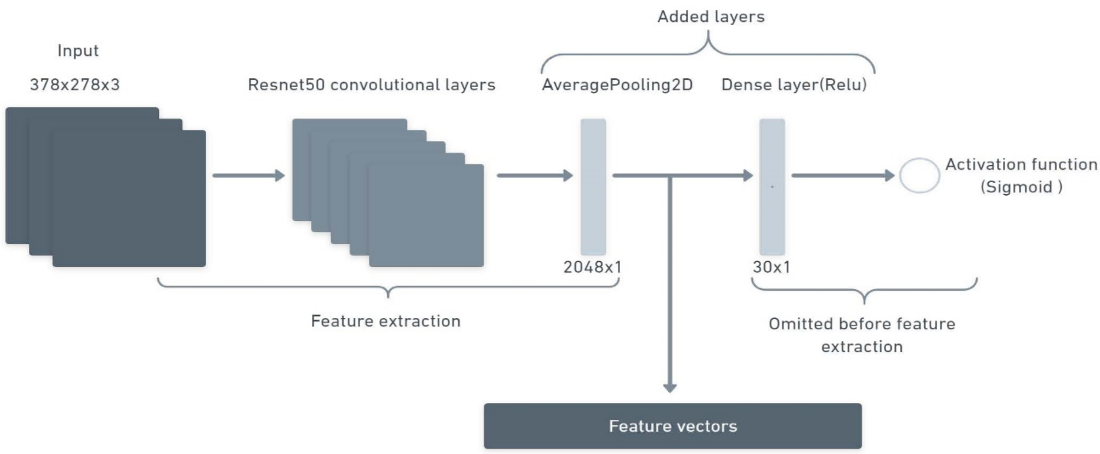

Fig. 4: CT Scan feature extraction for the proposed model

After the added layers weights adjustments, the model is altered by eliminating the ReLU and the Sigmoid function and keeping the average pooling attached to the original layers. Then, all CT image features are extracted.

After generating the features, one is assigned as a label in a column after the last feature. Table 3 shows that ResNet-50 produces 2048, and with Sigmoid, the new total is 2049 features.

3.7. Diagnosis Based Classification

To discover the performance of classification models on the feature vectors extracted from Fig. 4, we implemented eight machine learning models. Each one of those trained on the generated features, 2048 features, and a target variable. The training phase is similar to that of the initial model having 80% of training and the rest, 20% for testing. Cross-validation is applied on some models, and it crosses validate training samples. Testing the model was not involved in the model selection phase.

Table 3: Number of features for each ResNet50.

Model	# Of non-trainable parameters	# Of trainable parameters	#Of features
ResNet50	23,587,712	61,501	2048

K-Nearest Neighbors (KNN) algorithm finds the nearest COVID-19 case if new data is being predicted. Using cross-validation with K folds= 15, training data was divided into 15 parts, one for validation and 14 for training for each K-fold, and loop over both K of KNN from 1-20 to find the best K-nearest neighbor. After training and validation finished, 17 was the best K for the KNN classifier. Subsequently, based on the best K, the algorithm is trained and tested using the designated set.

An ensemble learning is multiple classifiers that are trained and combined to generate a prediction for a particular problem. Hence, we have used experimented with various machine learning models and combined the best-yielded predictors.

To overcome overfitting in Logistic Regression, different penalty values, C value, enlarge the magnitude of Logistic Regression parameters. The goal is to choose the best fit for data and decrease the error of the prediction. Therefore, we iterate over a range of values to find the best C, the regularization parameter using the same cross-validation method above. C values used were 0.001, 1, 1000, and these were specifically chosen because we needed to see the classifier performance when the regularization is at different intensity.

The Quadratic Classifier (QDA) classifier utilized a regularization parameter of 1. SVM's kernel adopted is Radial Basis Function (RBF) that produces a non-linear hyperplane to separate the classes, and regularization and gamma variables harvested using cross-validation and resulted in 0.001 for gamma and 1 for the regularization parameter.

