IoT-enabled stacked ensemble of deep neural networks for the diagnosis of COVID-19 using chest CT scans

Mohammad Shorfuzzaman¹

Received: 23 April 2021 / Accepted: 7 June 2021 / Published online: 21 June 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2021

Abstract

The ongoing COVID-19 (novel coronavirus disease 2019) pandemic has triggered a global emergency, resulting in significant casualties and a negative effect on socio-economic and healthcare systems around the world. Hence, automatic and fast screening of COVID-19 infections has become an urgent need of this pandemic. Real-time reverse transcription polymerase chain reaction (RT-PCR), a commonly used primary clinical method, is expensive and time-consuming for skilled health professionals. With the aid of various AI functionalities and advanced technologies, chest CT scans may thus be a viable alternative for quick and automatic screening of COVID-19. At the moment, significant advances in 5G cellular and internet of things (IoT) technology are finding use in various applications in the healthcare sector. This study presents an IoT-enabled deep learning-based stacking model to analyze chest CT scans for effective diagnosis of COVID-19 encounters. At first, patient data will be obtained using IoT devices and sent to a cloud server during the data procurement stage. Then we use different fine-tuned CNN sub-models, which are stacked together using a meta-learner to detect COVID-19 infection from input CT scans. The proposed model is evaluated using an open access dataset containing both COVID-19 infected and non-COVID CT images. Evaluation results show the efficacy of the proposed stacked model containing fine-tuned CNNs and a meta-learner in detecting coronavirus infections using CT scans.

Keywords Internet of things (IoT) · Chest CT scans · COVID-19 Diagnosis · Stacking model · Fine-tuned CNNs · Deep learning

Mathematics Subject Classification 68 · 68T07

✉ Mohammad Shorfuzzaman
m.shorf@tu.edu.sa

¹ Department of Computer Science, College of Computers and Information Technology, Taif University, Taif 21944, Saudi Arabia

1 Introduction

In recent years, there has been rapid progress in the fields of information technology (IT) and digital electronics, as well as an unprecedented rise in the growth of 5G internet of things (IoT) systems. These technologies can be used to diagnose patients in a variety of ways. With the on-going COVID-19 pandemic, the world has gone through a worldwide emergency causing significant casualties and impact on healthcare and socio-economic structures globally. The virus originally believed to be started in December 2019 in Wuhan, China, has rapidly spread all over the globe. As of December 7, 2020, there are approximately 67 million infected cases and more than 1.5 million mortalities all over the world [7]. The overall infected cases are constantly increasing due to the dearth of proper treatment and vaccines and the excessive rate of public transmission. World Health Organization (WHO) has announced the outbreak as a global pandemic on March 11, 2020 [39]. COVID-19 is triggered by SARS-CoV-2 and is transmitted between individuals primarily through infection due to direct contact. The major symptoms found in COVID-19 patients are highly variable and commonly include fever, cough, shortness of breath, and so on. Individuals without any symptoms can even spread the virus and can stay infectious for a longer period.

Accurate and timely identification of coronavirus infections is very crucial to place infected patients in quarantine and prescribe proper line of treatment. This in turn will facilitate timely restrain of the outbreak and ensure public health and wellbeing. However, the panic for COVID-19 has risen manifold due to the lack of a quick and precise diagnosis system. Consequently, curbing the transmission of the virus has become very challenging.

Detection of COVID-19 is predominantly performed by real-time reverse-transcription polymerase chain reaction (RT-PCR) which works based on the detection of nucleic acid in lower and upper respiratory specimens. The test requires the specimen collection through an oropharyngeal or nasopharyngeal swab or a saliva sample for the detection of viral RNA. However, the use of RT-PCR test is limited due to several reasons. Firstly, it generates a high rate of false negative alarms and a patient initially assessed COVID-19 negative later could be tested positive [8]. Hence, multiple tests may be needed to verify a case which can take a maximum of two days. Secondly, the PCR test kits are short in supply as compared to the global demand due to the overwhelming infection rate. As a result, many coronavirus patients remain unidentified due to time-consuming and manual testing of PCR and are likely to infect others inadvertently. In such cases, alternative testing methods based on AI (Artificial Intelligence) techniques for automatic diagnosis of COVID-19 would be very useful to the clinicians. This will also facilitate the screening of COVID-19 cases on large scale.

Typical symptoms found in patients infected with coronavirus include attacks in different types of lung cells and inflammatory reaction to them. This type of inflammatory reaction can be effectively identified from radiology images such as chest X-Ray (CXR) or chest computed tomography (CT). Earlier studies have demonstrated that radiographic features present in infected chest CT images come as a form of ground-glass opacities (GGO) [6]. These visual elements specific to the novel coronavirus can be used by health practitioners to detect COVID-19 infection with the help of

computer-aided diagnosis. There have been substantial works based on deep learning techniques that use radiology images of different modalities including chest CT and CXR images in diagnosing diseases in the smart healthcare domain. Specifically, several studies [18,25,30,31] in the literature have developed various deep learning-based approaches to detect COVID-19 from the chest CT scan and CXR images and subsequently observe disease progress in the future. While these studies demonstrate promising results, they face several challenges. First, the diversity of COVID-19 radiographic features makes the deep learning models struggle in attaining superior diagnosis accuracy that complies with the clinical standard [35]. Second, the lack of a huge amount of training data typically required by deep learning models poses a serious challenge to models' generalizability to unseen data. It is challenging to collect a large amount of COVID-19 positive training data in this pandemic circumstance. Besides, some studies require manual segmentation of lungs or lesion masks which requires domain knowledge and is time-consuming. Hence, it is important to develop models that are effective with limited training data and do not necessitate domain knowledge in interpreting the diagnosis results.

