Convolutional capsule network for COVID-19 detection using radiography images

Shamik Tiwari | Anurag Jain

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

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

Abstract
Novel corona virus COVID-19 has spread rapidly all over the world. Due to increasing COVID-19 cases, there is a dearth of testing kits. Therefore, there is a severe need for an automatic recognition system as a solution to reduce the spreading of the COVID-19 virus. This work offers a decision support system based on the X-ray image to diagnose the presence of the COVID-19 virus. A deep learning-based computer-aided decision support system will be capable to differentiate between COVID-19 and pneumonia. Recently, convolutional neural network (CNN) is designed for the diagnosis of COVID-19 patients through chest radiography (or chest X-ray, CXR) images. However, due to the usage of CNN, there are some limitations with these decision support systems. These systems suffer with the problem of view-invariance and loss of information due to down-sampling. In this paper, the capsule network (CapsNet)-based system named visual geometry group capsule network (VGG-CapsNet) for the diagnosis of COVID-19 is proposed. Due to the usage of capsule network (CapsNet), the authors have succeeded in removing the drawbacks found in the CNN-based decision support system for the detection of COVID-19. Through simulation results, it is found that VGG-CapsNet has performed better than the CNN-CapsNet model for the diagnosis of COVID-19. The proposed VGG-CapsNet-based system has shown 97% accuracy for COVID-19 versus non-COVID-19 classification, and 92% accuracy for COVID-19 versus normal versus viral pneumonia classification. Proposed VGG-CapsNet-based system available at https://github.com/shamiktiwari/COVID19_Xray can be used to detect the existence of COVID-19 virus in the human body through chest radiographic images.

KEYWORDS
capsule network, convolutional neural network, COVID-19, decision support system, deep learning, visual geometry group, X-ray

1 INTRODUCTION

As per National Center for Biotechnology Information (NCBI) report, there are 219 species of viruses, which can infect human being. A corona virus is a group of viruses, that generally causes minor problems like cough and cold. Majority of corona viruses are harmless, but the illness caused due to the novel corona virus COVID-19 is harmful and may result even in a patient's death. It is called corona due to the crown-like protein embossed
structures on its surface. Corona virus is a single-stranded RNA virus, which is more prone to mutation as compared to DNA-based viruses. COVID-19 spreads more rapidly relative to other viruses. Whenever this virus comes in contact with human beings, it establishes a very strong bond with the cell membrane of the person through its protein spikes. Its incubation period is about 5 days. COVID-19 virus generates a huge amount of mucus on the chest and causes swelling in the breathing pipe. If the mucus is not cleared then this results in damage to the lungs, which in turn can be fatal.2,3 Till February 12, 2021, around 23.83 lakh people have lost their lives due to COVID-19.4 Scientists all around the world are looking for a vaccine that can prevent the COVID-19 virus from penetrating the human cell membrane.

There are two aspects of the fight against the corona virus: First its detection and second its cure. If we can detect corona virus at an early stage, then we can stop the widespread of this pandemic. In developing countries, where health services are not easily available at a low price, there are financial constraints associated with the diagnostic test cost. Due to the widespread of COVID-19 cases in different countries, there was a sudden increase in the number of chest radiography during April 2020.5 Through analysis of the chest X-ray of COVID-19 infected patient, it has been found that at an early stage, symptoms are mild and there is a small ground-glass opacity and nodule in the lungs. With the progress of the disease, symptoms aggravate and this results in large ground-glass opacity around the nodule. Moreover, there are multiple consolidations in bilateral pulmonary. Halo, reversed halo, and paving stone signs are deposited around the nodule. At the severe stage, there are a lot of diffuse lesions in both the lungs, leading to the formation of the pulmonary fibrosis formation and lungs become white.6 This has motivated the authors to design a system to diagnose COVID-19 through X-ray images.

Due to data analytic techniques like machine learning and deep learning, the design of a decision support system is not as complex as it was two decades ago. The model of deep learning is inspired by the filtering and classification approach used by the human brain. Deep learning is a subclass of machine learning which consists of multiple processing layers. Layers are used for input analysis and classification of information. Input can be in text, sound, or image form. Due to these features of deep learning, it has application in medical image classification.7,8 Although convolutional neural network (CNN) is best suited for image classification in deep learning, there are some conceptual limitations associated with CNN. In CNN, information about the entity position which is used by the network for recognition is lost during max pool operation. Also, CNN does not consider a few spatial relationships among the simpler object. To overcome these conceptual limitations of CNN, authors have used the convolutional capsule network (CapsNet) for X-ray image classification.9

In this paper, the authors have proposed a deep learning model named visual geometry group capsule network (VGG-CapsNet) by combining VGGNet and CapsNet to acquire more detailed information from X-ray data. The major contributions of the proposed work are as follows:

1. Ensemble the CapsNet and VGG16Net to design a hybrid network framework named VGG-CapsNet. The input features are fetched using VGG16Net and the output of VGG16Net is given as input in CapsNet for classifying work.
2. The proposed VGG-CapsNet method is compared with CapsNet model.
3. Separate experiments are carried out for COVID-19 detection as a multi-class classification problem and as a binary classification problem. For multiclass classification, three classes are considered, namely, viral pneumonia, normal, and COVID-19. While in binary classification COVID-19 detection, images from normal and viral pneumonia classes are merged into a single class referred to as “Non-COVID19.”