The Decision Tree utilizes the criterion of Gini, which is a measurement of split impurity, and the max depth of the tree is 12. Random Forest, on the other hand, used a max depth of 5, and the maximum number of trees is 12. The feature vectors were scaled using the min-max scaler. It normalizes features to be on a scale of 0 and 1. The maximum value transformed to 1 while the minimum changed to 0. The other is set in the range of 0-1. Moreover, twenty trees were trained and boosted using the AdaBoost algorithm, with the original values of the feature vectors.

Multilayer Perceptron (MLP) is a deep learning algorithm that applies backpropagation to the training phase. The used was log-loss function used LBFGS for optimization and ReLU as activation functions; The number of hidden layers and nodes is 5, with a maximum iteration of 500 times. This was chosen based on the fact that LBFGS needs the number of iterations less than or equal to lose function calls. For instance, when the maximum iteration was set to 10, which is not sufficient, the fit accuracy on the testing decreases by approximately 11%. The reason is that the algorithm still did not reach the local minima, and the stall needs more iterations. If the local minima are achieved, even if not all the 500 iterations are executed, the model training is terminated.

Each model above was trained independently. After each step training phase, we report the precision, recall, and F1-score, ensuring the models' results are maximized on the testing data by models' tuning. After the results of each model were obtained, three models with the highest accuracy, the least false positive, and false negative were selected. The outstanding models used hard voting for each class based on the input. On the other hand, hard voting was implemented. For instance, if a COVID-19 CT scan slice was fed into the voting model, each one of the selected models, SVM, Logistic Regression, and MLP, would vote either 0 or 1. Again, 0 for negative and 1 for positive case.

Each of the ensemble voting models has a weight, which is shown in Fig. 5. The weight determines the importance of each electing model. The weights are chosen since each model has different accuracy, false positive and false negative. Thus, they were ranked from the highest, SVM with 3, MLP with 2, and Logistic Regression with 1 to signal each model's voting importance for the prediction. In all cases, majority voting is possible; however, if there is a tie between voters, SVM vote for a positive class where the other votes for the negative, for instance. The final decision of the ensemble will be the negative class. This became an advantage for our model and increased the prediction accuracy.

4. Results

In this section, we summarize the results of the used model. Tables 5 and 4 show models' accuracy, precision, recall, and F1-score for each, in addition to the proposed ensemble model, RESML, in table 5. It stands for ResNet, as a feature extractor, Support-vector Machine(SVM), Multilayer Perceptron(MLP), and Logistic Regression (LR) models used in the design of an ensemble model.

Results in Table 4 were produced by extracting features of the images using the 30 nodes layer from Fig. 4. Only two features have values, and the rest are extracted by ResNet as zeros. QDA algorithm was implemented and trained, but the results were below the average, scored 26.24%. As we observed, the outcome is not significant. It was eliminated as a possible model. Table 4 shows the models that we trained on two features that were obtained by a modified ResNet50 for the CT image. The best model with two features is Random Forest, Decision Tree, and AdaBoost. They produce the same results of all metrics; thus, only Random Forest is reported in Table 4. Even though some models resulted in a higher accuracy compared to results in Table 5, the performance could not be maximized. When we used the ensemble method, the results of the voting did not achieve the RESML performance as shown in Table 5.