Motivated by the challenges faced by the earlier studies, in this study, we propose an IoT-enabled integrated stacking ensemble framework which assembles several deep CNN (convolutional neural network) models to speed up the investigation of CT scans in robust diagnosis of COVID-19 patients. At first, patient data will be obtained using IoT devices and sent to a cloud server using 5G networks. In the stacking ensemble approach, model averaging is adopted where multiple sub-models, preferably deep and, are combined to obtain final prediction results. The ensemble model performance can be enhanced by taking the weighted contribution of each sub-model to the stacking model. This can be further improved by training a completely new model to combine the predictions from various individual sub-models in the best possible manner. This method is known as stacked generalization [43,46] which was originally introduced to minimize the generalization error rate for one or more generalizers used on a learning dataset. Thus, the stacked generalization of deep CNN models provides the benefit of harnessing the strengths of a range of models on a prediction task and yields improved classification results than any of the sub-models in the ensemble. Specifically, we use three different fine-tuned CNN models called ResNet50V2, DenseNet121, and Xception as sub-models which are stacked together using a meta-learner for ultimate categorization of COVID-19 encounters from input CT images. Furthermore, we have used chest CT images from a public dataset consisting of 2484 samples to train our stacked ensemble model. Fig. 1 shows some positive and negative COVID-19 samples from the dataset. In summary, this work makes the following contributions:

- We propose an IoT-enabled stacking ensemble framework of deep learning models to facilitate the analysis of chest CT scans in an automatic diagnosis of COVID-19 patients
- We provide the process of leveraging transfer learning capabilities of fine-tuned pre-trained deep CNN models in identifying COVID-19 encounters from non-COVID cases.
- A comparative study is presented to investigate the effectiveness of the stacked ensemble model and individual base CNN sub-models.

Fig. 1 Samples of chest CT images with **a** coronavirus infection and **b** no apparent infection

- We present extensive experimental analysis to demonstrate the performance of the studied models. The proposed stacked model achieves an accuracy of 96.58% in classifying COVID-19 and non-COVID CT images with a high degree of precision (99.16%), specificity (99.16%), and AUC score (96.6%).
- We also show the flexibility of the proposed stacking ensemble model which can simply be integrated with other off the shelf deep learning models to obtain further improvement in diagnosis performance.

In the rest of the paper, we first present recent studies related to our work. Then, we present methodology and dataset description with implementation details in Sects. 3 and 4, respectively. Performance results with discussion are presented in Sect. 5. Lastly, we provide conclusions and future work in Sect. 6.

2 Related studies

In the recent past, deep learning techniques have been widely used in image processing and computer vision applications [24,48]. Specifically, there have been substantial research efforts introducing Internet of Things (IoT) and deep learning techniques for healthcare applications [2,12,14,22,27,32]. The researchers have achieved promising results in diagnosing lung abnormalities from radiology images using latest deep learning approaches. To tackle the challenge of the on-going COVID-19 pandemic, researchers have shown intensive interest in developing deep learning-based systems for automatic diagnosis of COVID-19 using radiology imaging. To this end, this section reviews recently proposed systems that have leveraged deep learning-based methods to detect COVID-19 infections from clinical images such as chest CT scans and CXR.

The fact that CT imaging can aid in the rapid diagnosis of COVID-19 infections is corroborated by several earlier studies [4,17]. Research results from some other studies show evidence that the diagnosis of COVID-19 is effective even for asymptomatic patients [33]. This is achieved by detecting several clinical radiographic features such

as loculated pleural effusion, ground-glass opacities, and consolidation noticed in chest CT scans of COVID-19 patients [15]. Chen et al. [5] presented one of the earliest studies that construct a deep learning-based AI system to detect COVID-19 pneumonia from high-resolution CT images. The model is built using UNet++ [49] which is a very effective architecture for medical image segmentation. The authors used ResNet-50 with all pre-trained (on ImageNet dataset) weights as the backbone of UNet++. Model training and validation were performed using over 46,000 CT images collected from 106 admitted patients. Evaluation results with two different test datasets show the effectiveness of the model with a maximum per-patient accuracy of 95.24% and a reduction of radiologists' reading time by 65%.

A deep learning-based study presented in [47] offers early screening of COVID-19 from healthy and influenza-A viral pneumonia (IAVP) using pulmonary CT scans. The proposed approach started with pre-processing the CT images to extract pulmonary regions and then used a 3D CNN segmentation model to segment multiple candidate infection regions. A location-attention classification algorithm was used to classify these image patches into IAVP, COVID-19, and irrelevant to infection groups. The model finally used the Noisy-OR Bayesian function to calculate the type of infection and confidence score for each image. Model training and evaluation were done using 618 CT images from all three categories mentioned above. The images were obtained from three COVID-19 hospitals in China. Evaluation results demonstrated the effectiveness of the model for early screening of COVID-19 with a modest rate of accuracy (86.7%). In another effort, Mishra et al. [23] proposed a deep learning system to detect COVID-19 in CT images using various off-the-shelf CNN models. They have also proposed a decision fusion approach where predictions from multiple models are combined to find the final prediction result. However, the performance results demonstrate only mediocre detection accuracy (86%) and AUC score (0.883).

Hasan et al. [9] introduced a hybrid system for the classification of COVID-19 patients from CT scans using a combination of automatic and handcrafted features extracted from deep learning models and Q-deformed entropy algorithm, respectively. The curated features are then fed to a long short-term memory (LSTM) classifier to discriminate COVID-19 cases from other pneumonia and healthy cases. The proposed model achieved the highest accuracy of 99.68% using CT images collected from a dataset consisting of 321 patients. In a more recent effort, Harmon et al. [8] proposed an AI-based approach to detect COVID-19 pneumonia using multinational chest CT datasets. The authors have developed a number of deep learning methods and trained them using CT images collected from a multinational cohort of 1280 patients. A lung segmentation model was developed using AH-Net architecture [20] to localize complete lung areas which are then fed to multiple classification models that perform 3D classification using multiple slices at fixed resolution and using one whole volume with fixed size as well. Evaluation results using an independent test dataset of 1337 patients showed that the model can achieve maximum accuracy of 90.8%. Some other works [5, 19] have also developed COVID-19 diagnostic tools using 2D and 3D CNNs based on CT scans. Moreover, some researchers adopted segmentation techniques for the rapid identification of COVID-19 using CT scans [28].

