COVID-19 Detection from CT Scan Images using Transfer Learning Approach

Navjot Singh Bajaj
Pavinder Yadav
20mma015@nith.ac.in
pavinder_phdmath@nith.ac.in
National Institute of Technology Hamirpur
Hamirpur, Himachal Pradesh, India

Nidhi Gupta*
National Institute of Technology Kurukshetra
Kurukshetra, Haryana, India
nidhi.gupta@nitkkr.ac.in

ABSTRACT

In the past years, since 2020, the outbreak of COVID-19 has alarmed the world with the speed and its spread around the world. This raised the demand of early, accurate and automated detection system for the COVID-19 as there is a scarcity of manpower in medical field. This attracted many researches using deep learning to build COVID-19 detection model. For the diagnosis of COVID-19, computed tomography scanning are being used as more accurate, non-invasive and efficient method in real-time. In this work, we have proposed a model using six different image classification techniques of deep learning on CT scan images and compared the accuracy to find the most suitable and reliable model for transfer learning to achieve best result on ResNet50 as 97.19% training and 98.05% testing accuracy. The model will automate the process of detection of the COVID-19, leading to the advancement in the field of smart health-care.

CCS CONCEPTS

• **Computing methodologies → Neural networks**.

KEYWORDS

Deep Convolutional Neural Network, Computer Vision, Lung Segmentation, COVID-19 Detection, Medical Imaging

ACM Reference Format:
Navjot Singh Bajaj, Pavinder Yadav, and Nidhi Gupta. 2024. COVID-19 Detection from CT Scan Images using Transfer Learning Approach. In *2024 The 8th International Conference on Machine Learning and Soft Computing (ICMLSC 2024), January 26–28, 2024, Singapore, Singapore*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3647750.3647774

1 INTRODUCTION

COVID-19 is a respiratory infectious disease caused by SARS-Cov-2 virus. It was first detected in the year of 2019 in Wuhan, China and started spreading all over the world rapidly and later declared as a pandemic by the World Health Organization (WHO) in March 2020. This led to global and social disturbance on a large-scale. The common symptoms of COVID-19 are fever, dry cough, fatigue, breathing difficulties, loss of smell and taste, and headache. The severity of COVID-19 includes pneumonia and acute respiratory distress syndrome. The most of the method for diagnosis is real-time reverse transcription polymerase chain reaction (RT-PCR) [4]. Despite the standard detection method, RT-PCR has been observed too sensitive for the contamination process, that is, even the presence of DNA traces could produce false results [5]. Seeing the limitations of RT-PCR [7], the computed tomography (CT) scanned images are considered more trustworthy by the radiologists and medical experts.

The computed tomography (CT) scanned images is basically a picture of the inside the chest, which is generated by small amount of ionizing radiation. The main purpose of the CT scanned images is to examine the lungs, heart and chest wall and detect any abnormality such as pneumonia, COVID-19 and cancer. The demand of CT scanned images have been increased due to the efficiency, accuracy, and feasibility in real-time [11].

Recently, deep learning has become the most popular and powerful tool for the prediction and classification tasks in the computer vision applications. It is a technique that simulates the human brain to carry out any assigned task. The need of minimum human intervention makes it more accurate. Due to the hasty increase in the computational power and efficient storage systems, the deep learning techniques are in demand [15]. Convolutional Neural Network (CNN) is a deep learning method mainly used to analyze and process images such as classification, recognition, etc. Image classification is the process of grouping similar pixels in an image according to some particular conditions or properties. These methodologies are utilized in a wide range of applications, such as traffic monitoring [17], weapon detection [18], self-driving vehicles [6], natural language processing [13], and many others.

In this article, six image classification models are used to detect COVID-19 from the CT scanned images and the accuracy is analyzed. On the basis of images obtained from chest CT scans, we proposed constructing an automated and intelligent method for making the distinction between COVID-19 patients and non-patients. Before taking each image through the training process using deep learning models, we pre-processed each. In addition to this, amount of the data has been augmented, and a feature selection strategy has been employed in order to identify relevant and significant trainable characteristics for the CNN model. The models

*Corresponding Author.

