A lightweight capsule network architecture for detection of COVID-19 from lung CT scans

Shamik Tiwari | Anurag Jain

School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India

Correspondence
Anurag Jain, School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India. Email: anurag.jain@ddn.upes.ac.in

Abstract
COVID-19, a novel coronavirus, has spread quickly and produced a worldwide respiratory ailment outbreak. There is a need for large-scale screening to prevent the spreading of the disease. When compared with the reverse transcription polymerase chain reaction (RT-PCR) test, computed tomography (CT) is far more consistent, concrete, and precise in detecting COVID-19 patients through clinical diagnosis. An architecture based on deep learning has been proposed by integrating a capsule network with different variants of convolution neural networks. DenseNet, ResNet, VGGNet, and MobileNet are utilized with CapsNet to detect COVID-19 cases using lung computed tomography scans. It has found that all the four models are providing adequate accuracy, among which the VGGCapsNet, DenseCapsNet, and MobileCapsNet models have gained the highest accuracy of 99%. An Android-based app can be deployed using MobileCapsNet model to detect COVID-19 as it is a lightweight model and best suited for handheld devices like a mobile.

KEYWORDS
CapsNet, COVID-19, deep learning, DenseNet, lung CT scan, MobileNet, ResNet, VGG16

1 INTRODUCTION

The COVID-19 virus that was discovered in November 2019 is responsible for a global pandemic. This virus has a genetic similarity with the SARS virus discovered in 2002. This outbreak had begun in China and now spread all over the world. This virus was generated in bats, and through them, it had entered the human body. This virus can spread in the air through water droplets coming out from diseased lungs while sneezing, speaking, or cough and remain active for up to 3 h. The virus can also survive on surfaces of objects. On the aluminum surface, it can survive for 8 h. On cardboard, it can last for 24 h. On plastic and stainless steel, it can live for 3 days, while on wood and glass, it can live for 5 days. People aged above 60, children below 12 years of age, asthmatic patients, people having a weak immune system, and pregnant women are more prone to COVID-19.1,2 COVID-19 symptoms include a high fever, exhaustion, shortness of breath, cough, and a loss of taste and smell. This causes serious lung damage later on, which is referred to as acute respiratory distress syndrome (ARDS) in medical terminology. Symptoms of COVID-19 in a person are developed after 5 days of infection. These 5 days are called the incubation period, and during this period, COVID-19 diseased lung becomes a moving source of infection. At present, one COVID-19 diseased lung is infecting other 2.2 persons.

To confirm the disease, a doctor can use RT-PCR test. It can identify a small amount of viral ribonucleic acid (RNA) too. However, during the early stages of COVID-19, this test fails to diagnose the virus. A lung CT scan can be utilized by a clinician to detect lung illness caused by COVID-19 infection. Serological testing is another
method, which can be utilized to identify COVID-19. This method detects the presence of antibodies developed by the immune system to notice the existence of the COVID-19 virus. The RT-PCR test has a sensitivity of 30%–70%, which is lower than the lung CT-scan test.\(^3,4\) As of May 30, 2021, globally there are 170,650,028 humans affected by this virus, and 3,549,264 have lost their life. COVID-19 instances are on the rise as a result of the virus’s late detection in the human body.\(^5\)

1.1 Motivation

Due to less availability of testing facilities and healthcare workers, highly dense developing countries like Brazil, India, Argentina are struggling against the COVID-19 virus. Researchers are aiming to develop a reliable, cost-effective method for diagnosing COVID-19 that uses readily available radiography equipment to overcome this issue. COVID-19 is diagnosed mostly through lung X-ray scans or computed tomography (CT) scans. In the COVID-19 diseased lung's lung X-ray scan, a section of the lung changes color from black to a hazy gray tint. It looks similar to the frosted window glass of winter.\(^6\) However, these lung X-ray scans are insensitive to COVID-19 and can result in false-negative results. Lung CT scans are more sensitive than lung X-ray scans and provide a considerably more detailed perspective. These factors have motivated authors to design the COVID-19 detection system through lung CT scans.\(^7\)

1.2 Features of computed tomography scans

The following are the findings on a deep investigation of CT-scan pictures of COVID-19 diseased lungs by medical practitioners.

On analysis of CT scan of COVID-19 patient lungs at an early stage (after 2–3 days of infection), it has found that some ground-glass opacity (GGO) is present in lungs. This GGO is scattered throughout the lungs. GGO represents a little air sac filled with fluid and visible as a gray shade in the CT-scan image. It usually occurs in the lower lobe.\(^8\) It is shown in Figure 1A. In the advanced stage of COVID-19 (after 8–9 days of infection), more fluid will collect in the lungs along with severe lesions, and gray glass appearance in the right lower lobe will turn into solid white consolidation.\(^9\) It is shown in Figure 1B.

The intensity of GGO and paving pattern in the right lobe grows after 12 days of infection, and it begins to develop in the left lobe. Figure 2 depicts this (a). If the infection persists after 16–17 days, it develops into a more serious case of COVID-19, causing enlargement of the interstitial space between the lung lobules. It makes the

FIGURE 1 (A) Lung CT scan of COVID-19 diseased lung after 4 days showing GGO in the right lobe (B) Lung CT scan of COVID-19 diseased lung after 8 days showing more GGO with severe lesions and solid white consolidation\(^8,9\)

FIGURE 2 (A) Lung CT-scan of COVID-19 diseased lung after 12 days showing the severity of GGO increases in the right lobe and also start appearing in the left lobe, (B) Lung CT-scan of COVID-19 diseased lung after 17 days showing more solid white consolidation and the thicker wall\(^8,9\)
Among these models, the MobileCapsNet is suggested—a deep learning architecture that integrates ConvNets and CapsNet is designed. Four prominent ConvNet architectures namely VGG, DenseNet, ResNet, and MobileNet are utilized with CapsNet. These architectures have utilized ConvNets efficiently in the feature extraction process. CapsNet is being utilized while wrapping features into capsules. It has ensured that the model will not be reliant on large datasets and that it will be able to distinguish COVID-19 from non-COVID-19 using lung CT scans.