A few other studies have also utilized chest X-Ray for the detection of COVID-19 utilizing deep learning approaches. In one of the earliest open-source efforts, Wang

et al. [45] introduced an AI-based framework called COVID-Net to detect COVID-19 cases from CXR images. The authors also released a publicly available benchmark dataset called COVIDx consisting of 13,975 CXR images from 13,870 patients. They have provided evaluation results both from quantitative and qualitative perspectives. An explainability method was used to show how the model is making decisions for likely COVID-19 infections. However, the drawback of the study is that the dataset used for model training and testing exhibits class imbalance with only a few (approximately 100) COVID-19 images. A coronavirus detection framework based on meta-learning known as MetaCOVID is introduced in [20] for COVID-19 detection using n-shot classification. The authors have introduced a collaborative method to extract CNN based features with contrastive loss and then used a Siamese neural network for final prediction. Evaluation results demonstrated the effectiveness of the approach with a limited training dataset.

In another recent effort, Horry et al. [11] suggested a COVID-19 detection framework using multimodal image data with transfer learning. The authors have used images of three different modalities such as Ultrasound, CXR, and CT scans. A pre-processing pipeline for image data was developed to apply histogram equalization to images to reduce the effect of sampling bias by using N-CLAHE method. Test results demonstrated that the framework can achieve high detection accuracy using VGG-19 transfer learning model. Islam et al. [16] suggested a deep learning-based approach by combining the power of long short-term memory (LSTM) with CNN to detect coronavirus infection using chest X-Ray images. In the study, the authors have used a dataset consisting of 4575 CXR images including 1525 COVID-19 positive images. Findings from the model evaluation showed that the suggested hybrid architecture outperforms a CNN model with high accuracy (99.4%) and specificity (99.2%). Hossain et al. [13] proposed an explainable AI-based secured framework to control the ongoing pandemic. The framework leverages the low-latency and high bandwidth features of 5G network for the identification of infected cases from CXR and CT images. Three transfer learning models that are known as ResNet50, Deep Tree, and InceptionV3 were used to assess the efficacy of the proposed framework. Similarly, some other works [4,17] have developed COVID-19 diagnostic tools using different deep learning techniques and CXR images. Some other studies [3,40] also compared the prediction performance of using both CT scans and CXR images in COVID-19 diagnosis.

Besides, a few research efforts [37,45] have contributed by developing models with interpretability results for wider acceptability of the models among front line clinical professionals. Yet, some research efforts [18,21,26,29] introduced the privacy-aware energy-efficient framework for data collection, data fusion, visualization, and secure communication in COVID-19 application environments.

Current studies in the literature primarily use off-the-shelf or custom CNN models for the diagnosis of COVID-19 patients from chest CT scans and CXR images. On the contrary, we propose in this study an IoT-enabled deep learning stacked ensemble model combining several fine-tuned CNN models to minimize the generalization error rate for one or more generalizers used on a learning dataset. Hence, the proposed stacked generalization of deep CNN models offers the benefit of combining the

strengths of a range of models to produce better performance results than any of the sub-models in the ensemble for effective COVID-19 screening.

3 Methodology

The section begins with a formal problem definition for the classification task at hand with IoT-enabled stacked ensemble architecture. Subsequently, we explain different elements of our proposed stacking model with their core technology to precisely comprehend the complete detection process.

3.1 Problem definition

Stacked generalization refers to the method of using a high-level model also called meta-learner to combine multiple lower-level models for improved final prediction. Specifically, the stacked ensemble of multiple CNN models permits us to combine the capability of each model (e.g., wide and deep) that has been trained for a particular task such as classification. Typically, different models learn for the classification task at hand and the outputs of these models are first collected to form a new dataset that contains the prediction probabilities of each model for every instance of data in the original training set. This new dataset is considered as the data for a second learning problem which is solved by using a second learning model called meta-learner. Thus, the original data and the models used in the first step are referred to as “level-0 data” and “level-0 models” (or sub-models), respectively. Likewise, the cross-validated data and the learning model in the second step are called “level-1 data” and “level-1 generalizer” (or meta-learner). Finally, we are going to have one stacked multi-headed system which is intended to function dependably for the categorization of unseen CT scans. An example of our stacked generalization problem for CT scan-based COVID-19 classification is given in Fig. 2.

Provided a CT scan dataset, $C = \{(x_n, y_n), n = 1, \dots, N\}$, where x_n and y_n represent the n -th input image and its target class, respectively. Dataset is split into K equal parts $C_1, C_2, \dots, C_K$. Moreover, let $C^{(k)}_{test}$ and $C^{(k)}_{train}$ be test and training set for the k -th fold of K -fold cross-validation where $C^{(k)}_{train} = C - C^{(k)}_{test}$. Additionally, we presume L overall deep learning sub-models (level-0 models) where l -th model denoted by M_l , for $l = 1, \dots, L$ is invoked on the training dataset. Considering every single image, x_n , in C , let $p_{li}(x_n)$ is the probability for i -th class label generated by model M_l and the vector of probabilities generated by this sub-model can be denoted as below where t refers to the total number of class labels:

$$P_{ln} = [p_{l1}(x_n), p_{l2}(x_n), \dots, p_{lt}(x_n)], 1 \leq l \leq L \quad (1)$$