ICMLSC 2024, January 26–28, 2024, Singapore, Singapore

ACM ISBN 979-8-4007-1654-6/24/01
https://doi.org/10.1145/3647750.3647774

for image classification are trained using just a selection of the features available, and then their accuracy is evaluated and compared in considerable detail. Our algorithm can read the structure of a chest CT scan in an instant, can utilize hidden patterns to identify patients with COVID-19, and it can decrease the number of human preprocessing steps that are required. The experimental findings reveal that the x-ray image dataset has high levels of accuracy, precision, and recall, as well as an F-1 score.

2 DEEP LEARNING ARCHITECTURE

Deep learning methods are the form of machine learning that makes use of computing models inspired by the structure of the human brain to carry out the tasks of classification and prediction. The neural network is an integral part of deep learning. In general, it has more than three layers. Convolutional Neural Network (CNN) is a form of artificial neural network, mainly used to process pixel data. It derives its name from the process termed as convolution. In mathematics, convolution is an operation on two functions (f and g) which produce third function ($f * g$) that expresses how the shape of one is modified [19]. In CNN, convolution (also known as convolutional layer) is a process where a small matrix of numbers (kernel or filter) is taken and passed over the image and the image gets transformed depending on the values of the filter. This transformed image is termed as feature map. The formula to calculate the value of feature map values is defined by Equation 1,

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k] \quad (1)$$

where, f denotes the input image, h denotes the kernel, m and n denotes the rows and columns of the feature map matrix, respectively.

A CNN architecture has several convolutional and pooling layers followed by the output layer called the fully connected layer. The final classification and flattening of the vector, that is the arrangement of 3D volume of numbers into a 1D vector, received from the preceding layer takes place in this layer.

2.1 Image Classification

Image classification is the process of categorizing and sorting the pixels within an image based on specific rules and conditions [12]. It has accelerated technological advancement in the many fields including auto-mobile, health-care, manufacturing, etc. There are two types of image classification, namely single-label classification and multi-label classification. Single-label classification model is used where each image contains only one label. For example, image of an apple, ring, etc. multi-label classification is a categorizing technique where more than one label is present in the image. For example, an image of a road can contain different labels like vehicles, traffic lights, shops, person etc.

The image classification models are developed by varying either the size of the filter used in various layers, the number of convolutional, pooling and fully connected layers or both. The image classification models used in this work are,

- ResNet50
- DenseNet121
- DenseNet169
- DenseNet201
- VGG16
- VGG19

2.1.1 *ResNet50.*

Residual Network (ResNet) is a neural network architecture that uses the technique of skip connections. The skip connection is a technique where the original input value is added to the output of the layer. This is useful to tackle the problem of vanishing gradient while training the neural network [8]. In the Fig. 1, it is visible that x is original input, $F(x)$ is output from first layer, and $F(x) + x$ will be the input for next layers.

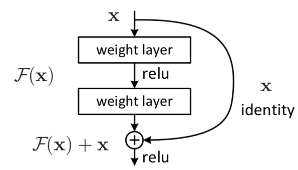


Figure 1: Illustration of Skip Connection.

ResNet50 is the form of ResNet model that has 48 convolution layers as shown in Fig. 2 .

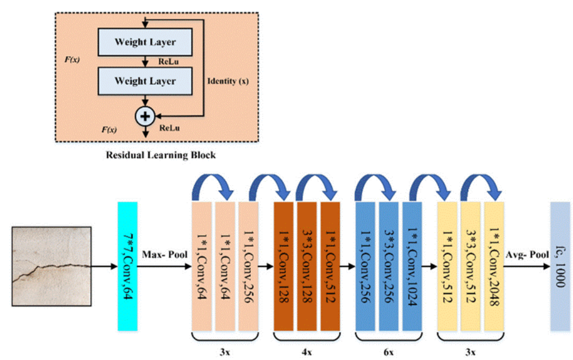


Figure 2: ResNet50 Architecture [8].