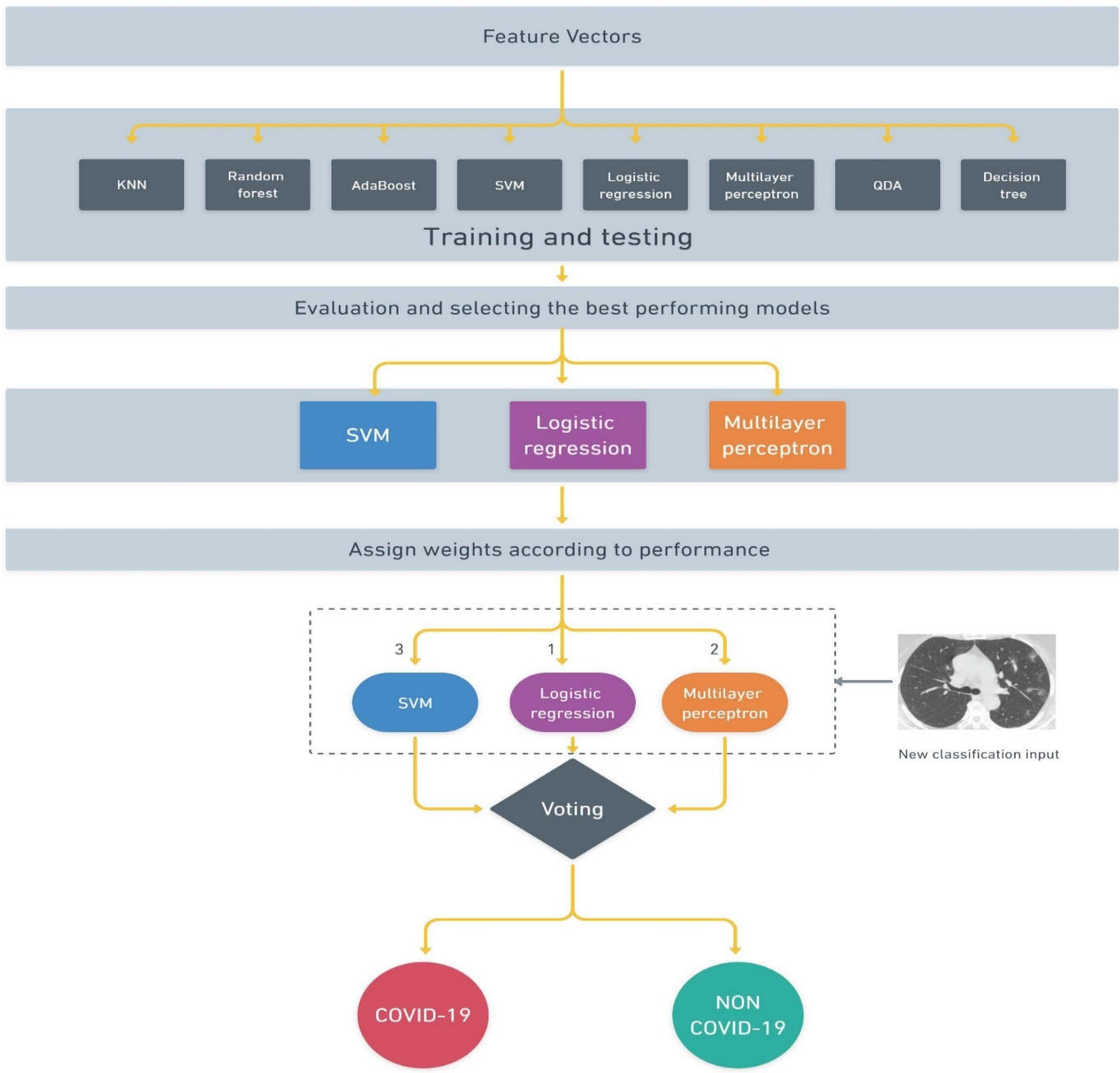

Fig. 5: Ensembling and the weighting method utilizing the extracted 2048 features

Model	Accuracy	Precision	Recall	F1-score
KNN	94.68%	98.87%	94.19%	96.47%
MLP	95.01%	99.09%	94.40%	96.69%
SVM	95.18%	98.44%	95.26%	96.83%
Random Forest	96.34%	98.68%	96.55%	97.60%
Logistic Regression	95.01%	98.22%	95.26%	96.72%

Table 4: Diagnosis with two features out of the 30 nodes of the ReLU layer

It can be seen from Table 5, the best model in terms of accuracy is the proposed model, while the inferior is QDA. It is also observed that the proposed model has the best performance among all other models in every metric, except the recall of the MLP model is slightly better.

Table 5: Diagnosis with 2048 features including RESML. The highest values are formatted in bold.