Now, given the vector of probabilities generated by each sub-model for a data instance, x_n , we combine them for all L sub-models with the actual class label as below:

$$C_{cv} = [y_n, P_{1n}, P_{2n}, \dots, P_{Ln}, \dots, P_{Ln}] \quad (2)$$

Fig. 2 Example of stacked generalization problem

Ultimate outcomes of classification are attained from C_{cv} by using the level-1 meta learner, M_{meta} :

$$p_{final} = M_{meta}(C_{cv}) \quad (3)$$

3.2 Proposed IoT-enabled stacking CNN model

We propose an IoT-enabled deep learning framework (as shown in Fig. 3) for the diagnosis of COVID-19 from chest CT scans. It includes several components, including chest CT image capturing from mobile CT scanners, cloud deployment of a stacked ensemble model, large-scale chest CT scan collection for online model training, and results from inference. In the current study, we particularly focus on the stacked ensemble CNN model that is an integral part of our proposed IoT framework. Figure 4 shows the block diagram of our proposed ensemble model for an automatic diagnosis of COVID-19 cases. Initially, prediction probabilities are generated from validation fold CT images using three different fine-tuned CNN sub-models called ResNet50V2 [10], DenseNet121 [41], and Xception [38] which are stacked together at level-0. Later, we combine the predictions generated from these networks and feed them to a meta-learner at level-1 for the final classification of COVID-19 cases. Finally, we investigate how good are the feature representations that are obtained from these networks using the t-SNE visualization technique. A thorough explanation of the system is provided in the next subsections.

Fig. 3 Proposed IoT-cloud framework for an automatic diagnosis of COVID-19 from chest CT scans

Fig. 4 Blocked diagram of proposed integrated stacked CNN model (Please zoom out for superior view)

3.2.1 CNN sub-model architectures and fine-tuning

In the proposed stacking ensemble system, we have utilized the above stated off-the-shelf CNN models and fine-tuned them for the generation of level-0 prediction probabilities from an input validation fold CT images. We use ResNet CNN as our first sub-model in the stacked architecture. It is observed that conventional sequential deep learning models face vanishing gradient problems where accuracy gets saturated at some point and decreases unexpectedly with an additional increase in depth. ResNet model deals with this issue by bypassing through less essential layers with the assis-

tance of residual units during model training using a regular SGD optimizer. In our ensemble network, we have used ResNet50V2 consisting of 50 weight layers with a substantial drop in model size as well as the FLOP count. Our second model is Xception which is an extreme version of its pioneer called Inception [42]. The idea is to deal with each output channel separately by using a mapping of spatial correlations. Further inter-channel correlation is captured by performing 1×1 convolutions. Finally, we use the DenseNet121 CNN model which is densely connected and requires fewer parameters as compared to a traditional CNN. Unlike ResNets, DenseNets have extremely constricted layers and use simply twelve filters along with a very few feature-maps. DenseNet121 can find gradient values directly from the loss function which improves the time required for training. This significantly reduces computation cost and makes this a superior option.

As part of fine-tuning, we delete the classifier part of the transfer learning (CNN) models and include our custom prediction layer which consists of a global average pooling layer (GAP) followed by dual fully connected (FC) layers consisting of 256 neurons and a single neuron, respectively. As opposed to a flattening layer, a GAP layer can better address the overfitting problem by lowering the volume of parameters used in the model. In global average pooling, a feature map with dimension $h \times w$ is converted to a single value by computing the mean of all the pixel values in the feature map and thus obtains $1 \times 1 \times d$ tensor from a 3-D tensor with dimension $h \times w \times d$.

Furthermore, we avoid re-training the CNN models completely by partially fine-tuning and updating the weights of the pre-trained layers. In connection with this, hyperparameter tuning is done by appropriately choosing the learning rate and optimizer to reduce binary cross-entropy loss. We decide to re-train one-third of the upper-level convolutional layers since they learn features that are mostly related to the target classification task while layers in the lower level in the networks are generally believed to learn common features. Thus, we obtain fine-tuned CNN sub-models to be used in the stacked ensemble which requires less training time yet shows better performance.

As commonly used, we have leveraged Adam optimizer with mini-batch to train the entire stacked model. In summary, our stacked model works as a single deep learning model which is multi-headed to accept identical input images from training data. Intermediate prediction probability vectors generated from the first level CNN sub-models are combined and fed through a meta-learner (in the second level) for final classification of the input images. The complete training, validation, and testing of the stacked model are done using a cross-validation technique. Algorithm 1 summarizes the complete integrated stacked mechanism.

Algorithm 1: Integrated stacked ensemble network for categorization of chest CT scans

Input: Training data $C = x_i, y_i \mid 1 \leq i \leq N$, Test data $C_{holdout}$, CNN sub-models

Output: Results from a stacking ensemble model after classification

for $k = 1$ to $k\text{-fold}$ **do**

 Split C into C_k^{train} , C_k^{valid} for k -th fold

 Produce prediction probability vectors from CNN sub-models (L):

for $l = 1$ to L **do**

```

    Make predictions  $P^{(l)}$  based on  $C_k^{train}, C_k^{valid}$ 
end for
 $P^{(s)} = \text{Concatenation}([P^{(1)}, P^{(2)}, \dots, P^{(L)}])$ 
    Build a new dataset,  $C_{cv}$  comprising the probability scores and class
    labels:
    for  $i = 1$  to  $N$  do
         $C_{cv} = \{P_i^{(s)}, y_i\}$ 
    end for
    Learn a meta-classifier,  $M_{meta}^{(k)}$  based on the newly built dataset,  $C_{cv}$ 
    Validate  $M_{meta}^{(k)}$  with  $C_k^{valid}$ 
end for
    Perform classification using hold-out test data:
    output = classify ( $M_{meta}, C_{holdout}$ )
return output

```

3.2.2 Feature representation

To support qualitative analysis, we investigate how well the features are distributed in the feature space to understand the class separability. Since convolutional layers produce high dimensional output, we need to adopt a dimensionality reduction technique to visualize them in 2D space. To achieve this, we use t-SNE (t-Distributed Stochastic Neighbor Embedding) [44] which is a popular technique for exploring and reducing high dimensional data. t-SNE does this by calculating affinities between data points and preserving these affinities in the reduced low-dimensional space.