Among these models, the MobileCapsNet is suggested as the best model to design COVID-19 diagnosis smart applications due to its latency, size, and accuracy. The structure of the paper is as follows: Introduction to COVID-19, motivation for this work, and features of lung CT-scan images have already been discussed in Section 1. Work done in the area of COVID-19 detection through lung CT-scan images is discussed in Section 2. The methodology of the offered work is given in Section 3. The simulation environment and results are given in Section 4. Section 5 contains a discussion of the findings and their interpretation. Section 6 concludes with a conclusion and future scope.

2 | REVIEW OF LITERATURE

CT scans are a more reliable, quick, and accessible way for diagnosing COVID-19, and several researchers have recently developed a variety of models for diagnosing COVID-19 using CT scans. Furthermore, due to the widespread use of deep learning in the medical imaging field, the majority of COVID-19 diagnosis models are deep learning-based. As a result, the authors discuss the various findings of researchers in the diagnosis of COVID-19 utilizing lung CT-scan pictures in this section. Using lung CT-scan images, Li et al.\textsuperscript{10} created a system for detecting the COVID-19 virus. The framework was developed using 3D ConvNet by the authors. Authors have utilized 4536 CT-scan pictures of 3322 patients of age group 34–64, having 1838 male and 1484 female. From August 2016 to February 2020, these images were collected from six Chinese hospitals. Exactly 1296 images of COVID-19 diseased lungs, 1733 images of pneumonia patients, and 1325 images of non-pneumonia patients were among the 4536 lung CT-scan images. Their framework has achieved 90% sensitivity, 96% specificity, and 96% AUC.

Ozturk et al.\textsuperscript{11} have proposed a machine learning-based model to diagnose COVID-19 from lung X-ray and CT scans. The authors have utilized a lung X-ray and CT scans having four images of ARds diseased, 101 images of covid diseased, two images of pneumocystis-pneumonia diseased, 11 images of SARS diseased, six images of streptococcus diseased, and two images of normal persons. The authors have utilized the classical image augmentation approach to handle the class imbalance issue. Through different feature extraction techniques, they have obtained 78 features. Seventy-eight features are shrunken through stacked autoencoder (SAE) and principal component analysis (PCA) into 20 features. Finally, an SVM is utilized for image classification. The relevance of lung CT-scan pictures in detecting COVID-19 infection in a person has been debated by Zheng et al.\textsuperscript{12} The authors designed the prediction method utilizing a 3D lung CT scan and the concept of deep learning. First, the authors have utilized a pretrained U-Net to segment the lung region in CT scans. Then they have utilized 3D deep neural network to classify the segmented lung region into COVID-19 and non-COVID-19. The authors have utilized 499 lung CT scans collected from December 2019 to January 2020 to train their model. Further, they have utilized 131 lung CT scans collected from January 24, 2020 to February 6, 2020 to test their model. Authors have claimed to obtain 90.7% sensitivity, 91.1% specificity, and 90.1% accuracy.

Shan et al.\textsuperscript{13} have suggested a deep learning algorithm for automatic segmentation and quantification of the COVID-19 diseased areas in the patient lung using lung CT scans. For segmentation of the infectious region in lung CT scans, the authors have utilized VB-NET neural network. The authors have utilized 249 lung CT scans of COVID-19 diseased lungs to train the model. Further, they have utilized 300 lung CT scans of COVID-19 diseased lungs to validate the trained model. The authors have claimed that the proposed model is very fast and accurate relative to the existing manual quantification method.

Wang et al.\textsuperscript{14} have discussed the need for an alternative method to diagnose COVID-19 due to fewer testing facilities in developing countries. The authors have mentioned that these can be utilized ahead of the pathological test due to radiographic changes in lung CT scans. Thus, it can save lots of time and control the spread of disease. The authors have collected a pool of 453 lung CT scans of COVID-19 diseased and pneumonia diseased lungs. They have utilized 217 images for the model's training, and the remaining images were utilized for model testing. The authors have claimed to achieve 82.9% accuracy with 80.5% specificity and 84% sensitivity.

Zhang et al.\textsuperscript{15} have discussed the challenges involved in the early exposure of COVID-19 diseased lungs. The...
authors have established an AI-based system using 3777 lung CT scans to differentiate among COVID-19 diseased, pneumonia, and normal lungs. It is also mentioned that the proposed system can be utilized as an assistant to overloaded radiologists and physicians.

Singh et al.\textsuperscript{16} have given a ConvNet model to categorize lung CT scans in the COVID-19 diseased and non-diseased categories. The authors have utilized multi-objective differential evolution for tuning of starting parameters of convolutional neural networks. Furthermore, the authors have utilized 20-fold cross-validation to avoid overfitting. The authors have claimed that the model has shown good accuracy relative to ANN, ANFIS based classifiers. Rajinikanth et al.\textsuperscript{17} have proposed an image-based system to fetch the COVID-19 diseased section in the CT scans of the lungs. Firstly, the authors have utilized the threshold filter to remove artifacts. It has helped in the extraction of lung regions from CT scans. In the second step, the authors have utilized Otsu thresholding and Harmony search optimization for image enhancement. In the third step, the authors have extracted the diseased lung region. In the last step, the authors have utilized a region of interest feature to compute the level of severity. The primary objective of the proposed model is to assist health practitioners in assessing the severity of the disease.

Based on the findings of the literature study, CT scans can play a significant job in the early detection of COVID-19 infection. The next part provides an overview of the suggested model for detecting COVID-19 from lung CT-scan pictures.