Model	Accuracy	Precision	Recall	F1-score
1. KNN	93.19%	98.18%	92.90%	95.47%
2. MLP	95.85%	96.41%	**98.28%**	97.34%
3. QDA	66.45%	97.47%	58.06%	72.77%
4. SVM	97.34%	98.49%	98.06%	98.28%
5. Decision Tree	86.71%	93.06%	89.46%	91.23%
6. Random Forest	92.52%	95.06%	95.27%	95.17%
7.Logistic Regression	95.51%	96.20%	98.06%	97.12%
8. AdaBoost	91.20%	94.78%	93.76%	94.27%
9. ResNet-50	86.21%	92.26%	89.68%	90.95%
10. Xception	80.73%	86.74%	88.60%	87.66%
11. DenseNet201	65.78%	89.12%	63.44%	74.11%
12. **RESML**	**97.84%**	**99.13%**	98.06%	**98.59%**

Table 6: Proposed models by different literature that have data leakage compared with RESML. Highest values are formatted in bold.

Method	Accuracy	Precision	Recall	F1-Score
1. xDNN [29]	97.38%	**99.16%**	95.53%	97.31%
2. ResNet [29]	94.96%	93.00%	97.15%	95.03%
3. GoogleNet [29]	91.73%	90.20%	93.50%	91.82%
4. VGG-16 [29]	94.96%	94.02%	95.43%	94.97%
5. AlexNet [29]	93.75%	94.98%	92.28%	93.61%
6. Decision Tree [29]	79.44%	76.81%	83.13%	79.84%
7. AdaBoost [29]	95.16%	93.63%	96.71%	95.14%
8.SepNorm+Contrastive [33]	90.83%	95.75%	85.89%	90.87%
9. Single COVID-Net [33]	89.09 %	94.58 %	83.78%	88.97 %
10. MS-Net [33]	87.98 %	93.78 %	84.91%	88.73 %
11. Joint COVID-Net [33]	78.42 %	80.82 %	74.07%	77.86 %
12. RESML	**97.84%**	99.13%	**98.06%**	**98.59%**

In table 6, a comparison between different studies is presented. The models from model 1 to model 7 are reported by [29], while models 8 to 11 are reported by [33]. All these models, 1 to 11, suffer from data leakage, where patients' CT-images are present in training and testing datasets. This leads to the proposed models being inclined to have assistance from CT-slices present in training since it is related to what is being classified. Testing images shouldbe unseen, meaning there should be no relationship with patients in the training data. The information leakage indeed makes the model outperforms its real performance. Despite that, the proposed model RESML, mentioned in the last row, still surpasses all of them, except the precision metric, the xDNN surpasses RESML with a minimal value. We can induce from this comparison that RESML surpasses all the mentioned models if we consider a dataset with no information leakage.

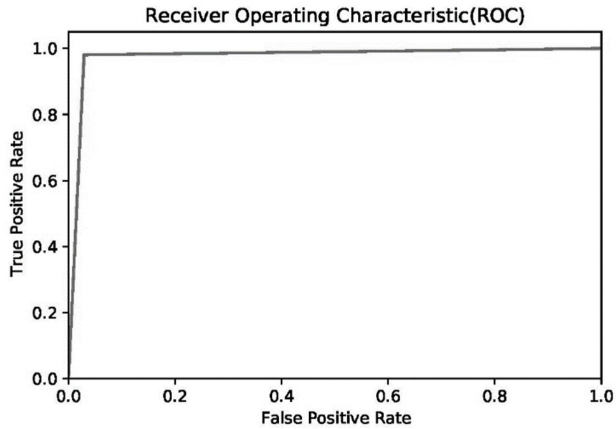

Fig. 6: ROC of the RESML

The Fig. 6 curve sketches the ROC of the RESML model. It shows that RESML covers most of the area under the curve with a low false-positive rate and a high true-positive rate for COVID-19 cases diagnosed. This produces better results than other existing classifiers.

5. Discussion

The proposed approach consists of two consecutive models, ResNet50, for feature extraction, and the ensemble model was designed to classify the COVID- 19 cases. Having heterogeneous classifiers using voting increases the overall performance. Among all experiments, and by using this method, the proposed model achieved an accuracy of 97.84%. The error rate in the classification is the lowest among the other single classifiers. For instance, out of 100 CT scans, approximately 3 to 4 get classified erroneously. However, the accuracy gives only an overview of the performance, but it does not, alone, prove the model's performance. Other metrics are needed to evaluate the models; consequently, we report four metrics, including each model'saccuracy.