2.1.2 *DenseNet121.*

DenseNet or Densely Connected Convolutional Network is an image classification modeling which each layer obtains increased input from all prior layers, such that every level is interconnected with every other level and transmits its very own feature maps to every following levels [9]. DenseNet concatenates and utilizes image features from all preceding levels as inputs, resulting in reduced required parameters and repetitive image features discarded due to feature reuse. Therefore, the i^{th} layers accepts the image features from all the previous layers, $y_0, \ldots y_{i-1}$, as the input as in equation below. The architecture of DenseNet121 is shown in Fig. 3.

$$y_i = F_i([y_0, y_1, y_2, ..., y_{i-1}]) \quad (2)$$

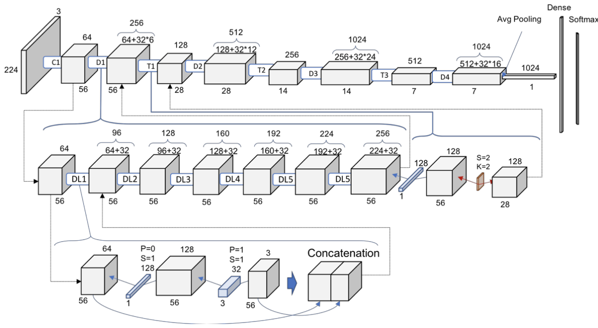


Figure 3: DenseNet121 Architecture [9].

2.1.3 DenseNet169. DenseNet169 is a 169 layer deep DenseNet architecture [9].As shown in Fig. 4, The DenseNet-169 contains basic convolution layer with 64 filters of size 7X7 and Max Pooling Layer of size 3X3, followed by dense block 1 which has 2 convolutions (1X1 and 3X3, respectively) repeated 6 times, 1 transition layer having one 1X1 convolutions and one 2X2 AvgPool, 2 dense block has 2 convolutions (1X1 and 3X3, respectively) repeated 12 times, 2 transition layer (one 1X1 Conv and one 2X2 AvgPool), 3 dense block with 2 convolutions (1X1 and 3X3, respectively) repeated 32 times, 3 transition layer (one 1X1 Conv and one 2X2 AvgPool, 4 dense block having 2 convolutions (1X1 and 3X3, respectively) repeated 32 times, global Average Pooling layer of size 7X7 accepts all the feature maps of the network to perform classification and 1 fully connected layer with 1,000 nodes that makes a total of 169 layers.

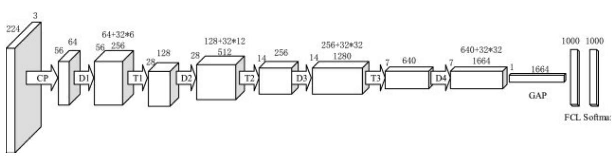


Figure 4: DenseNet169 Architecture [9].

2.1.4 DenseNet201. DenseNet201 is another enhanced variant of DenseNet architecture having 201 deep layers [16]. As shown in Fig. 5, the 201 layers constitute the DenseNet201 architecture, are basic convolution layer with 64 filters of size 7X7 and Max Pooling Layer of size 3X3, followed by 1 dense block which has 2 convolutions (1X1 and 3X3, respectively) repeated 6 times, 1 transition layer having one 1X1 convolutions and one 2X2 AvgPool, 2 dense block has 2 convolutions (1X1 and 3X3, respectively) repeated 12 times, 2 transition layer (one 1X1 Conv and one 2X2 AvgPool), 3 dense block with 2 convolutions (1X1 and 3X3, respectively) repeated 48 times, 3 transition layer (one 1X1 Conv and one 2X2 AvgPool, 4 dense block having 2 convolutions (1X1 and 3X3, respectively) repeated 32 times, global Average Pooling layer of size 7X7 accepts all the feature maps of the network to perform classification and 1 fully connected layer with 1,000 nodes.