3 | MATERIAL AND METHODS

A deep learning-based framework has been designed by integrating a capsule network (CapsNet) with DenseNet, ResNet, VGGNet, and MobileNet to decide COVID-19 from non-COVID-19 through lung CT scans. These hybrid architectures are named as DenseCapsNet, ResCapsNet, VGGCapsNet, and MobileCapsNet, respectively. This section contains the detail of the different model architecture utilized in the design of the proposed model. This section also gives a detailed CT-scan image dataset utilized for model design.

3.1 | DenseNet-121

Huang et al.\textsuperscript{18} have proposed the architecture of DenseNet where there is a connection among each pair of layers in a feed-forward fashion as the network of layers is very dense, so it is called DenseNet. A feature map of all preceding layers is considered a separate input for every layer, whereas its feature map is passed on as input to all subsequent layers. On the large-scale data set like ILSVRC 2012 (ImageNet), DenseNet attains accuracy similar to ResNet by consuming less than half of trainable parameters and FLOPs. Here, 121 represents the number of layers, among which 117 are convolution layers, three transition layers, and one classification layer. Each convolutional layer corresponds to the combined sequence of operations having batch normalization (BN)-ReluConv. The classification sub-network consists of $7 \times 7$ global mean pooling, $1000D$ dense layers, and a softmax layer.\textsuperscript{19}

3.2 | ResNet50

ResNet50 is a 48-layer residual network with one max-pooling layer and one mean pooling layer. It is a variation of the ResNet model. It performs $3.8 \times 10^{49}$ floating points operations. It is a commonly utilized ResNet model.\textsuperscript{20} The architecture of ResNet-50 is composed of four stages. First, images whose height and width are multiple of 32 and 3 will be taken as input to the network. The authors have considered the images of size $224 \times 224 \times 3$. Using kernel of size $7 \times 7$ and respectively, initial convolution and max-pooling have been implemented. After this, stage one has three residual blocks, each containing three layers. Kernel of sizes 64, 64, and 128 is utilized to implement convolution operation in all three layers. As the convolution operation in the residual block is implemented with stride 2, with the progression of the stage, the width of the channel gets double, and the input size gets half. Bottleneck design is utilized in a deeper network like ResNet 152. Corresponding to every residual function F, three convolution layers $1 \times 1$, $3 \times 3$, $1 \times 1$ are piled one over the other. Convolution layer $1 \times 1$ is responsible for reducing and restoring dimension, while $3 \times 3$ is left as the bottleneck for smaller input/output dimensions. In the end, there is mean pooling layer, which is followed by a dense layer with 1000 neurons.\textsuperscript{21,22}

3.3 | VGG16

In 2014, Simonyan et al.\textsuperscript{23} proposed the VGG16 model to win the ILSVR competition by achieving an accuracy of 92.7% on a very huge dataset with more than 14 million images. The model was recognized among the top five models of this prestigious competition. RGB image has a dimension of $224 \times 224$, which is given as input to a group of the convolutional layer where a small kernel
size filter $3 \times 3$ or even $1 \times 1$ is utilized to capture information from all fields. A convolutional stride of 1 pixel with a $3 \times 3$ filter is utilized to preserve the spatial resolution. Five max-pooling layers were utilized to carry out spatial pooling with a window size of $2 \times 2$ pixels and stride 2. Each max-pooling layer is followed by a collection of convolutional layers. It is followed by three completely connected layers, the first two of which have 4096 channels and the last having 1000 channels. After that, the softmax layer is utilized to get the final classification result. As an activation function, the rectified linear unit (ReLU) is present in all hidden layers. In the classification work, VGG outperformed the previous generation model. But still, some disadvantages are there associated with VGG16. The training process is very slow, and deployment is tiresome due to its size.\(^{24}\)

### 3.4 MobileNet

MobileNet is a ConvNet model for mobile vision and classification. Though there are other prominent architectures, MobileNet is a preferable choice since it only requires a small amount of computing power to operate or apply transfer learning.\(^{24}\) MobileNet is a simplified architecture for mobile and embedded vision applications that build lightweight deep convolutional neural networks using depthwise separable convolutions. MobileNet is built on depthwise separable convolutions, with the exception of the first layer. A full convolutional layer is the initial layer. Batch normalization and ReLU nonlinearity are applied to all layers. The final layer, on the other hand, is a fully linked, nonlinear layer that feeds the softmax for classification. Stride convolution is utilized for both depthwise convolution and the first fully convolutional layer in downsampling. When depthwise and pointwise convolution are considered independent layers, MobileNet has a total of 28 layers. The main distinction between the MobileNet design and typical convolution neural networks is using a single $3 \times 3$ convolution followed by ReLU and batch normalization instead of a single $33\%$ convolution followed by ReLU and batch normalization. MobileNet separated the convolution into a $3 \times 3$ depthwise convolution kernel and a $1 \times 1$ pointwise convolution.\(^{25}\)

### 3.5 Capsule network

The ConvNet is the key architecture for the achievement gained by the deep learning models in the area of classification. But there are some issues associated with the architecture of ConvNet, which degrades its performance in some cases. ConvNet-based model focuses on the features of an image and does not pay much attention to spatial information. Also, due to the usage of pooling operation, there is a loss of spatial information, affecting the accuracy of the ConvNet-based model. So it can be concluded that due to the translation-invariant nature of ConvNet architecture, it cannot record the changing position of an object.