Let X be a matrix consisting of all the samples in the dataset, and Y be a target matrix containing the low-dimensional representation. The similarity between two data points in the original high dimensional space can be expressed as a conditional probability:

$$P_{j|i} = \exp\left(\frac{-||x_i - x_j||^2}{2\sigma^2}\right), \text{normalized s.t. } \forall i \sum_k P_{k|i} \quad (4)$$

The affinity metric can be obtained by using a symmetric variant of Equation (4) where the affinity of U to V and V to U are the same:

$$P_{ij} = P_{i|j} + P_{j|i}, \text{normalized s.t. } \sum_i \sum_j P_{ij} = 1 \quad (5)$$

Similarly, affinities in low-dimensional space are calculated considering a student- t distribution for d dimensions as follows:

$$Q_{ij} = \left(1 + \frac{||y_i - y_j||^2}{d-1}\right)^{-\frac{d}{2}}, \text{normalized s.t. } \sum_i \sum_j Q_{ij} \quad (6)$$

Table 1 Distribution of samples in the dataset from both categories of CT scans

Class	No. of samples training (60%)	Validation (20%)	Testing (20%)	Total (100%)
COVID-19	752	250	250	1252
Non-COVID	738	246	246	1230

Given the affinities for every pair of data points both in high and low dimensional spaces, the goal is to keep them closer as much as possible. A loss function is used to estimate the distances between the similarities. T-SNE uses Kullback-Leibler divergence as a loss function since the similarities are defined using probabilities:

$$KL(P||Q) = \sum_i \sum_j P_{ij} \log \frac{P_{ij}}{Q_{ij}} \quad (7)$$

4 Dataset description and implementation details

In this section, we begin with the description of the dataset used in the study followed by the implementation details of our proposed stacking ensemble network. There are research efforts in the literature [3] that suggest that CT scans show better prediction performance than CXR images in diagnosing COVID-19 [47]. Hence, we decided to use CT images instead of using CXR or other types of image data. We use a publicly available SARS-CoV-2 CT scan dataset which includes 1252 CT images collected from coronavirus infected patients and 1230 CT scans from individuals that are not infected by the coronavirus. The data is obtained from a hospital in the city of Sao Paulo, Brazil. Overall, the dataset contains 2482 CT images of patients from both categories such as COVID-19 and non-COVID. A 5-fold cross-validation method is used to validate the stacked ensemble model as well as individual sub-models. Hence, we obtain five equal parts of the dataset in the image level that are used in the cross-validation process. The distribution of samples in the dataset for training (60%), validation (20%), and test (20%) is shown in Table 1. Training and validation datasets are used for cross-validation during model training while the performances of the studied models are evaluated using the test set.

4.1 Pre-processing

Given that the samples were gathered at various times with different medical setups, the quality of the images differs significantly. Nonetheless, we avert performing substantial pre-processing of the input CT images to obtain enhanced model generalizability. Consequently, this causes our ensemble network to be even more powerful to artifacts as well as impurities contained in the images while computing salient features from the images. Hence, we have just utilized several basic pre-processing jobs such as resizing, normalization, and augmentation of images to improve model training. The dataset

Table 2 Model configuration and augmentation features

Parameters	Values
Total Epochs	50
Size of batch	32
Optimizer	Adam
Loss function	Binary crossentropy
Learning rate (initial)	0.001
Range of rotation	15
Shear range	10%
Zoom range	10%
Width and height shifting	10%
Flip (horizontal)	Yes
Fill mode	Nearest
Re-scaling	1/255

contains images with various dimensions ranging from 365×465 to 1125×859 pixels. Therefore, all the images in the complete dataset are re-scaled to a unique dimension of 64×64 . Furthermore, we carry out image normalization which refers to the process of changing the range of pixel values and can accelerate model convergence by removing attribute biases as well as achieving a dataset with uniform distribution. Min-max scaling method is used to rescale the pixel values to the range of $[0, 1]$. Lastly, we use image augmentation to deal with the problem of the limited size of the dataset as well as to improve performance while ensuring that the model does not overfit. Table 2 provides a list of features that are used for image augmentation during model training.

4.2 Implementation details

We have used TensorFlow to implement the stacked ensemble model and the pre-trained CNN sub-models. In particular, the functional API from Keras which offers more flexibility in building complex models with multiple inputs or outputs is used to create the ensemble network. Additionally, we use a NumPy utility function to store the input CT images in a compressed file. We take the benefit of free GPU offered by Google Colab for model training and performance evaluation. All the necessary packages are pre-installed and come with Jupyter notebook environment.

After compiling the sub-models into a multi-headed single deep model using Keras functional API, a dense layer with ReLU activation function and with 256 neurons is added to the stacked model. Lastly, prediction results are generated using a dense layer and sigmoid activation function at the end of the model. We apply the commonly used loss function also known as binary cross-entropy for model learning which facilitates faster model convergence. Furthermore, we have used Adam optimizer for model training and validation. Adam is an adaptive learning rate optimization algorithm particularly designed for training deep neural network models. It can be considered as a combination of RMSProp and SGD (Stochastic Gradient Descent) with momentum

where it leverages the squared gradients to scale the learning rate as done in RMSprop and exploits the benefit of momentum by utilizing the moving average of the gradient. This gives Adam optimizer a performance boost over other learning schedules. We set the initial learning of 0.001 for Adam optimizer. The subsequent decay is calculated by dividing the initial learning rate by the total number of epochs to update the learning rate during the training process. During the model training, the performance is monitored and the model with the best performance is saved based on some validation metrics. To achieve this, we use a Keras callback known as *ModelCheckpoint*.

For model evaluation, we include accuracy, sensitivity, specificity, precision, F1-score, and AUC score in our performance metric. Accuracy refers to the proportion of instances where the prediction labels and the ground truth labels are the same. Sensitivity or recall explains how good the model is at detecting positive cases from all the true positive instances in the dataset. On the contrary, specificity explains how good the model is at detecting negative cases from all the true negative instances in the dataset. For instance, the proportion of the healthy population which are accurately classified as showing COVID-19 negative. Precision explains how many of the positively identified cases are relevant. F1-score is calculated as a harmonic mean of both precision and sensitivity. Lastly, AUC (Area Under Curve) explains how good the model is at separating the true positive and negative cases.