The rest of the research paper is structured as follows: Methodology used by other researchers for diagnosis of COVID-19 from chest radiography is described in Section 2. Detail of VGG-CapsNet methodology and the dataset used is discussed in Section 3. In Section 4, the authors have given the details of the simulation result and their discussion is covered in Section 5. It is followed by concluding remarks and future scope in Section 6.

2 | LITERATURE REVIEW

In this section, the authors have discussed the methodology adopted by different researchers to find the presence of COVID-19 virus in the human body through X-ray images.

Apostolopoulos and Mpesiana10 have offered a COVID-19 detection model using chest radiography. The model was developed using deep learning with CNN (specifically transfer learning procedure). For the training of the model, the authors have utilized the chest X-ray image of a normal person, pneumonia infected person, and COVID-19 infected persons. These images are publicly available on medical repositories. The authors have trained and tested the model through two datasets of
X-ray images, each having 224 COVID-19 patients X-ray, 714 pneumonia patients' X-ray, and 504 normal people X-ray. Through simulation results, the authors have found that their model has attained an accuracy level of 96.78% in finding COVID-19 patients. In the future, the authors have also planned to include the X-ray images of SARS, EBOLA infected patients in the training process. Hall et al.11 have discussed the significance of chest radiography images in the detection of COVID-19. The authors have developed a deep learning model by the combination of pre-trained Resnet 50, VGG 16, and CNN. For training and testing, the authors have used publicly available chest radiography images of 135 COVID-19 patients and 320 viral and pneumonia infected patients. About 102 COVID-19 patient X-ray images and 102 pneumonia patient X-ray images were used for the pre-training of Resnet 50. This combination of three different classifiers was tested on 33 unseen COVID-19 X-ray patient images and 218 other virus and pneumonia patient images. The total accuracy of the hybrid model was 91.24%. Zhang et al.12 have mentioned the effectiveness of chest radiography in the detection of COVID-19 patients as this virus is affecting the lungs. The authors have established a deep learning-based anomaly detection model for cost-effective and fast screening purposes. They have used the 100 chest radiographic images accessible on Github. Among the image dataset, 70 chest radiographic images were of COVID-19 patients while the remaining 30 were of patients suffering from other infections. Also, to effectively utilize the features of deep learning, the authors have used 1008 images of pneumonia patients. Their simulation results have shown the accuracy of 96% in COVID-19 cases, while 70.65% accuracy on non-COVID cases.

Ucar and Korkmaz13 have discussed the issues associated with the detection of COVID-19. Authors have used the artificial intelligence-based structure to detect COVID-19 through chest radiography. They have used Squeezenet with Bayesian optimization for deep learning model development. Also, to improve accuracy, they have incorporated hyper-parameters and augmented datasets. Authors have claimed to achieve an accuracy of 98.3% in the detection of normal, pneumonia, and COVID-19 cases. Khalifa et al.14 have mentioned the inflame of infection in the lungs caused by COVID-19. Authors have used generative adversarial network (GAN) with Resnet 18, AlexNet, GoogleNet, and SqueezeNet deep transfer learning models. For training and testing of models, authors have utilized a dataset having 5863 chest radiography images of normal and pneumonia infected patients. Through simulation results, the authors have concluded that a combination of GAN with Resnet18 has achieved 99% on the scale of precision, recall, and F-1 score. Salman et al.15 have discussed the role of high-definition X-ray images in reducing the burden of radiologists and detecting COVID-19. They have used a set of 260 images available on Kaggle and Github for developing a CNN-based deep learning model. The used image dataset contains chest radiography of 130 normal patients and 130 COVID-19 infected patients. Through simulation results, authors have claimed to achieve accuracy equivalent to an expert radiologist.

Sethy et al.16 have suggested a deep learning-based strategy to identify COVID-19 using chest radiography. The authors have proposed a hybrid approach by combining the support vector machine and resnet50. They have used the chest radiography images available on Github and Kaggle. They have claimed to achieve 95.38% accuracy. Maghdid et al.17 have focused on the development of publicly available chest radiography image datasets. The authors have also tried to provide the COVID-19 detection method using transfer-learning based on deep learning models. The authors have designed two prediction models using vein-based CNN and Alexnet-based CNN, respectively. They have claimed to achieve 94.1% accuracy from a vein-based CNN model and 98% accuracy from the Alexnet-based CNN model. Bassi and Attux18 have proposed a chest radiography classification model for the detection of COVID-19. They have developed a model using DenseNet121 CNN using chest radiography images of COVID-19 patients, pneumonia patients, and normal person. The model was trained twice, first through image set and second through chest radiography database. Through a simulation study, authors have claimed to achieve 97.8% test accuracy for the COVID-19 class. Ozturk et al.19 have developed a deep learning-based model for the detection of COVID-19 using chest radiography. They have used Darknet CNN classifier having 17 layers with the YOLO object detection system. The model can be used for binary and multi-class classification of chest radiographic images. While finding the COVID-19 patient, the authors have shown to achieve 98.08% and 87.02% accuracy in binary and multi-class classification, respectively.