The classifiers were trained on the same set separately; nevertheless, it is evident that the performance of each varies since each model tries to capture the variability of the training data differently. For instance, 66% was obtained by QDA, which is a single classifier, which is the lowest, while the best single model, SVM, scores 97.34%. On the other hand, for precision andrecall, they had decent outcomes for most of the models except the recall of QDA, 58.06%, while the proposed model is the highest at 99.13%.

Moreover, it was observed that the best-performing model is ResNet50 based on the original models' experiments. However, this is not true when allowing all layers of ResNet50 to be trainable to predict the infection. The loss value decreases up to a point where the model performance declines significantly; as a result, the loss gets high again. Thus, it was determined to set the original model's layers as non-trainable and limit the weight adjustments to the last added layer using regularization. Fig. 4 pinpoints the added trainable layers and shows the process of how getting feature vectors waspossible.

The model's evaluation is done based on training and testing separation in the whole process; as noted earlier, no patient data is present in both training and testing. Besides, the same set was used to construct further models. One of the significant drawbacks of mixing CT data and dividing them with no separation is patient data leakage between training and validation. This was prevented throughout the process as the built model needs to be generalized.

6. Conclusions

This paper explores various deep and machine learning techniques and proposes a new hybrid model for diagnosing COVID-19 based on lung CT scan; We formulated the following hypothesis, is it possible to create a model that utilizes both deep learning and machine learning algorithms for COVID-19 diagnosis. The hypothesis is to use deep learning for feature selection and an ensemble machine learning model for classification. To examine our hypothesis, we performed some experiments and found that ResNet50 is more suitable for the feature selection phase. We performed other experiments to check the appropriate machine learning algorithms that can be grouped to create the ensemble model. Experiments revealed that SVM, MLP, and Logistic Regression surpass the other machine learning models. We performed some preprocessing steps on the dataset before performing the two phases of feature selections and classification. Namely, we applied data augmentation and recoloring of the CT images. Experiments showed that using the original data produces more accurate results. Our final model uses ResNet50 for feature selection, and the selected features are passed to an ensemble model forclassification.

The proposed model is compared with eight state-of-the-art techniques and surpassed them using the accuracy, precision, recall, and F1-score together as a performance measure. RESML is promising in COVID-19 diagnosis, and we claim that it will help in the early detection of +ve cases.

As future work, the proposed model may be adjusted for the diagnosis of other lung diseases using CT scans. For the sake of interpretation, it is better to have a small number of features. So, more deep learning models will be investigated to check its ability to select fewer features for the classification phase. Also, we will apply our model to more COVID-19 datasets. Also, the proposed model can be tested against other classification problems that depend on CT scans, like the diagnosis of breast and braincancer.

Acknowledgments

This project is funded by the Islamic University of Madinah, Kingdom of Saudi Arabia, represented by the Deanship of Scientific Research under Takamol program, project number 95.