2.1.5 VGG16. VGG (Visual Geometry Group) consisted of blocks composed of 2D convolution and max pooling layers. It is still one of the most used image classification model [14]. VGG16 is a variant

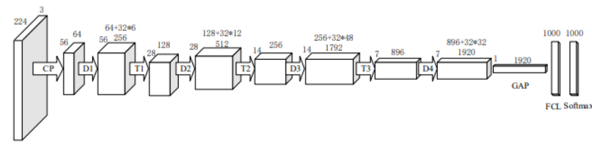


Figure 5: DenseNet201 Architecture [16].

of VGG model with 16 convolutional layers. As shown in Fig. 6, the 16 layers of VGG16 are two convolution using 64 filters along with max pooling layer, two convolution using 128 filters and max pooling layer, three convolution using 256 filters followed by max pooling layer, two sets of three convolutions using 512 filters and max pooling layer. At the end, three Fully Connected layers are present in the network, first two have 4,096 nodes and the last is the output layer with Soft-Max activation with 1,000 nodes.

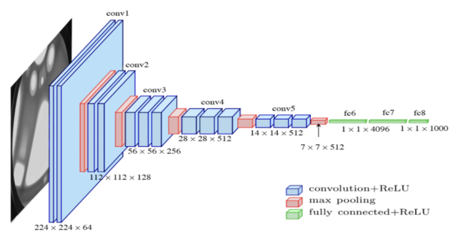


Figure 6: VGG16 Architecture [14].

2.1.6 VGG19. VGG19 is other form of VGG architecture that is 19 layers deep [3]. As shown in Fig. 7, the 19 layers of VGG19 are two convolution using 64 filters along with max pooling layer, two convolution using 128 filters and max pooling layer, four convolution using 256 filters followed by max pooling layer, two sets of four convolution using 512 filters and max pooling layer. At the end three Fully Connected layers are present in the network, first two have 4,096 nodes and the last is the output layer with Soft-Max activation with 1,000 nodes.

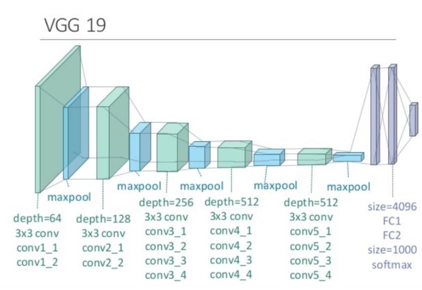


Figure 7: VGG19 Architecture [3].

3 PROPOSED METHODOLOGY

The dataset is collected from three different sources. The first one is the data collected by Joseph Paul Cohen and Paul Morrison and Lan Dao for their research [1], the second one is collected by Kermany and team for their research Medical Diagnoses and

Proposed Framework steps

1. Reading the CT-Scan images from the dataset.
2. Augmenting the images to increase the volume of the dataset.
3. Performing image pre-processing.
   - Resize the image to 224*224*3.
   - Split the dataset into two parts training part and a testing part (the dataset was split into 75% and 25% for training and testing, respectively)
4. Extracting the features and classifying the images using the deep learning proposed model:
   - ResNet50
   - DenseNet121
   - DenseNet169
   - DenseNet201
   - VGG16
   - VGG19
5. Optimizing and fit functions were used to train and validate the proposal models (Optimiser- Adam, Epoches-30 and Loss function- categorical_crossentropy)
6. Calculate the validation performance metrics, training and testing accuracies and losses.
7. Comparing the resulted metrics with other frameworks.

Figure 8: The Proposed Framework Steps Used.

Treatable Diseases by Image Based Deep Learning [10] and the last one is collected by Radiological Society of North America [2]. A quick algorithm overview of the steps of the framework that have been used is shown in Fig. 8.