5 Evaluation results and discussion

Experimental results obtained from the evaluation of the proposed stacked ensemble model and all the CNN sub-models are presented in this section. First, a quantitative evaluation is performed by comparing all the studied sub-models with the ensemble model. Then, we perform a qualitative evaluation representing extracted features to investigate how good are the features that are generated by various sub-models used in the stacked architecture.

5.1 Quantitative results

We perform a set of experiments for the performance evaluation of our stacked ensemble model as well as to compare its performance with other studied pre-trained sub-models. Table 3 provides the overall performances of all models using the test dataset. Besides, performance results for each class obtained from all the studied models using the same metrics are also presented in Table 4. The proposed ensemble technique constantly attains the best performance as compared to the sub-models in terms of accuracy, precision, specificity, and area under the curve (AUC). The stacked ensemble network attains accuracy and specificity of 96.58 and 99.16%, respectively. These two performance measurement

criteria are very important to assess the effectiveness of any diagnostic method in medical settings. The advantage of combining prediction probabilities from different fine-tuned CNN sub-models in the stacked ensemble framework is evident from this result. The high value of specificity (99.16%) obtained from the model evaluation

Table 3 Performance results obtained from ResNet50V2, Xception, DenseNet121, and the proposed stacked model using the test dataset

Model	Acc.	Prec.	Rec.	Spec.	F1-score	AUC
ResNet50V2	95.87	94.98	98.01	94.72	96.47	96.40
Xception	94.97	92.80	97.61	92.28	95.15	94.90
DenseNet121	93.96	91.39	97.21	90.65	94.21	91.90
Stacked model	96.58	99.16	94.02	99.16	96.52	96.60

Table 4 Class-wise performance results for all the studied models using the test dataset

Model	Class	Accuracy	Precision	Recall	F1-score
ResNet50V2	COVID-19	98.00	93.0	98.0	96.0
	Non-COVID	93.69	98.0	93.0	95.0
Xception	COVID-19	97.60	93.0	98.0	95.0
	Non-COVID	92.27	97.0	92.0	95.0
DenseNet121	COVID-19	97.21	91.0	97.0	94.0
	Non-COVID	90.65	97.0	91.0	94.0
Stacked model	COVID-19	94.02	99.0	94.0	97.0
	Non-COVID	99.18	94.0	99.0	97.0

implies the strength of our model at avoiding false alarms. In addition, the high precision (99.16%) of our ensemble model signifies that the positive COVID-19 cases are classified with high relevance. Nevertheless, amongst the individual sub-models ResNet50V2 outperforms other models in terms of all evaluation metrics on the hold-out test dataset. Surprisingly, ResNet50V2 demonstrates superior sensitivity (98.01%) as compared to the proposed ensemble model (94.02%).

We also observe the impact of the stacked ensemble model over individual CNN sub-models using class-specific results as shown in Table 4. It is noticed that the sub-models exhibit comparatively poor performance in categorizing non-COVID samples though demonstrate average performance in classifying COVID-19 positive samples. More specifically, the sub-models achieve better recall and accuracy scores in diagnosing infected cases. As expected, the proposed stacked model makes use of the advantage of a mix of several fine-tuned CNN sub-models by maintaining the power of ResNet50V2 to make up for the limitations of Xception and DenseNet121 in enhancing the classification results. These results are supposedly significant considering that precisely categorizing CT images for both the subject groups (COVID-19 and non-COVID) are truly crucial for a reliable diagnostic tool.

As per the learning curves, the studied models demonstrate a moderate learning process throughout the training duration by sustaining a consistent decline in both training and validation losses. Furthermore, training and validation in the integrated stacked model as shown in Figs. 5 and 6 appear to converge much better than the CNN sub-models considering the identical length of epochs. Despite the fact that our dataset consists of minimal instances, the learning curves tend to demonstrate that the models

Fig. 5 Training and validation losses for all the models

are not vulnerable to overfitting. This is largely achieved due to the generalizability of the stacked model, data augmentation, and the usage of dropout technique as a regularization employed to the stacked model.

To obtain a greater perception of the effectiveness of the studied models, we provide the receiver operating characteristic (ROC) curve and confusion matrix for all the models in Fig. 7 and Table 5, respectively. The ability of a model for separating the true positive and negative cases is manifested by a ROC curve where the true positive rate (TPR) is plotted against the false positive rate (FPR). Our stacked ensemble model outperforms other sub-models and achieves a mean AUC (area under the curve) value (as shown in Table 3) of 0.966 for both target labels. The individual CNN sub-models exhibit similar classification performance for both COVID and non-COVID classes having somewhat lower AUC scores than the stacked model. It is noticed that the ensemble model produces incredibly few false positive (FP) infected cases as opposed to the individual CNN sub-models. The reduced FP count implies that the number of incorrectly identified infected encounters is less and it improves precision and specificity rates. It is crucial to reduce the FP to avoid unnecessary monetary burdens on healthcare providers. However, the reduced count of FP is obtained with the expense of a relatively increased number of FN which causes a slight decrease in sensitivity obtained by the ensemble model as shown in Table 3. All the sub-models generate

Fig. 6 Training and validation accuracies for all the models

Fig. 7 ROC curves for the stacked model and various CNN sub-models

relatively fewer FN cases but with increased FP counts. In practice, keeping the count of FN cases low is important since incorrectly identifying a COVID-19 patient as healthy will severely hinder proper treatment for the patient. The proposed integrated stacked model makes a trade-off between the number of FP and FN cases and thus

Table 5 Confusion matrix for various evaluated models with the test dataset

Model	TP	FP	TN	FN
ResNet50V2	246	15	231	5
Xception	245	19	227	6
DenseNet121	244	23	223	7
Stacked model	236	2	244	15

offers a fair diagnosis performance. From the overall evaluation results, we conclude that the proposed stacked ensemble model emerges as the best performing among all the studied models.