Mei et al.20 have discussed the issues associated with reverse transcriptase-polymerase chain reaction (RT-PCR) test for the diagnosis of COVID-19. One of the prime issues associated with the RT-PCR test is less availability of kits and delay in the result of the test. Besides, the issue of false-negative is also associated with it. The authors have also discussed the role of chest CT scan images in finding COVID-19 infection. However, they have also mentioned the limitations associated with a chest CT scan while finding COVID-19 infection. Using artificial intelligence-based algorithms, authors have designed a system to diagnose the COVID-19 infection by...
integrating clinical symptoms, laboratory test results, exposure history, and findings of chest CT scans. The authors have developed three different models. The first model was a deep learning-based model which was developed using CNN. The model was trained through chest CT images only. The second model was developed using a support vector machine and random forest algorithms. This model was trained through clinical information. The third model was designed using CNN, which was trained by combining the clinical and radiographic information. The authors have tested all the three models on 279 patients and claimed to achieve .92 area under the curve from the third model. Li et al21 have developed a deep learning-based model called COVNet (COVID-19 detection from a neural network) to detect COVID-19 using chest CT scan images. For training and testing of the designed model, the authors have used the 4352 chest CT scan images of 3322 patients. Among the 3322 patients, 1838 were male, and 1484 were female. The authors have collected this dataset from August 2016 to February 2020. These data include images of COVID-19 patients, pneumonia patients, and normal persons. Authors have claimed to achieve .96 area under the curve for COVID-19 CT scan images and .95 for pneumonia patients.

Toraman et al22 have discussed the problems in detecting COVID-19 due to similarity with lung infection symptoms. They have also discussed other factors that motivate them for the development of a computer-based tool to detect COVID-19. The authors have also mentioned the role of information embedded in radiographic images while finding COVID-19. Using a deep learning-based model named CapsNet, authors have designed a system to detect COVID-19. The system has been trained and tested using chest radiography images. Authors have used 231 X-ray images of COVID-19 patients, 1050 X-ray images of a normal person, and 1050 X-ray images of pneumonia patients. Authors have claimed to achieve 97.24% accuracy and 84.22% accuracy in binary and multi-class classification, respectively, while detecting COVID-19. Shoeibi et al23 have described the characteristics of the COVID-19 virus. The authors have also discussed the significance of X-ray and CT scan images in the diagnosis of COVID-19. They have discussed the role of artificial intelligence-based methods like deep learning and machine learning in the detection of COVID-19. They have further mentioned that deep learning has gained more popularity relative to machine learning-based methods due to automatic feature extraction, selection, and classification. The authors have conducted a detailed review of different deep learning-based models found in the literature for the detection of COVID-19 using chest radiography. In the end, the authors have also mentioned the challenges while developing a deep learning-based COVID-19 diagnosis system.

3 MATERIAL AND METHODS

In this section, details of deep learning models, proposed architecture and image dataset utilized have been described. The CNN and CapsNet is the basic building block of the proposed model.

3.1 Convolutional neural network

CNN is the key to deep learning success. It has a variety of applications in the computer vision domain. It can identify the different objects in an input image. The key feature of CNN is that it requires less pre-processing task relative to other classification algorithms for object detection in an image. CNN uses two operations, namely, pooling and convolution to fetch the embedded features of an image. Then these extracted features are used for the training of the model and later on for the classification of images. Details of each layer are as follows.24,25

3.1.1 Convolutional layer

This layer acts as a filter, which views the area of the image in the size of $3 \times 3$ pixels or $5 \times 5$ pixels, respectively. A dot product is applied between the viewed area and predefined weight in the layer. This procedure of the dot product between viewing area and weight is called convolution. Results found after dot product are summed up. A resultant number which is the outcome of convolution operation is used to represent the viewed image area. Mathematically, if we represent the image by function $f$, filter by function $h$, and output by function $C$, respectively, then for filter size of $m \times n$, convolution layer functionality can be described through the following equation.

$$C[m,n] = (f \times h)[m,n] = \sum_j \sum_k h[j,k]f[m-j,n-k]$$

3.1.2 Activation layer

The outcome of the convolutional layer is a matrix that is relatively very small than the original image. The activation layer introduces non-linearity in the input received from the convolutional layer, which helps in the training of the network through backpropagation. There can be
multiple activation layers. If we represent the input to the activation layer by function $C[i]$, function for activation layer by $Z[i]$ and output by $A[i]$, then it can be represented by Equation (2). Here the activation function can be Sigmoid, Tanh, ReLU, and so forth.

$$A[i] = Z[i] \left( C[i] \right)$$  \hspace{1cm} (2)

### 3.1.3 Pooling layer

In this layer, the size of the matrix received from the activation layer is further reduced by selecting either maximum or minimum value from each group of values. Reduction in size, speed up the training procedure. Like the activation layer, we can have multiple pooling layers. If we represent the pooling layer function through $P[i]$ then it can be represented through Equation (3). Here pooling function can be the max, min, and so forth.

$$T[i] = P[i] \left( A[i] \right)$$  \hspace{1cm} (3)

### 3.1.4 Fully connected layer

This layer receives a one-dimension vector from the previous layer. It gives the output as a set of probabilities corresponding to the different labels attached to the image. The label that receives the highest value of probability represents the image identified by the model.