3.1 Dataset and Pre-Processing

The original dataset consisted of 3,995 COVID-19 infected and 4,038 normal CT scanned images for training purpose. This makes total of 8,033 images in the training dataset. There are 532 COVID-19 infected and 541 normal X-Rays in total 1,073 images consists testing dataset. Data augmentation is a process of increasing the volume and diversity of data without collecting new data. In the work, several data augmentation techniques are being used as rotation, horizontal flipping and shearing. In this work, the dataset are augmented using the rotation, magnification, shifting, shearing, and flipping. An image from the original dataset gets augmented into four images, as seen in the Fig. 9. After augmentation, the training dataset have in total 18,032 images, where 9,224 are COVID-19 infected and 8,808 are normal chest X-Rays, while among rest of 5,450 images, 2,725 images are COVID-19 infected and rest 2,725 images are normal in the testing dataset.

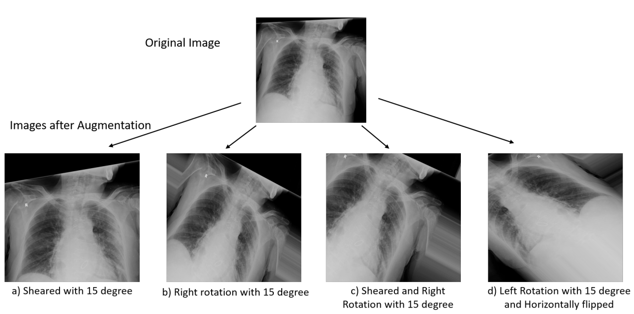


Figure 9: Data Augmentation.

3.2 Feature Extraction and Selection

Feature extraction is the process of extracting the parameters from the images on which the model is going to be trained. Significantly, in the model, the number of features is same as the number of pixels in the image. On the extraction of features from the dataset using CNN model, we observed 2,37,88,418 parameters. Feature selection is a process of selecting those parameters that meaningfully contribute in giving the final output value. In other words, feature selection is the process of removing the irrelevant or partially relevant features that reduces the efficiency of the performance of the model. The Recursive Feature Elimination (RFE) procedure evaluates features based on certain assessment of their value. At each stage, each feature is evaluated to see how important it is, and the ones that aren't as important are discarded. In the sequential elimination process, recursion is needed because the relative value of each feature could change when it is looked at across a new set of features. RFE was used to attain the highest accuracy on every classification model. In this work, RFE has been used for feature selection to obtain 2,00,706 trainable features as the output.

4 RESULTS AND DISCUSSION

The Transfer Learning approach has been used to train deep learning architectures used in this research. Transfer Learning is the process in which the network,pre-trained on one database, successfully transfers its learned features to a new database, thereby helping in increasing the overall performance of the model. The ResNet50, DenseNet121, DenseNet169, DenseNet201, VGG16 and VGG19 models were trained on the x-ray images database, as the baseline model, and the learning was transferred to CNN.

The images from the augmented dataset are read and resized to 224X224X3. The dataset is split into training and testing datasets with a ratio of 3:1. The features are extracted from the dataset and go through a selection process known as feature selection. The x-ray images are then classified on the basis of these selected features. The accuracy of all the six models on the training and testing dataset is listed in the Table 1.

Table 1: Training and Testing Accuracy of the Models.

Model	Training Accuracy	Testing Accuracy
ResNet50	97.19%	98.05%
DenseNet121	89.50%	98.20%
DenseNet169	89.87%	97.66%
DenseNet201	90.38%	95.31%
VGG16	93.12%	95.51%
VGG19	96.88%	96.48%