Capsule Network (CapsNet) was introduced in 2017 by Sabour et al.,\(^{26}\) which is very useful in computer vision. It has offered a shift in paradigm in neural computation. In CapsNet, the traditional approach of scalar neural computation has been replaced by the vectorized approach. A capsule is a group of neurons that encode spatial information like angle, scale, and position and record the object probability present at that position. It minimizes the loss of spatial information and improves the CapsNet-based model's accuracy compared with the ConvNet-based model. This makes it translation equivariance and helps learn and characterize the patterns in various object recognition applications. It has motivated authors to use CapsNet in place of ConvNet while designing the proposed model.\(^{27}\)

Broadly, the architecture of CapsNet can be divided into two parts, namely encoder and decoder, respectively. Each part contains three layers. Figure 3 shows the architecture of the encoder and decoder in CapsNet. An encoder takes the input image and converts it into a 16-dimensional vector. Details of three layers of the encoder are as follows:

1. Convolutional neural network: This layer extracts basic features of an input image.
2. PrimaryCaps network: This layer fetches more detail patterns from the basic features of an input image. The number of capsules in this layer depends on the dataset.
3. CTscanCaps Network: The number of capsules in this layer is also not fixed. This layer contains the instantiation parameters.

The decoder takes a 16-dimensional vector as an input, and through three fully connected layers, it reconstructs the image. To perform the different functionality in different layers of encoder and decoder, a CapsNet has four main components. Their details are as follows:

1. Matrix Multiplication: To encode the information related to spatial relationships, a weighted matrix multiplication operation is performed on the output of the first layer in the encoder. It gives an estimate of the probability of correct classification.
2. Scalar Weighing: To match the weight of lower-level capsules with higher-level capsules, scalar weighing is implemented.

3. Dynamic Routing Algorithm: To ensure communication between different layers, dynamic routing algorithms are utilized. Though this will increase the space and time complexity, it also improves the accuracy. The algorithm of dynamic routing (Algorithm 1) is shown after Equation (3).

4. Vector to Vector Nonlinearity: This component is utilized for compressing information in the last step. Squash function $S_j$ defined in Equation (1) is utilized for this purpose. It takes the final vector and compresses a length in size less than one.

$$ S_j = \frac{||w_j||^2 w_j}{1 + ||w_j||^2 ||w_j||}. $$ (1)

where, $w_j$ is the weighted summation of capsules output, it is given by Equation (2)

$$ w_j = \sum_i c_i \tilde{v}_{ji}, $$ (2)

and $\tilde{v}_{ji}$ shows the affine transformation, it is defined in Equation (3).

$$ \tilde{v}_{ji} = W_{ji}v_j. $$ (3)

### Algorithm 1

**Dynamic Routing Algorithm**$^{26}$

Dy_Routing ($\tilde{v}_{ji}, x, y$)

1. ∀ caps node $i$ in layer $y$ and capsule node $j$ in layer $(y + 1)$

   $$ bij \leftarrow 0 $$

2. repeat the following $x$ times.

3. ∀ caps node $i$ in $y$th layer: $c_i \leftarrow$ softmax($b_i$).

4. ∀ caps node $j$ in $(y + 1)$th layer: $w_j \leftarrow \sum_i c_i \tilde{v}_{ji}$

5. ∀ caps node $j$ in $(y + 1)$th layer: $S_j \leftarrow$ squash($w_j$)

6. ∀ caps node $i$ in layer $y$ and caps node $j$ in layer $(y + 1)$:

   $$ bij \leftarrow bij + \tilde{v}_{ji}, S_j $$

return $v_j$

### 3.6 Proposed CapsNet architectures

Four different models are designed by integrating a capsule network (CapsNet) with DenseNet, ResNet,
VGGNet, and MobileNet to differentiate COVID-19 from mon-COVID-19 through lung CT scans. The design of these models is shown in Figure 4. These hybrid models are named as DenseCapsNet, ResCapsNet, VGGCapsNet, and MobileCapsNet, respectively. In the proposed architectures, one of the pretrained models (DenseNet, ResNet, VGGNet, MobileNet) has been utilized to calculate the initial feature map of an input image. Usage of these models at the initial level will help gather low-level features. Two changes are made in these pretrained models.

- The first change is to lower down the input shape
- The second change is to remove the last dense layer.

The weights of all the previous layers are frozen. Then, the collected feature map is passed to CapsNet model through a convolutional layer to classify the CT-scan image. This convolutional layer consists of 256 filters of $9 \times 9$ size and ReLU activation function, and stride 1 is applied. As mentioned in Section 3.5, the CapsNet architecture consists of a layer of primary capsules and CTscanCaps. Finally, CTscanCaps output is utilized to determine the class of the input CT-scan image.

3.7 CT-scan image dataset

The SARS-COV-2 CT-Scan dataset is accessible at Kaggle, which consists of CT scan from multiple patients. It contains 1252 CT pictures from patients who tested positive for SARS-CoV-2 (referred to as COVID +ve) and 1230 CT scans from patients who did not test positive (referred to as COVID –ve). The total number of CT scans in the dataset is 2482. Few sample CT scans from this dataset are presented in Figure 5. The entire dataset is divided into three subgroups for train, validation, and test sets, with 1737, 524, and 221, respectively, images in each subset. The training dataset is applied to learn the model, the validation dataset is utilized to validate the models during training, and the test dataset remains unseen for the models. The complete dataset is utilized to calculate the performance measures as provided in the results section.

4 EXPERIMENTS AND RESULTS

Experiments are done using deep learning library Kears with Intel Core i3-6006U@2.00 GHz CPU on 64-bit Windows 10 OS. In this section, we explore the lung CT image dataset and experiments with the proposed deep learning models. The proposed CapsNet architectures consist of a ConvNet followed by CapsNet. First, a CT-scan image gets passed through a pretrained ConvNet model to fetch initial feature maps. Then these features are given as input to the CapsNet model through a convolution layer to attain the final classification. The VGG16, DenseNet121, ResNet50, and MobileNet ConvNet architectures are integrated with CapsNet model separately. Two changes are carried out in these pretrained models before combining them with CapsNet classifier.