5.2 Feature representation for CNN sub-models

For a better understanding of the class separability of individual CNN sub-models, we also investigate how well the features generated by these models are distributed. A dimensionality reduction technique is necessary to visualize the high dimensional output produced from convolutional layers. As stated earlier, we utilize t-SNE (t-Distributed Stochastic Neighbor Embedding) [16] which is a nonlinear method for dimensionality reduction to prepare high-dimensional data to visualize them in a low dimensional space. As opposed to PCA (Principal Component Analysis), t-SNE is a nonlinear technique that models similar features using nearby points and dissimilar features using distant points with high probability.

Figure 8 illustrates the representation of features extracted from test CT images by various CNN sub-models in level-0 using t-SNE. It is noticed that ResNet50V2 shows superior feature representation in comparison with other sub-models by displaying a strong separation between image features belonging to different classes. It is interesting to note that, the Xception model occupies a comparatively dense space for feature representation. However, all the sub-models show an area of overlap between COVID-19 and non-COVID target classes.

In summary, a performance comparison between the proposed integrated stacked model and individual fine-tuned CNN sub-models is presented in this study. It is necessary to point out that the CT scan dataset used in this work is very minimal since the process of collecting a large number of open access CT images during this ongoing pandemic is still in its initial phases.

Provided the volume of work accomplished until now for the automated screening of coronavirus cases utilizing deep learning systems, the contribution of AI in supporting frontline health practitioners for effective and rapid diagnosis of COVID-19 can easily be realized. This research serves as one step towards a clearer comprehension of the characteristics of ongoing pandemic and offers a sophisticated deep learning-based solution for effective and rapid identification of COVID-19 cases. Nevertheless, we can positively advocate that our proposed stacked ensemble network is in no way a substitute for a human health practitioner but instead we anticipate that our experimental results provide a useful contribution towards an increasing acceptance of AI-based diagnostic tools in medical settings. Although we cannot merely depend

Fig. 8 Representation of features utilizing t-SNE in 2D space considering both predicted classes using CNN sub-models. **a** ResNet50V2, **b** Xception, **c** DenseNet121

on the diagnosis results found from CT scans to recommend the treatment plan for a patient, initial testing can support clinicians to isolate positive cases till a thorough checkup is completed.

6 Conclusion

In this study, we propose an IoT-enabled end-to-end integrated stacked deep learning method to precisely detect COVID-19 encounters using CT images. Initially, patient data are obtained using IoT devices and sent to a cloud server using 5G networks. Specifically, we develop a stacked ensemble model that exploits the benefit of a combination of multiple deep CNN models to speed up the assessment of chest CT images in automated screening of COVID-19 patients. We use three different fine-tuned CNN models called ResNet50V2, DenseNet121, and Xception as sub-models which are stacked together using a meta-learner for final categorization of COVID-19 encounters from input CT images. Our proposed stacked model attains an accuracy of 96.58% for the categorization of COVID-19 and non-COVID CT images with a high value of specificity. Besides, the stacked model exhibits a superior AUC (0.966) value in contrast with all other studied CNN sub-models suggesting its strong ability to dis-

tinguish between the COVID-19 positive and negative cases. In the future, we plan to use a curated dataset of CT images containing more than two classes to better generalize the model's ability to diagnose potential COVID-19 cases. Furthermore, we intend to obtain better prediction results by utilizing the segmented lung area through state-of-the-art segmentation networks.

Acknowledgements This work was supported by the Taif University Researchers Supporting Project number (TURSP-2020/79), Taif University, Taif, Saudi Arabia.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Abdulsalam Y, Hossain MS (2020) COVID-19 networking demand: an auction-based mechanism for automated selection of edge computing services. *IEEE Trans Netw Sci Eng.* <https://doi.org/10.1109/TNSE.2020.3026637>
2. Alamri A, et al (2014) Evaluating the impact of a cloud-based serious game on obese people. *Comput Hum Behav* 30:468–475.
3. Benmalek E, Elmhamdi J, Jilbab A (2021) Comparing CT scan and chest X-ray imaging for COVID-19 diagnosis. *Biomed Eng Adv.* 1:100003. <https://doi.org/10.1016/j.bea.2021.100003>
4. Bernheim A, Mei X, Huang M, Yang Y, Fayad ZA, Zhang N, Diao K, Lin B, Zhu X, Li K et al (2020) Chest CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. *Radiology*, pp 200463
5. Chen J, Wu L, Zhang J et al (2020) (2020) Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography. *Nature Sci Rep* 10:19196. <https://doi.org/10.1038/s41598-020-76282-0>
6. Chung M, Bernheim A, Mei X et al (2020) CT imaging features of 2019 novel coronavirus (2019-ncov). *Radiology* 295(1):202–207
7. COVID-19 dashboard, coronaBoard, URL: <https://coronaboard.com/>. Accessed Dec 07, 2020
8. Harmon SA, Sanford TH, Xu S et al (2020) Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun* 11:4080. <https://doi.org/10.1038/s41467-020-17971-2>
9. Hasan AM, Al-Jawad MM, Jalab HA et al (2020) Classification of COVID-19 coronavirus, pneumonia and healthy lungs in CT scans using Q-Deformed entropy and deep learning features. *Entropy* 22(5):517
10. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 770–778
11. Horry MJ et al (2020) COVID-19 detection through transfer learning using multimodal imaging data. *IEEE Access* 8:149808–149824
12. Hossain MS, Muhammad G (2018) Emotion-aware connected healthcare big data towards 5G. *IEEE Internet Things J* 5(4):2399–2406
13. Hossain MS, Muhammad G, Guizani N (2020) Explainable AI and mass surveillance system-based healthcare framework to combat COVID-19 like pandemics. *IEEE Netw* 34(4):126–132
14. Hu L, Qiu M, Song J, Hossain MS, Ghoneim A (2015) Software defined healthcare networks. *IEEE Wireless Commun* 22(6):67–75
15. Huang C, Wang Y, Li X et al (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 395(10223):497–506
16. Islam MZ, Islam MM, Asraf A (2020) A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Inf Med Unlocked* 20:100412. <https://doi.org/10.1016/j.imu.2020.100412>
17. Li Y, Xia L (2020) Coronavirus disease 2019 (COVID-19): role of chest CT in diagnosis and management. *Am J Roentgenol* 2020:1–7