References

1. J. Shabbir, T. Anwer, Artificial intelligence and its role in near future, arXiv preprint arXiv:1804.01396(2018).
2. Worldometers.info, Covid-19 coronavirus pandemic. URL https://www.worldometers.info/coronavirus/
3. P. B. van Kasteren, B. van Der Veer, S. van den Brink, L. Wijsman, J. de Jonge, A. van den Brandt, R. Molenkamp, C. B. Reusken, A. Meijer, Comparison of seven commercial rt-pcr diagnostic kits for covid-19, Journal of Clinical Virology 128 (2020) 104412.
4. Y. Fang, Fang y, zhang h, xie j, et al, Sensitivity of chest CT for COVID-19: comparison to RT-PCR. Radiology 200432 (2020).
5. M. Kaur, S. Tiwari, R. Jain, Protein based biomarkers for non-invasive covid-19 detection, Sensing and Bio-Sensing Research 29 (2020) 100362.
6. P. Angelov, E. Almeida Soares, Explainable-by-design approach for covid- 19 classification via ct-scan, medRxiv(2020).
7. T. Ai, Z. Yang, H. Hou, C. Zhan, C. Chen, W. Lv, Q. Tao, Z. Sun, L. Xia, Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases, Radiology 296 (2) (2020) E32–E40.
8. J. Chen, L. Wu, J. Zhang, L. Zhang, D. Gong, Y. Zhao, S. Hu, Y. Wang, X. Hu, B. Zheng, et al., Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: a prospective study, MedRxiv (2020).
9. Y. Gang, X. Chen, H. Li, H. Wang, J. Li, Y. Guo, J. Zeng, Q. Hu, J. Hu, H. Xu, A comparison between manual and artificial intelligence– based automatic positioning in ct imaging for covid-19 patients, European Radiology (2021) 1–10.
10. Y. Zhou, F. Pei, M. Ji, L. Wang, H. Zhao, H. Li, W. Yang, Q. Wang, Q. Zhao, Y. Wang, Sensitivity evaluation of 2019 novel coronavirus (sars- cov-2) rt-pcr detection kits and strategy to reduce false negative, PLOS ONE 15 (11) (2020) 1–12. doi:10.1371/journal.pone.0241469. URL https://doi.org/10.1371/journal.pone.0241469
11. T. Ai, Z. Yang, H. Hou, et al., Correlation of chest ct and rt-pcr testing in coronavirus disease 2019 (covid-19) in china: a report of 1014 cases [e- pub ahead of print], Radiology https://doi. org/10.1148/radiol 2020200642 (2020).
12. D. Chen, X. Jiang, Y. Hong, Z. Wen, S. Wei, G. Peng, X. Wei, Can chest ct features distinguish patients with negative from those with positive initial rt-pcr results for coronavirus disease (covid-19)?, American Journal of Roentgenology 216 (1) (2021)66–70.
13. X. Xie, Z. Zhong, W. Zhao, C. Zheng, F. Wang, J. Liu, Chest ct for typical 2019-ncov pneumonia: relationship to negative rt-pcr testing, Radiology (2020) 200343.
14. J. Bullock, K. H. Pham, C. S. N. Lam, M. Luengo-Oroz, et al., Mapping the landscape of artificial intelligence applications against covid-19, arXiv preprint arXiv:2003.11336(2020).
15. Ulhaq, A. Khan, D. Gomes, M. Paul, Computer vision for covid-19 control: A survey, arXiv preprint arXiv:2004.09420 (2020).
16. S. H. Kassania, P. H. Kassanib, M. J. Wesolowskic, K. A. Schneidera, R. Detersa, Automatic detection of coronavirus disease (covid-19) in x-ray and ct images: A machine learning based approach, Biocybernetics and Biomedical Engineering 41 (3) (2021)