The first change is to lower the shape for the input images to 128, 128, 3, and the second change is to remove the last dense layer. The last layers, also referred to as top layers, are removed, and the weights of all the previous layers are frozen. This change is required since the dense layers at the end can only receive fixed-size inputs that...
the input shape has previously determined and all computations in the convolutional layers. Any modification to the input shape adjusts the input’s shape to the dense layers, making them incompatible weights. All CapsNet models receive the input of shape $128 \times 128 \times 3$. These images are resized to $128 \times 128$ pixels. These models are compiled with the **adam** optimizer. It merges the benefits of two SGD (Stochastic gradient descent) extensions, that is, root mean square propagation and adaptive gradient algorithm. The margin loss function is utilized that can be defined as in Equation (4):\(^3\)

$$
L_k = T_k \max(0, m^+ - \|v_k\|)^2 + \alpha(1 - T_k) \max(0, m^+ - \|v_k\| - m^-)^2.
$$

(4)

Where $T_k = 1$ if class $k$ feature exists and $m^+ = 0.9$ and $m^- = 0.1$, to ensure that the vector length will lie in the specified bounds. The $\alpha$ (down-weighting function) is utilized for numerical steadiness, with a suggested value of 0.5. VGGCapsNet, DenseCapsNet, and ResCapsNet models have utilized categorical cross-entropy loss function as defined in Equation (5):\(^3\)

$$
L = -\frac{1}{n} \sum_{k=1}^{n} n_k \log \frac{e^{W_k x_k + b_k}}{\sum_{i=1}^{n} e^{W_{ij} x_k + b_i}}.
$$

(5)

The MobileCapsNet model utilizes binary cross-entropy loss function as defined in Equation (6):\(^3\)

$$
L = -\frac{1}{n} \sum_{k=1}^{n} y_k \log(m_w(x_k)) + (1 - y_k) \log(1 - m_w(x_k)),
$$

(6)

Where $x_k$ = $k$th training sample, $y_k$ = class label of $k$th training sample, $b$ = bias term, $n$ = sample count, $m_w$ = model with weights $W$, $W$ is the weight matrix, $W_{ij}$ = $j$th column of $W$ $W_{ij}$ = $y_k$th column of $W$. All architectures are trained for 100 iterations (epochs), with 64 and 8 batch sizes respectively for train and validation datasets. To avoid overfitting, a call-back mechanism termed as early stopping method is utilized. This call-back allows specifying an arbitrarily large number of learning

---

**Figure 5** Sample Images for (A) CT scan of patients diseased by COVID-19 and (B) CT scan of patients with other pulmonary illnesses who are not diseased by COVID-19

**Table 1** Hyper-parameter values utilized for the COVID-19 classification models

| Hyper-parameters | Values |
|------------------|--------|
| Learning rate    | 0.0015 |
| Beta1            | 0.9    |
| Beta2            | 0.999  |
| Epsilon          | 0.1    |
| Decay            | 0.0    |
| Batch size       | 64     |
| Epochs           | 100    |
| Optimizer        | adam   |
| Metric           | accuracy |
| Monitor          | validation loss |
epochs and stop learning once the model performance stops improving on a hold-out validation dataset. The stage at which training should be stopped is determined by epoch count (patience parameter). The patience is set at 30 epochs for all models. A more extensive search is carried out in these classification models to search the set of hyper-parameters that attained the best performance; considered values are provided in Table 1.

Model performance learning curves on the train and validation datasets can be utilized to diagnose if a model

**FIGURE 6**  Progress of (A) validation and training accuracy (B) validation and training loss throughout the learning of the VGGCapsNet model. Accuracy and loss change abruptly during the first 20 epochs and become stable after 40 iterations. Early stopping is utilized to restore the best weight at epoch 60, specifically where the validation loss is minimum.

**FIGURE 7**  Progress of (A) validation and training accuracy (B) validation and training loss throughout the learning of the DenseCapsNet model. Accuracy and loss change abruptly during the first 40 iterations and become stable after 50 iterations. Early stopping is utilized to restore the best weight at epoch 45, specifically where the validation loss is minimum.
is underfitting, overfit, or well-fit. Figures 6–9 present the learning curves of VGGCapsNet, DenseCapsNet, ResCapsNet, and MobileCapsNet models. Learning curves are line plots that depict how learning performance changes over time as a function of experience. From these plots, it can be observed that the learning of the MobileCapsNet model is more stable and well-fit. It is also observed that the number of training epochs of MobileCapsNet is much fewer than that of VGGCapsNet, DenseCapsNet, and ResCapsNet.

It is required to evaluate the model's performance after it has been built, that is, how well it predicts the outcome of new observations test data that have not been utilized to train the model. To determine how many observations are correctly classified or wrongly, a confusion matrix is created. It illustrates the number of correct and wrong

![Figure 8](image8.png)

**FIGURE 8** Progress of (A) validation and training accuracy (B) validation and training loss throughout the learning of the ResCapsNet model. Accuracy changes abruptly during the epochs while the loss becomes stable after 40 epochs. Early stopping is utilized to restore the best weight at epoch 58, specifically where the validation loss is minimum.

![Figure 9](image9.png)

**FIGURE 9** Progress of (A) validation and training accuracy (B) validation and training loss throughout the learning of the MobileCapsNet model. Accuracy and loss change abruptly during the first 20 epochs and become stable after 30 iterations. Early stopping is utilized to restore the best weight at epoch 18, specifically where the validation loss is minimum.
FIGURE 10  Confusion matrix separately for each classification model (A) VGGCapsNet model, (B) DenseCapsNet model, (C) ResCapNet Classification model, and (D) MobileCapsNet model

TABLE 2  Performance measures for COVID-19 detection using VGGCapsNet and DenseCapsNet models

| Class/metric       | VGGCapsNet model | DenseCapsNet model |
|--------------------|------------------|--------------------|
|                    | Precision | Sensitivity | F-score | Precision | Sensitivity | F-score |
| COVID-19           | 0.98       | 0.99        | 0.99    | 0.99      | 1.00        | 0.99    |
| Non-COVID-19       | 0.99       | 0.98        | 0.99    | 1.00      | 0.97        | 0.99    |
| Macro average      | 0.99       | 0.99        | 0.99    | 0.99      | 0.99        | 0.99    |
| Weighted average   | 0.99       | 0.99        | 0.99    | 0.99      | 0.99        | 0.99    |
| Avg. accuracy      | 0.99       |             |         |           |             | 0.99    |