### 3.2 Capsule network

CapsNet is offered by Hinton and his colleagues as an alternative to CNN.

Conventional CNN architecture contains pooling layers for down-sampling. However, pooling is applied on the feature map to confirm that CNN identifies similar objects in images with different scales and to reduce the memory requirement of the model using down-sampling. It leads to the spatial invariance in CNNs that ultimately converts into one of the major flaws of CNN.

The major challenges of CNNs that steered the development of CapsNet are their incapability to identify texture, pose, and shape distortions of the complete image or a part of the image. In short, CNNs are invariant due to the pooling operation. They are not equivariant and due to this, it lacks equivalence. Moreover, CNN loses some features in the image due to down-sampling. It consequently requires a large size of training data for loss compensation. It has a lengthier training time relative to CapsNet because CNN has deep architecture in-depth, whereas CapsNet has deep architecture in width with limited parameters. Some prominent methods for CapsNet design are transforming autoencoders, dynamic routing-based vector capsules and expectation maximization routing-based matrix.

In CapsNet, Capsule represents a vector that denotes the probability of the existence of an entity, and orientation indicates the properties possessed by the entity. Every dimension of the capsule can be described as a property like texture, size, deformation, position, velocity, hue, albedo, and so forth. The squashing function (non-linear function) is used in training procedures for easy classification of different capsules. This function acts as activation at the capsule level. The squashing function restricts the capsule length under 1 and it can be easily used for the classification of present entities. Different hierarchical caps in CapsNet are similar to layers available in the basic neural network. The number of capsules in the lower level cap is large and they have small dimensionality, whereas the number of the capsule is less in higher-level cap and large dimensionality. The reason for this kind of organization is that at lower level capsules, each capsule represents a small input area, so due to a large number of capsules at the lower level, spatial relationships can be easily represented. While at the upper level, a higher number of dimensions are used to store the location information. Due to this, dimensionality increases while moving from lower level to higher-level capsule. After the determination of dimensionality increment, we can establish the relationship between different capsules and caps. Higher-level capsules are represented as a weighted summation of lower-level capsules. Dynamic routing helps in the conversion of the lower-level capsule to higher-level capsule.

There are multiple layers of capsules in CapsNet. As shown in Figure 1, the proposed CapsNet architecture, comprises a single ReLU convolution layer, followed by a layer of primary capsules (squashed and resized o/p of the convolutional layer), and a layer of X-rayCaps (i.e., Capsules signifying three classes of X-ray images).

The convolution layer transforms pixel intensities to the activities of local feature finders, which are further passed as input to the primary capsules. In contrast to max-pooling layers of CNN, the convolutional layer with strides larger than 2 is used for dimension reduction. The output of X-rayCaps is utilized to decide the class of input image. The entire capsules of X-rayCaps layer have a connection with every capsule in the primary capsule layer. Routing-by-agreement is a method that enables better learning in comparison to the max-pooling operation.
based routing. In the case of dynamic routing by agreement method, a feedback procedure that raises the support of those capsules that reach the agreement most with the parent output, thereby further solidifying its support.\textsuperscript{31}

Notably, such activation functions in deep networks are used to approximate non-linear relationships by basic mathematical functions. These functions are usually applied to scalar values. Notwithstanding, a special form of nonlinear activation function termed as a squash function is used in CapsNet to normalize the magnitude of the vectors. This squash function is defined as

\[
\hat{v}_j = \frac{||\hat{s}_j||^2 s_j}{1 + ||\hat{s}_j||^2} \quad (4)
\]

where, \(s_j\) is the weighted sum output of capsules given by

\[
s_j = \sum_i c_{ij} \hat{u}_{j|i}
\]

and \(\hat{u}_{j|i}\) represents the affine transformation defined as

\[
\hat{u}_{j|i} = W_{\theta} u_i
\]

Dynamic Routing Algorithm\textsuperscript{24}

Dy_Routing (\(\hat{u}_{j|i}, x, y\))

\begin{itemize}
  \item \textit{Step 1}: \(\forall\) capsule node \(i\) in layer \(y\) and capsule node \(j\) in layer \((y+1)\): \(b_{ij} \leftarrow 0\)
  \item \textit{Step 2}: for \(x\) iterations do
  \item \textit{Step 3}: \(\forall\) capsule node \(i\) in layer \(y\):
    \[ c_i \leftarrow \text{softmax}(b_i) \]
  \item \textit{Step 4}: \(\forall\) capsule node \(j\) in layer \((y+1)\):
    \[ s_j \leftarrow \sum c_i \hat{u}_{j|i} \]
  \item \textit{Step 5}: \(\forall\) capsule node \(j\) in layer \((y+1)\):
    \[ v_j \leftarrow \text{squash}(s_j) \]
  \item \textit{Step 6}: \(\forall\) capsule node \(i\) in layer \(y\) and capsule node \(j\) in layer \((y+1)\):
    \[ b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j \]
  \item return \(v_j\)
\end{itemize}

3.3 Proposed VGG-CapNet architecture

In earlier work, researchers have used separately CNN and CapsNet models for COVID-19 detection as discussed in the introduction section. In this work, the proposed
model extends the design of CapsNet with VGG16 network to design the VGG-CapNet model to obtain more detailed information on X-ray data for COVID-19 diagnosis. In the proposed model, VGG16 network is used to calculate the initial feature map and then these feature maps are passed to CapsNet model for classification.