867–879. doi:https://doi.org/10.1016/j.bbe.2021.05.013. URLhttps://www.sciencedirect.com/science/article/pii/S0208521621 00067X
17. P. Afshar, S. Heidarian, N. Enshaei, F. Naderkhani, M. J. Rafiee, Oikonomou, F. B. Fard, K. Samimi, K. N. Plataniotis, A. Mohammadi, COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning 8 (1) 121. doi:10.1038/ s41597-021-00900-3. URL https://www.nature.com/articles/s41597-021-00900-3
18. M. Barstugan, U. Ozkaya, S. Ozturk, Coronavirus (COVID-19) classification using CT images by machine learning methodsarXiv:2003. 09424. URL http://arxiv.org/abs/2003.09424
19. H. Mohammad-Rahimi, M. Nadimi, A. Ghalyanchi-Langeroudi, M. Taheri, S. Ghafouri-Fard, Application of machine learning in diagnosis of COVID-19 through x-ray and CT images: A scoping review 8 185. doi:10.3389/fcvm.2021.638011. URL https://www.frontiersin.org/article/10.3389/fcvm.2021. 638011
20. S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, B. Xu, A deep learning algorithm using CT images to screen for corona virus disease (COVID-19) 2020.02.14.20023028doi:10.1101/2020.02.14.20023028. URLhttps://www.medrxiv.org/content/10.1101/2020.02.14. 20023028v5
21. L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, et al., Artificial intelligence distinguishes covid-19 from community acquired pneumonia on chest ct, Radiology (2020).
22. D. Apostolopoulos, T. A. Mpesiana, Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks, Physical and Engineering Sciences in Medicine (2020)1.
23. C. Qin, D. Yao, Y. Shi, Z. Song, Computer-aided detection in chest radiography based on artificial intelligence: a survey, Biomedical engineering online 17 (1) (2018) 113.
24. 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 (2020).
25. C. Jin, W. Chen, Y. Cao, Z. Xu, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, J. Feng, Development and evaluation of an ai system for covid-19 diagnosis, medRxiv (2020).
26. O. Gozes, M. Frid-Adar, H. Greenspan, P. D. Browning, H. Zhang, W. Ji, A. Bernheim, E. Siegel, Rapid ai development cycle for the coronavirus (covid-19) pandemic: Initial results for automated detection & patient monitoring using deep learning ct image analysis, arXiv preprint arXiv:2003.05037 (2020).
27. S. Serte, H. Demirel, Deep learning for diagnosis of covid-19 using 3d ct scans, Computers in biology and medicine 132 (2021) 104306.
28. M. Barstugan, U. Ozkaya, S. Ozturk, Coronavirus (covid-19) classification using ct images by machine learning methods, arXiv preprint arXiv:2003.09424 (2020).
29. E. Soares, P. Angelov, S. Biaso, M. H. Froes, D. K. Abe, Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification, medRxiv(2020).
30. Zhao, Y. Zhang, X. He, P. Xie, Covid-ct-dataset: a ct scan dataset about covid-19, arXiv preprint arXiv:2003.13865(2020).
31. V. Gudivada, A. Apon, J. Ding, Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations, International Journal on Advances in Software 10 (1) (2017) 1–20.
32. R. M. Pereira, D. Bertolini, L. O. Teixeira, C. N. Silla Jr, Y. M. Costa, Covid-19 identification in chest x-ray images on flat and hierarchical classification scenarios, Computer Methods and Programs in Biomedicine (2020) 105532.
33. Z. Wang, Q. Liu, Q. Dou, Contrastive cross-site learning with redesigned net for covid-19 ct classification, IEEE Journal of Biomedical and Health Informatics 24 (10) (2020) 2806–2813.

Abdulrahman Alhaidari: is a lecturer at the Islamic University of Madina and received his master's degree from the University of Pittsburgh, USA, with a distinction. He is a passionate computer science specialist with a research focus on Machine Learning, IoT, and Cybersecurity. His ultimate goal is to contribute to the ongoing research that will hopefully have a significant impact on the local society and around the globe.

Mustafa Elnainay: received the B.Sc. and M.Sc. degrees in computer engineering from Alexandria University, in 2001 and 2005, respectively, and the Ph.D. degree in computer engineering from Virginia Tech, in 2009. He is Professor of computer engineering with the Computer and Systems Engineering Department, Alexandria University, Egypt. He is currently on leave and affiliated with the Faculty of Computer Science and Engineering, AlAlamein International University. His research interests include wireless and mobile networks, cognitive radio, and cognitive networks, as well as software testing automation and optimization. His focus is on the use of artificial intelligence to solve problems in different domains, including communications and networking, indoor localization, and software engineering. He has served as a Reviewer, TPC member, and TPC chair/track chair for various international journals and conferences.

Emad Nabil: received the B.Sc. degree (Hons.) from the Computer Science Department, in 2004, the M.Sc. degree in soft computing and bio-inspired algorithms, in 2008, and the Ph.D. degree in optimization and machine learning, in 2012. He is currently an Assoc. Prof. at the Faculty of Computers and Artificial intelligence, Cairo University, Egypt. He is also an Assoc. Prof. at the Faculty of Computer and Information Systems, Islamic University of Madinah, Saudi Arabia. He is interested in Machine Learning, Deep Learning, Computer Vision, Natural Language Processing, Optimization, Bio-Inspired Algorithms, Health Informatics, and Bioinformatics.