TABLE 3  Performance measure for COVID-19 detection using ResCapsNet and MobileCapsNet models

| Class/Metric       | ResCapsNet Model | MobileCapsNet Model |
|--------------------|------------------|--------------------|
|                    | Precision | Sensitivity | F-score | Precision | Sensitivity | F-score |
| COVID-19           | 0.96      | 0.98        | 0.97    | 0.97      | 1.00        | 0.99    |
| Non-COVID-19       | 0.98      | 0.96        | 0.97    | 1.00      | 0.97        | 0.99    |
| Macro average      | 0.97      | 0.97        | 0.97    | 0.97      | 0.97        | 0.97    |
| Micro average      | 0.97      | 0.97        | 0.97    | 0.99      | 0.99        | 0.99    |
| Weighted average   | 0.97      | 0.97        | 0.97    | 0.99      | 0.99        | 0.99    |
| Accuracy           | 0.97      |             |         |           |             | 0.99    |
predictions categorized by outcome by comparing observed and expected outcome values. The confusion matrix, individually, for each architecture is shown in Figure 10.

Aside from raw classification accuracy, numerous other metrics, like precision, sensitivity, and F-score, are commonly employed to evaluate the effectiveness of a classification model. Sensitivity and specificity are two
key measures in medical science that describe how well a classifier or diagnosis performs. The choice between sensitivity and specificity depends on the context. Usually, we are concerned with one of these metrics. The confusion matrix is utilized to calculate performance matrices, namely accuracy, precision, sensitivity (recall), and F-score, in order to evaluate the efficiency of suggested models. The following equations are utilized to determine these performance metrics:

\[
\text{Precision} = \frac{T^{\text{Positive}}}{T^{\text{Positive}} + F^{\text{Positive}}},
\]

\[
\text{Sensitivity} = \frac{T^{\text{Positive}}}{T^{\text{Positive}} + F^{\text{Negative}}},
\]

\[
F - \text{Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Sensitivity}}},
\]

\[
\text{Accuracy} = \frac{T^{\text{Positive}} + T^{\text{Negative}}}{T^{\text{Positive}} + T^{\text{Negative}} + F^{\text{Negative}} + F^{\text{Positive}}},
\]

Tables 2 and 3 show the findings in terms of the aforesaid performance metrics, respectively. Table 2 provides VGGCapsNet and DenseCapsNet models, and Table 3 presents results for ResCapsNet and MobileCapsNet models. From these results, it can be observed that all the proposed deep learning architectures offer effective accuracies.

To examine the more precise analysis of the projected models, the ROC with their respective AUC are plotted in Figure 11. In medical diagnosis, ROC/AUC plots are a method to examine the accuracy of indicative tests and decide the best threshold value for differentiating between positive and negative test results. From these ROC plots, it is observed that proposed deep learning models except ResCapsNet are good enough to discriminate against the COVID-19 affected lung from the normal one. Furthermore, the AUC value for all the models is 1.0 for both classes, which is remarkable.

5 DISCUSSION

The mean accuracy using VGGCapsNet, DenseCapsNet, and MobileCapsNet models is 0.99, as provided in Section 4. The precision, sensitivity, F-score are also remarkable for these models. It confirms that the proposed CapsNet models can be utilized for COVID-19 diagnosis with high sensitivity. Though the accuracies achieved through VGGCapsNet, DenseCapsNet, and MobileCapsNet are the same, the MobileCapsNet can be preferred for such diagnosis applications due to its low latency. MobileNet uses deeply separable convolutions. A pointwise convolution follows the depthwise convolution in a depthwise separable convolution. When compared with a network with regular convolutions of the same depth in the nets, the number of parameters is significantly reduced. As a result, deep neural architectures are made lighter. The numbers of non-trainable and trainable parameters are given in Table 4. MobileNet outperforms other similar models (DenseCapsNet, ResCapsNet, VGGCapsNet) in terms of latency and size while at par on the scale of accuracy. MobileCapsNet appears to use the fewest parameters, making it the best choice for mobile applications.

Table 5 shows the comparison of the proposed models with some other ConvNet and basic CapsNet architecture. It can be verified from Table 6 that the basic CapsNet performs unexpectedly poorly when applied on CT scans for classification. It happens because CapsNet was originally built for hand-writing numeric character identification. However, Table 5 also shows that the offered MobileCapsNet model has outperformed other architectures.