A feature extraction layer of fundamental CapsNet consists of one convolutional layer having 256 filters having 9 × 9 size with a stride of 1 and is activated with ReLU function to pass these features to the primary capsule layer. This feature extraction layer is modified in the proposed architecture. As demonstrated in Figure 2, the suggested architecture VGG-CapsNet is divided into two parts: VGG16 and CapsNet. First, an X-ray image is fed to the pre-trained VGG16 model, and convolutional layers are used to compute initial feature maps. After this, an initial feature map is passed into CapsNet to achieve the classification. In the proposed architecture, the initial feature detector is substituted with the “block5_pool” layer of the VGG16 model. This modification can be considered as a transfer learning approach, and, here, the objective is to transfer the low-level feature representation capability of the VGG16 model into the CapsNet so that it becomes able to identify the low-level features as decent as VGG16. The VGG16 architecture was first introduced by Simonyan, Zisserman in the year 2014. VGG16 model consists of 13 convolutional layers along with the max pooling, the two fully connected layers, and a Softmax classifier. The large size kernels are replaced with the multiple numbers of 3 × 3 filters in this network to extract complex features at a low cost.

### 3.4 Image dataset

The image dataset of chest radiography is used in this work. These data include 219, 1345, and 1341 images of COVID-19, pneumonia, and normal condition, respectively. The sample image from each class is shown in Figure 2. The dataset is divided into three parts, namely, train, test, and validation subsets. Since the dataset is unbalanced, an equal number of images are used from each class for evaluation. About 60 and 150 images are randomly selected to prepare validation and test datasets, respectively, for COVID-19 multiclass classification. For COVID-19 detection as a two-class classification, images of viral pneumonia and normal classes are merged into one class, that is, referred to as “non-COVID19.” Figure 3 presents the sample image from each class.

### 4 EXPERIMENT AND RESULTS

The objective of these experiments is to design and assess the generalization capabilities of both the classifiers, that is, CNN and CapsNet models. The performance of each

---

**Figure 2** Proposed visual geometry group capsule network (VGG-CapsNet) architecture for COVID-19 detection. The model consists of VGG16 to calculate the initial feature map and CapsNet in subsequent layers [Color figure can be viewed at wileyonlinelibrary.com]
model is evaluated using precision, recall, F1 score, and accuracy. These are the significant metrics to measure the performance of the trained model as defined below.33,34 All the images, including training, validation, and testing subsets are used to calculate the performance metrics.

\[
\text{Precision} = \frac{T^P}{T^P + F^P} \\
\text{recall} = \frac{T^P}{T^P + F^N} \\
F\text{-score} = \frac{(2 \times \text{recall} \times \text{precision})}{(\text{recall} + \text{precision})}
\]

here,

- \(T^N\) (True Negative): The actual value is false, and the COVID-19 detection model has projected false.
- \(F^P\) (False Positive): It is a Type-I error. The actual value is false, and the COVID-19 detection model has predicted true.
- \(T^P\) (True Positive): The actual value is true, and the COVID-19 detection model has predicted true.
- \(F^N\) (False Negative): It is a Type-II error. The actual value is true, and the COVID-19 detection model has predicted false.

Receiver operating characteristic (ROC) curves are also plotted for both models. It is a 2 - D representation for visualizing, establishing, and selecting classifiers based on their performance. A metric area under the ROC curve (AUC) is usually calculated to represent it using a single scalar value. The AUC is a segment of the area of the unit square. It lies in the range 0–1.0. A model is considered excellent if the AUC value is more than 0.95. In this COVID-19 detection problem, macro-average and micro-average are also calculated. Macro-average metric determines the performance independently for each class and then compute the mean by considering all classes likewise, whereas micro-average metric aggregates all chest radiography classes to compute the mean value.

4.1 Model 1: CNN-CapsNet model for COVID-19 detection

The CNN-CapsNet architecture consists of two parts: CNN and CapsNet. First, an X-ray image is passed into a CNN model, and the convolutional layers are used to extract initial feature maps, which are then fed into the CapsNet model to achieve the final classification. The CNN model-used CNN-CapsNet architecture has four convolutional layers. The number of kernels in each convolutional layer is 64, 64, 128, and 128 in that order. Each convolutional filter has a size of 3 x 3. As an activation function, ReLU is used on all layers and the second convolution layer is followed by an average-pooling layer. The CapsNet part of the CNN-CapsNet model has the architecture as discussed in Section 3.2.

Two separate models are designed here. One for COVID-19 detection as a binary class problem, that is, COVID-19 versus non-COVID-19. The second model is designed for classifying among three classes, namely, COVID-19, normal, and viral pneumonia. The models are compiled with the adam optimizer for 0.001 as the initial learning rate. A margin loss function is utilized to decide whether the radiographic features of a particular class exist. The loss function has the form given by the following equation:

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

(5)

where \(T_k = 1\) if a feature of class \(k\) is existing and \(m^+ = 0.9\) and \(m^- = 0.1\), to confirm that the vector length.