18. Lin H et al (2020) Privacy-enhanced data fusion for COVID-19 applications in intelligent Internet of edical things. *IEEE Internet Things Journal*. <https://doi.org/10.1109/JIOT.2020.3033129>
19. Li L, Qin L, Xu Z et al (2020) Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT. *Radiology*: 200905
20. Liu S et al (2018) 3D anisotropic hybrid network: Transferring convolutional features from 2D images to 3D anisotropic volumes. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G (eds) *Medical image computing and computer assisted intervention—MICCAI 2018*. MICCAI 2018. *Lecture Notes in Computer Science*, vol 11071. Springer, Cham. https://doi.org/10.1007/978-3-030-00934-2_94
21. Long Z, Alharthi R, Saddik AE (2020) NeedFull—a tweet analysis platform to study human needs during the COVID-19 pandemic in New York state. *IEEE Access* 8:136046–136055
22. Masud M, Hossain MS, Alamri A (2012) Data Interoperability and Multimedia Content Management in e-Health Systems. *IEEE Trans Inf Technol Biomed* 16(6):1015–1023.
23. Mishra AK, Das SK, Roy P, Bandyopadhyay S (2020) Identifying COVID19 from chest CT images: A deep convolutional neural networks based approach. *J Healthcare Eng*, 8843664. <https://doi.org/10.1155/2020/8843664>
24. Hossain MS, Muhammad G (2019) Emotion recognition using deep learning approach from audio-visual emotional big data. *Inf Fusion* 49:69–78
25. Muhammad G, Hossain MS (2021) COVID-19 and non-COVID-19 classification using multi-layers fusion from lung ultrasound images. *Inf Fusion* 72:80–88
26. Muhammad G, Hossain MS (2021) A deep learning-based edge-centric COVID-19-like pandemic screening and diagnosis system within a B5G framework using blockchain. *IEEE Netw* 35(2):74–81
27. Muhammad G, Hossain MS, Kumar N (2021) EEG-based pathology detection for home health monitoring. *IEEE J Select Areas Commun* 39(2):603–610
28. Ouyang X, Huo J, Xia L et al (2020) Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans Medical Imag* 39(8):2595–2605
29. Rahman MA, Hossain MS (2020) An Internet of medical things-enabled edge computing framework for tackling COVID-19. *IEEE Internet Things J*. <https://doi.org/10.1109/JIOT.2021.3051080>
30. Rahman MA et al (2020) B5G and explainable deep learning assisted healthcare vertical at the edge: COVID-19 perspective. *IEEE Netw* 31(4):98–105
31. Rahman MA et al (2021) A multimodal, multimedia point-of-care deep learning framework for COVID-19 diagnosis. *ACM Trans Multimedia Comput Commun Appl* 17:24
32. Saddik AE, Badawi H, Velazquez R, Laamart F et al (2019) Dtwins: a digital twins ecosystem for health and well-being. *IEEE COMSOC MMTC Commun Front* 14(2):39–46
33. Salehi S, Abedi A, Balakrishnan S, Gholamrezaezhad A (2020) Coronavirus disease 2019 (COVID-19): a systematic review of imaging findings in 919 patients. *Am J Roentgenol* 215:87–93
34. SARS-COV-2 Ct-Scan Dataset, <https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset>. Accessed 05 Sep 2020
35. Shi H, Han X, Jiang N et al (2020) Radiological findings from 81 patients with covid-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Inf Dis* 20(4):425–434
36. Shorfuzzaman M (2021) Hossain MS (2021) MetaCOVID: a Siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. *Pattern Recog* 113:107700
37. Shorfuzzaman M, Masud M (2020) On the detection of COVID-19 from chest X-ray images using CNN-based transfer learning. *Comput Mater Contin* 3:1359–1381
38. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: *3rd International conference on learning representations (ICLR 2015)*
39. Statement on the second meeting of the international health regulations (2005) In: *Emergency committee regarding the outbreak of novel coronavirus (2019-nCoV)*. World Health Organization. 30 Jan 2020. Archived from the original on 31 Jan 2020. Accessed 10 Aug 2020
40. Sverzellati N, Ryerson CJ, Milanese G et al (2021) Chest x-ray or CT for COVID-19 pneumonia? Comparative study in a simulated triage setting. *Eur Resp J* 57(5). <https://doi.org/10.1183/13993003.04188-2020>
41. Szegedy C, Liu W, Jia Y et al (2015) Goingd deeper with convolutions. In: *Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 1–9
42. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 2818–2826

43. Ting K, Witten IH (1999) Issues in stacked generalization. *J Artif Intell Res* 10(1):271–289
44. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9(2605):2579–2605
45. Wang L, Lin ZQ, Wong A (2020) COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Sci Rep Nature* 10:19549
46. Wolpert DH (1992) Stacked generalization. *Neural Netw* 5(2):241–259
47. Xu X, Jiang X, Ma C et al (2020) A deep learning system to screen coronavirus disease 2019 pneumonia. *Engineering* 6(10):1122–1129
48. Zhou Y, Dong H, Saddik AE (2020) Deep learning in next-frame prediction: a benchmark review. *IEEE Access* 8:69273–69283
49. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2018) Unet++: A nested u-net architecture for medical image segmentation. In: *Deep Learn Med Image Anal Multimodal Learn Clin Decision Support*, pp 3–11

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