In Table 6, the proposed work is also compared with recent research contributions, which have utilized lung X-ray scans or CT scans for diagnosis of COVID-19. Here, it can be noticed that most of the recent work has utilized X-ray scans, while in the proposed work, CT scans are utilized. CT scans are more informative relative to X-ray scans. The suggested model is robust and lightweight, while most of the researchers have either utilized existing machine learning, deep learning, or 2D ConvNet-based models. Accuracy in the proposed model is 99%, which is relatively higher than the recent works as presented in Table. The proposed model has utilized a big dataset of
| Source                         | Dataset                                                                 | Model detail                                                                                     | Number of classes                                                                 | Accuracy/result/outcome                                                                 |
|-------------------------------|------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------|
| Ozturk et al.11               | 126 images of lung X-ray scans and computed tomography scans           | • Feature Extraction based on principal component analysis for SVM for classification.          | Binary class classification: COVID and non-COVID                                     | • 86.54% accuracy with all features, 71.92% accuracy with shrunken features through SAE |
|                               |                                                                        | • Stacked Auto Encoder (SAE) for feature extraction and SVM                                      |                                                                                     | • 94.23% accuracy with shrunken features through PCA                                   |
| Tiwari and A. Jain29          | A dataset of 2905 lung X-ray scans                                     | • Visual Geometry Group Capsule Network.                                                        | Multi-class classification: Normal, COVID, and Pneumonia                             | • 97% accuracy in binary classification                                                  |
|                               |                                                                        | • Convolutional Neural Network Capsule Network                                                 |                                                                                     | • 92% accuracy in multiclass classification                                               |
| Zhou et al.37                 | Authors have utilized datasets of computed tomography scans collected from Italian Society of Medical and Interventional Radiology having 473 CT slices | • U-Net architecture with attention mechanism for segmentation of infection in lung due to COVID-19 through computed tomography scans | Localization of COVID-19 infection                                                   | • Proposed model segments a single CT slice in $0.29$ s.                                |
|                               |                                                                        |                                                                                                 |                                                                                     | • Dice score is 83.1%                                                                   |
|                               |                                                                        |                                                                                                 |                                                                                     | • Hausdorff distance was 18.8                                                           |
| Selvaraj et al.38             | A dataset of 80 lung computed tomography scans                         | • Deep neural network based model is utilized for segmentation of the COVID-19 diseased portion in the computed tomography scans | Localization of COVID-19 infection                                                   | • Authors have claimed an accuracy of 93.8%                                              |
| Feng et al39                  | Utilized a 350 lung computed tomography scans dataset                  | • Authors have utilized different algorithms like random forest, logistic regression, and SVM   | Binary class classification: Bacterial pneumonia and COVID-19                         | • Achieved 96.2% accuracy.                                                              |
| El-Dosuky et al.40            | Authors have utilized 594 genome sequences                              | • ConvNet with cockroach swarm optimization                                                     | Multi-class classification: COVID-19, influenza virus (Type A, Type B, and Type C)     | • Achieved an overall of 99% accuracy                                                    |
| Dhaka et al.41                | Publically available dataset of 3000 X-ray scans                       | • 2D ConvNet model                                                                               | Multi-class classification: Normal, Bacterial Pneumonia, COVID-19                      | • Achieved a mean accuracy of 97.71%                                                    |
| Khan42                        | 340 X-ray scans dataset                                                | • A combination of feature extraction algorithm, support vector machine, and K-means clustering algorithm | Binary class classification: COVID-19 and healthy                                     | • Achieved maximum accuracy of 94.12%                                                    |
| Soares et al.28               | SARS-COV-2 CT-Scan dataset                                            | • eXplainable deep learning (xDNN) model                                                        | Binary classification: COVID-19 and Non-COVID-19                                     | • Attained 97.38% accuracy                                                               |
lungs; therefore, the reliability of the proposed model is more relative to models analyzed in Table. Lightweight MobileCapsNet can be easily deployed on mobile phones also, while the models analyzed in Table 6 are not lightweight.

6 | CONCLUSION

As of May 30, 2021, globally, there are 170,650,028 humans affected by this virus, and 3,549,264 have lost their life. The reason for the increasing number of COVID-19 diseased lungs is fewer testing facilities. A model has been established to identify COVID-19 infection using lung CT scans to tackle this issue within available resources. As opposed to RT-PCR, lung CT scanning is a more reliable, practicable, and quicker way of diagnosing and assessing COVID-19, particularly in epidemic areas. CapsNet is a network structure designed to overcome the limitations of ConvNet. It uses the dynamic routing capsule vector to extract features and achieve classification. However, its application on CT-scan images, especially for COVID-19 is not extensively studied and discovered. In this paper, the authors have introduced redefined network structures called DenseCapsNet, ResCapsNet, VGGCapsNet, and MobileCapsNet to achieve classification based on CapsNet. By incorporating ConvNet architectures such as DenseNet121, ResNet50, VGG16, and MobileNet at the top of CapsNet, the capsule network performs better classification on CT scans with complex features. Simulation results have validated the efficiency of the offered models. Out of these architectures, Mobile CapsNet is suggested to prefer in the design of COVID-19 diagnosis application via CT scans. In future work, more modalities have to be considered to develop a powerful covid detection system.

ACKNOWLEDGMENTS

None declared.

DATA AVAILABILITY STATEMENT

The data that support this study’s results are publicly accessible at https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset and are also available in Reference 29.

ORCID

Shamik Tiwari © https://orcid.org/0000-0002-5987-7101
Anurag Jain © https://orcid.org/0000-0001-5155-022X