**FIGURE 3** Sample images from the dataset (A) COVID-19, (B) normal and (C) viral pneumonia
remains within practical limits. The $\alpha$ down-weighting function is presented for numerical stability and is recommended to be set at 0.5. Both models are trained for 100 epochs where the size for train batch and validation batch is set to 32 and 1, respectively. The X-ray images of size $64 \times 64$ are used for learning and testing. To avoid overfitting, an early stopping method is used to end learning. The early stopping mechanism stops the training when the validation score ends further improving. In this mechanism, a loss function is computed after each epoch on a validation set and training is stopped when the validation loss starts to rise which is the sign of overfitting. The stage at which to stop training depends on a patience parameter, corresponding to the number of epochs to wait to see if validation loss remains to improve. For both models, the patience is set to 10.

Figures 4 and 5 present the classification accuracy and loss curves in a train and the validation set during the training of CNN model for binary class classification and three class classification. The confusion matrix for each classification model is presented in Figure 6. Table 1 gives the results of this experiment in terms of performance metrics as discussed earlier.

4.2 Model 2: VGG-CapsNet model for COVID-19 detection

As described in Section 3.3, VGG-16 pre-trained CNN model is used as feature extractors to evaluate the effectiveness of convolutional features on the classification by the CapsNet in this experiment. The CNN developed by the VGG is widely employed for medical image classification problems. Moreover, after the trial implementation with the other models, the MobileNet and ResNets could not match the accuracy provided by the VGG16 and thus VGG16 model was a better choice. X-ray images are resized to size $299 \times 299$ as required by VGG-16 input layer. Two separate VGG-CapsNet models are designed here. One for COVID-19 detection as a binary classification problem, that is, COVID-19 versus non-COVID-19. The second model is designed for classifying among three classes, namely, COVID-19, normal, and viral pneumonia. The models are compiled with the sgd optimizer, where 0.00001, 0.9, and 0.001 are used as an initial learning rate, momentum, and decay, respectively. Binary cross-entropy is used as a loss function for a binary class model, which has the form given by the following equation:

$$L_{BCE} = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(h_w(x_i)) + (1-y_i)\log(1-h_w(x_i))$$

where $N$ is the number of training samples, $y_i$ is the target label for the $i$th training sample, $x_i$ is the input for the $i$th training sample, and $h_w$ is the model with weights $w$.

Categorical cross-entropy is used as a loss function for the three-class model, which has the form given by the following equation.
FIGURE 5  Evolution of classification accuracy and loss in a train and validation set during the convolutional neural network-capsule network (CNN-CapsNet) model’s training for three-class classification. Training loss decreases sharply during the first 10 iterations and stabilizes after 15 iterations. Training stops after 29 epochs after patience of 10 epochs [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 6  Confusion matrix for COVID-19 detection with COVID-19 and non-COVID-19 classes (A) COVID-19 detection using the convolutional neural network-capsule network (CNN-CapsNet) model, and (B) COVID-19 detection using visual geometry group capsule network (VGG-CapsNet) model [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1  Results in terms of performance metrics for CNN-CapsNet and VGG-CapsNet models for COVID19 versus non-COVID19 classification

| Class        | COVID-19 detection using CNN-CapsNet model | COVID-19 detection using VGG-CapsNet model |
|--------------|--------------------------------------------|-------------------------------------------|
|              | Precision  | Recall | F1 score | Precision | Recall | F1 score |
| COVID-19     | 1.00       | 0.90   | 0.95     | 0.96      | 0.98   | 0.97     |
| Non-COVID19  | 0.91       | 1.00   | 0.95     | 0.98      | 0.96   | 0.97     |
| Accuracy     | 0.95       |        |          |           |        | 0.97     |

Abbreviations: CNN-CapsNet, convolutional neural network-capsule network; VGG-CapsNet, visual geometry group capsule network.
\[ L_{CE} = -\frac{1}{N} \sum_{i=1}^{N} \log \left( \frac{e^{W_{y_i}x_i + b_{y_i}}}{\sum_{j=1}^{n} e^{W_{y_i}x_i + b_j}} \right) \]  

(7)

where \( W \) is the weight matrix, \( b \) is the bias term, \( x_i \) is the \( i \)th training sample, \( y_i \) is the class label for \( i \)th training sample, \( N \) is the number of samples, \( W_j \) and \( W_{y_i} \) are the \( j \)th and \( y_i \)th column of \( W \), respectively.

Both models are trained for 100 epochs where the size for train batch and validation batch is set to 32 and 2, respectively. The X-ray images of size 299 × 299 are used for learning and validation and testing. An early stopping practice to stop learning is used to avoid over-fitting. Early stopping is utilized during the learning process. A loss function is computed after each epoch on the validation set and once the validation loss begins to rise, training is stopped. For both models, the patience is set to 10. Figures 7 and 8 represent the classification accuracy and loss curves in a train and the validation set during the training of the VGG-CapsNet model for binary class classification and three class classification objectives. The confusion matrix for each classification model is presented in Figure 9. Table 2 provides the results of this experiment in terms of performance metrics.
FIGURE 9  Confusion matrix for COVID-19 detection with COVID-19, normal, and viral pneumonia classes (A) COVID-19 detection using the convolutional neural network-capsule network (CNN-CapsNet) model, and (B) COVID-19 detection using visual geometry group capsule network (VGG-CapsNet) model [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2  Results in terms of performance metrics for CNN-CapsNet and VGG-CapsNet models for COVID19 versus normal versus viral pneumonia classification

| Class            | COVID-19 detection using CNN-CapsNet model | COVID-19 detection using VGG-CapsNet model |
|------------------|-------------------------------------------|-------------------------------------------|
|                  | Precision | Recall | F1 score | Precision | Recall | F1 score |
| COVID-19         | 0.96      | 0.86   | 0.91     | 0.96      | 0.98   | 0.97     |
| Normal           | 0.89      | 1.00   | 0.94     | 0.87      | 0.96   | 0.91     |
| Viral pneumonia  | 0.88      | 0.86   | 0.87     | 0.92      | 0.82   | 0.87     |
| Accuracy         | 0.91      |        |          | 0.92      |        |          |

Abbreviations: CNN-CapsNet, convolutional neural network-capsule network; VGG-CapsNet, visual geometry group capsule network.

FIGURE 10  Receiver operating characteristic (ROC) curve obtained using convolutional neural network-capsule network (CNN-CapsNet) and visual geometry group capsule network (VGG-CapsNet) models demonstrating AUC, micro average, and macro average performance metrics for COVID-19 and non-COVID-19 classes separately [Color figure can be viewed at wileyonlinelibrary.com]
5 | DISCUSSION

As discussed in Section 4, four separate experiments are conducted using CNN-CapsNet and VGG-CapsNet models. The results are presented in terms of confusion matrix and performance metrics. It is evident from the result analysis that the VGG-CapsNet model has achieved higher accuracy than the CNN-CapsNet model for both objectives, that is, COVID-19 versus non-COVID19 classification and COVID-19 versus normal versus viral pneumonia classification. For the binary classification objective, the VGG-CapsNet model has provided 97% accuracy which is higher than the 95% accuracy achieved with CNN-CapsNet model. Also, VGG-CapsNet model has provided 92% accuracy which is higher than the 91% accuracy of CNN-CapsNet model for the multi-class classification objective. ROC plots are shown in Figures 10 and 11 also confirms the dominance of VGG-CapsNet model over CNN-CapsNet model. The micro-average area and macro-average area for VGG-CapsNet model are 0.99 and 1.0, respectively, for binary classification. The metric micro-average area is slightly improved in comparison to ConvNet model, which is 0.97. For multiclass classification, the micro-average area and macro-average area for VGG-CapsNet model are 0.99 and 0.99, respectively. These metrics are slightly improved in comparison to ConvNet model, which are 0.96 and 0.98, respectively. Finally, it can be concluded that VGG-CapsNet model is more robust to CNN-CapsNet model based on these performance metrics-based and ROC plots.

6 | CONCLUSION AND FUTURE SCOPE

Early detection of COVID-19 patients is crucial to prevent the widespread of this pandemic. In this work, deep learning-based models are designed to segregate COVID-19 patients from normal and other viral pneumonia patients automatically. Two deep learning models are designed in this work, namely, CNN-CapsNet and VGG-CapsNet. CapsNets is developed as a type of deep CNN that handles the limitations of traditional CNNs. CapsNet is the proficient development for computer vision applications that has been designed by substituting the usual scalar activations with vectors and by routing the neurons in a different style. The ability of CapsNet is enhanced by its integration with CNN and VGG16 models to compute initial feature maps. Performance analysis is carried out to select the best model. Experimental results have demonstrated that the VGG-CapsNet model performs better than CNN-CapsNet model. The proposed VGG-CapsNet has achieved an accuracy of 97% for segregating COVID-19 samples from the non-COVID-19 samples and 92% accuracy for segregating COVID-19 samples from normal and other viral pneumonia.
samples. The proposed model is expected to help clinicians to make decisions in clinical practice.

In the future, this work can be extended by designing a multi-modality deep learning model based on CT-scan and X-ray images for the robust detection of COVID-19 patients.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are openly available in kaggle at https://doi.org/10.34740/KAGGLE/DSV/1019469.