REFERENCES

1. Zu ZY, Jiang MD, Xu PP, et al. Coronavirus disease 2019 (COVID-19): a perspective from China. Radiology. 2020;296: E15–E25.
2. Singhal T. A review of coronavirus disease-2019 (COVID-19). Indian J Pediatr. 2020;87:1-6.
3. Pan L, Mu M, Yang P, et al. Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: a descriptive, cross-sectional, multicenter study. Am J Gastroenterol. 2020;115:766-773.
4. Yang W, Cao Q, Qin L, et al. Clinical characteristics and imaging manifestations of the 2019 novel coronavirus disease (COVID-19): a multi-center study in Wenzhou city, Zhejiang, China. J Infect. 2020;80:388-393.
5. https://www.worldometers.info/coronavirus/?utm_campaign=homeAdUOA?si May 30, 2021.
6. Hall LO, Paul R, Golfdof DB, Golfdof GM. Finding COVID-19 from lung X-rays using deep learning on a small dataset. arXiv Preprint arXiv:200402060. 2020.
7. Fang Y, Zhang H, Xie J, et al. Sensitivity of lung CT for COVID-19: comparison to RT-PCR. Radiology. 2020;296:E115-E117.
8. Bernheim A, Mei X, Huang M, et al. Lung CT findings in coronavirus disease-19 (COVID-19): relationship to duration of infection. Radiology. 2020;295:200463.
9. Ding X, Xu J, Zhou J, Long Q. Lung CT findings of COVID-19 pneumonia by duration of symptoms. Eur J Radiol. 2020;127:109009.
10. Li L, Qin L, Xu Z, et al. Artificial Intelligence distinguishes COVID-19 from community acquired pneumonia on lung CT. Radiology. 2020;200905. https://doi.org/10.1148/radiol.2020200905
11. Öztürk S, Özkaya U, Barştugan M. Classification of coronavirus (COVID-19) from X-ray and CT scans using shrunked features. Int J Imaging Syst Technol. 2021;31(1):5-15.
12. Zheng C, Deng X, Fu Q, et al. Deep learning-based detection for COVID-19 from lung CT using weak label. medRxiv. 2020.
13. Shan F, Gao Y, Wang J, et al. Lung infection quantification of COVID-19 in CT scans with deep learning. arXiv Preprint arXiv:200304655. 2020.
14. Wang S, Kang B, Ma J, et al. A deep learning algorithm using CT scans to screen for Corona virus disease (COVID-19). MedRxiv. 2020.
15. Zhang K, Liu X, Shen J, et al. Clinically applicable AI system for accurate diagnosis, quantitative measurements, and prognosis of COVID-19 pneumonia using computed tomography. Cell. 2020;181(6):1423-1433.
16. Singh D, Kumar V, Kaur M. Classification of COVID-19 patients from lung CT scans using multi-objective differential evolution-based convolutional neural networks. Eur J Clin Microbiol Infect Dis. 2020;39(7):1379-1389.
17. Rajinikanth V, Dey N, Raj ANJ, Hassanien AE, Santosh KC, Raja N. Harmony-search and Otsu based system for coronavirus disease (COVID-19) detection using lung CT scans. arXiv Preprint arXiv:200403431. 2020.
18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Paper Presented at: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017;4700-4708.
19. Zhang J, Lu C, Li X, Kim HJ, Wang J. A full convolutional network based on DenseNet for remote sensing scene classification. Math Biosci Eng. 2019;16(3):3345-3367.
20. Wen L, Li X, Gao L. A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications. 2019;1-14.
21. Farooq M, Hafeez A. Covid-resnet: a deep learning framework for screening of covid19 from radiographs. arXiv Preprint arXiv:200314395. 2020.
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Preprint arXiv:14091556. 2018.
23. Guan Q, Wang Y, Ping B, et al. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study. J Cancer. 2019;10(20):4876-4882.
24. Chaturvedi SS, Gupta K, Prasad PS. Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using MobileNet. Paper Presented at: International Conference on Advanced Machine Learning Technologies and Applications (pp. 165-176). 2020. Springer.
25. Srinivasu PN, SivaSai JG, Ijaz MF, Bhoi AK, Kim W, Kang JJ. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. Sensors. 2021;21(8):2852.
26. Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. Advances in Neural Information Processing Systems. Paper presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Long Beach, CA: 2017:3856-3866.
27. Deng F, Pu S, Chen X, Shi Y, Yuan T, Pu S. Hyperspectral image classification with capsule network using limited training samples. Sensors. 2018;18(9):3153.
28. Soares E, Angelov P, Biaso S, Frosst MH, Abe DK. SARS-CoV-2 CT-scan dataset: a large dataset of real patients CT scans for SARS-CoV-2 identification. medRxiv. 2020.
29. SARS-COV-2 Ct-Scan Dataset. https://www.kaggle.com/plameneduardo/sarscov2-ctscan-dataset
30. Tiwari S, Jain A. Convolutional capsule network for COVID-19 detection using radiography images. Int J Imaging Syst Technol. 2021;31:525-539.
31. Ho Y, Wookey S. The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. IEEE Access. 2019; 8:4806-4813.
32. Tiwari S. An ensemble deep neural network model for onion-routed traffic detection to boost cloud security. Int J Grid High Perform Comput. 2021;13(1):1-17.
33. Yacoubh R, Axman D. Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. Paper Presented at: Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems. 2020;79-91.
34. Tiwari S. Dermatoscopy using multi-layer perceptron, convolution neural network, and capsule network to differentiate malignant melanoma from benign nevus. Int J Healthc Inform Syst Inform. 2021;16(3):58-73.
35. Obuchowski NA, Bullen JA. Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. Phys Med Biol. 2018;63(7):07TR01.
36. Feng K, Hong H, Tang K, Wang J. Decision making with machine learning and ROC curves. Available at SSRN 3382962. 2019.
37. Zhou T, Canu S, Ruan S. Automatic COVID-19 CT segmentation using U-net integrated spatial and channel attention mechanism. Int J Imaging Syst Technol. 2021;31(1):16-27.
38. Selvaraj D, Venkatesan A, Mahesh VG, Joseph Raj AN. An integrated feature frame work for automated segmentation of COVID-19 infection from lung CT scans. Int J Imaging Syst Technol. 2021;31(1):28-46.
39. Feng J, Guo Y, Wang S, et al. Differentiation between COVID-19 and bacterial pneumonia using radiomics of lung computed tomography and clinical features. Int J Imaging Syst Technol. 2021;31(1):47-58.
40. El-dosuky MA, Soliman M, Hassanien AE. COVID-19 vs influenza viruses: a cockroach optimized deep neural network classification approach. Int J Imaging Syst Technol. 2021;31(2):472-482.
41. Dhaka VS, Rani G, Oza MG, Sharma T, Misra A. A deep learning model for mass screening of COVID-19. Int J Imaging Syst Technol. 2021;31(2):483-498.
42. Khan MA. An automated and fast system to identify COVID-19 from X-ray radiograph of the lung using image processing and machine learning. Int J Imaging Syst Technol. 2021;31(2):499-508.

**How to cite this article:** Tiwari S, Jain A. A lightweight capsule network architecture for detection of COVID-19 from lung CT scans. Int J Imaging Syst Technol. 2022;32(2):419-434. doi:10.1002/ima.22706