ORCID
Shamik Tiwari © https://orcid.org/0000-0002-5987-7101

REFERENCES
1. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. Philos Trans R Soc B Biol Sci. 2012;367(1604):2864-2871.
2. Cascella M, Rajnik M, Cuomo A, Dulebohn SC, di Napoli R. Features, evaluation and treatment coronavirus virus (COVID-19). StatPearls [internet]. Treasure Island (FL): StatPearls Publishing; 2020. https://www.ncbi.nlm.nih.gov/books/NBK554776/.
3. Guo YR, Cao QD, Hong ZS, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak—an update on the status. Mil Med Res. 2020;7(1):1-10.
4. https://www.worldometers.info/coronavirus/
5. Singhal T. A review of coronavirus disease-2019 (COVID-19). Indian J Pediatr. 2020;87(4):281-286.
6. Kong W, Agarwal PP. Chest imaging appearance of COVID-19 infection. Radiol Cardiothorac Imaging. 2020;2(1):e200028.
7. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-444.
8. Jain A, Tiwari S, Sapra V. Hands on Deep Learning with Python Programming. Germany: Lambert Academic Publishing House; 2018. 978-613-9-91551-4. https://www.lap-publishing.com/catalog/details/store/gb/book/978-613-9-91551-4/hands-on-deep-learning-with-python-programming.
9. Iesmantas T, Alzbutas R. Convolutional capsule network for classification of breast cancer histology images. International Conference Image Analysis and Recognition. Cham, Switzerland: Springer; 2018:853-860.
10. Apostolopoulos ID, Mpesiana TA. COVID-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. Phys Eng Sci Med. 2020;43(2):635–640.
11. Hall LO, Paul R, Goldgof DB, Goldgof GM. Finding COVID-19 from chest X-Rays using deep learning on a small dataset. arXiv preprint arXiv:2004.02060; 2020.
12. Zhang J, Xie Y, Li Y, Shen C, Xia Y. COVID-19 screening on chest X-Ray images using deep learning based anomaly detection. arXiv preprint arXiv:2003.12338; 2020.
13. Ucar F, Korkmaz D. COVIDiagnosis-net: deep Bayes-SqueezeNet based diagnostic of the corona virus disease 2019 (COVID-19) from X-ray images. Med Hypotheses. 2020;140 (2020):1–12.
14. Khalifa NEM, Taha MHN, Hassaniene AE, Elghamrawy S. Detection of corona virus (COVID-19) associated pneumonia based on generative adversarial networks and a fine-tuned deep transfer learning model using chest X-Ray dataset. arXiv preprint arXiv:2004.01184; 2020.
15. Salman FM, Abu-Naser SS, Alajrami E, Abu-Nasser BS, Alashqar BA, COVID-19 Detection Using Artificial Intelligence; United States: The DSpace Institutional Digital Repository System; 2020. http://dstore.alazhar.edu.ps/xmlui/handle/123456789/587.
16. Sethy PK, Behera SK. Detection of corona virus disease (COVID-19) based on deep features. Preprints, 2020030300; 2020.
17. Meghdid HS, Asaad AT, Ghafour KZ, Sadiq AS, Khan MK. Diagnosing COVID-19 pneumonia from X-Ray and CT images using deep learning and transfer learning algorithms. arXiv preprint arXiv:2004.00038; 2020.
18. Bassi PR, Attux R. A deep convolutional neural network for COVID-19 detection using chest X-rays. arXiv preprint arXiv: 2005.01578; 2020.
19. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR. Automated detection of COVID-19 cases using deep networks with X-ray images. Comput Biol Med. 2020;121:103792.
20. Mei X, Lee HC, Diao KY, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nat Med. 2020;26(8):1224-1228.
21. Li L, Qin L, Xu Z, et al. Using artificial intelligence to detect COVID-19 and community-acquired pneumonia based on pulmonary CT: evaluation of the diagnostic accuracy. Radiology. 2020;296(2):E65.
22. Toraman S, Alakus TB, Türkoglu İ. Convolutional CapsNet: a novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks. Chaos Solitons Fractals. 2020;140:110122.
23. Shoeibi A, Khodatars M, Alizadehsani R, et al. Automated detection and forecasting of COVID-19 using deep learning techniques: a review. arXiv preprint arXiv:2007.10785; 2020.
24. Tiwari S. A blur classification approach using deep convolution neural network. Int J Inf Syst Model Des. 2020;11(1):93-111.
25. Tiwari S. A comparative study of deep learning models with handcraft features and non-handcraft features for automatic plant species identification. Int J Agric Environ Inf Syst. 2020;11(2):44-57.
26. Hinton GE, Krizhevsky A, Wang SD. Transforming auto-encoders. International Conference on Artificial Neural Networks. Berlin, Heidelberg: Springer; 2011:44-51.
27. Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. Adv Neural Inf Proces Syst. 2017;10(3):3856-3866.
28. Hinton GE, Sabour S, Frosst N. Matrix Capsules with EM Routing; 2018:1–15. https://openreview.net/forum?id=HJWlF9bWb&noteid=rk5MadsMf&noteid=rk5MadsMf.
29. Paolletti ME, Haut JM, Fernandez-Beltran R, et al. Capsule networks for hyperspectral image classification. IEEE Trans Geosci Remote Sens. 2018;57(4):2145-2160.
30. Tiwari S. Dermatoscopy using multi-layer perceptron, convolution neural network, and capsule network to differentiate malignant melanoma from benign nevus. Int J Healthc Inf Syst Inform. 2021;16(3):58-73.
31. Kumar AD. Novel deep learning model for traffic sign detection using capsule networks. arXiv preprint arXiv:1805.04424; 2018.
32. Dadario AMV. COVID-19 X rays. Kaggle. https://doi.org/10.34740/KAGGLE/DSV/1019469.
33. Tiwari S. An analysis in tissue classification for colorectal cancer histology using convolution neural network and colour models. *Int J Inf Syst Model Des*. 2018;9(4):1-19.
34. Juba B, Le HS. Precision-recall versus accuracy and the role of large data sets. *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol 33; 2019:4039-4048.

**How to cite this article:** Tiwari S, Jain A. Convolutional capsule network for COVID-19 detection using radiography images. *Int J Imaging Syst Technol*. 2021;31:525–539. https://doi.org/10.1002/ima.22566