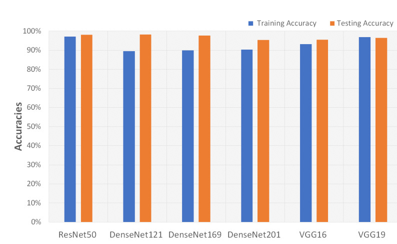
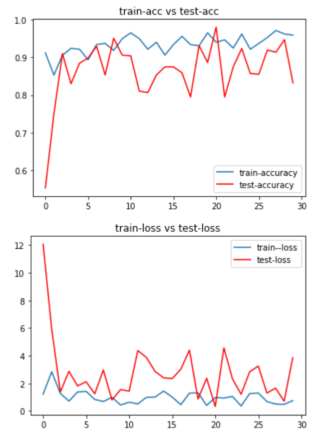
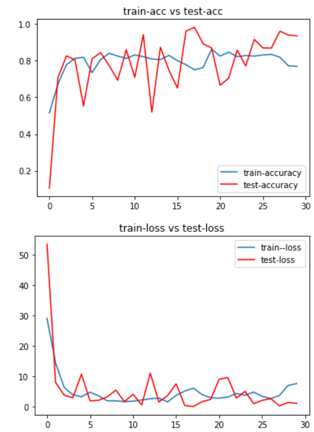


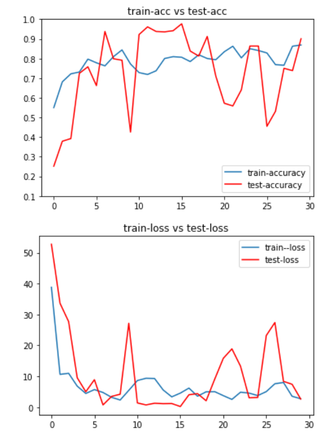
Figure 10: Accuracy Performance Metric of The Applied Deep Learning Models.



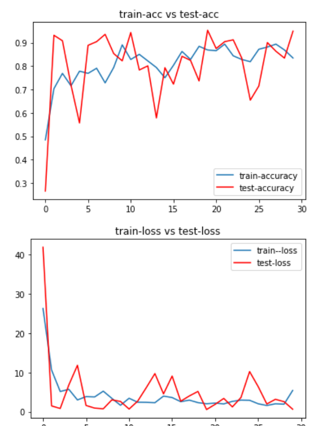
(a) ResNet50



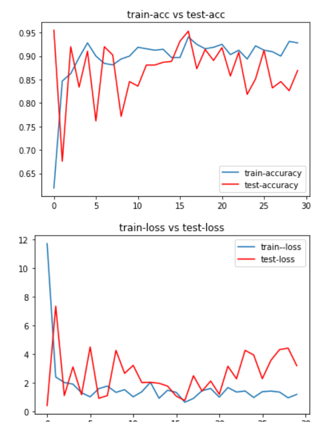
(b) DenseNet121



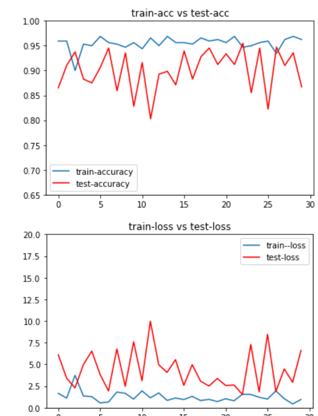
(c) DenseNet169



(d) DenseNet201



(e) VGG16



(f) VGG19

Figure 11: Accuracy and Data Loss Graphs of Models.

From the above Table, it is shown that ResNet50 gives good accuracy for training as well as testing dataset as of 97.19% and 98.05%, respectively. The maximum testing accuracy of 98.20% is achieved by DenseNet121, but the training accuracy of the same is 87.50%, which is not acceptable performance, respective to the detection model. Furthermore, after increasing more layers, using DenseNet169 and DenseNet201, the training accuracy has been examined and 87.87% and 89.38% accuracies are observed, respectively, while the testing accuracies are 97.66% and 95.31%, respectively. The lowest testing accuracy is achieved by DenseNet201. Therefore, this signifies that addition of more layers or deeper CNN models do not impact on the performance and hence do not improve the accuracy. Also, VGG19 model is able to achieve a good accuracy of 96.88% and 96.48% on the training and testing dataset both, while VGG16 achieved 93.12% and 95.51% on the training and testing dataset, respectively.

The reason behind the difference between the training and testing accuracy of DenseNet121, DenseNet169 and DenseNet201 is due to the augmented data as there may be some images that are creating redundancy and some images which are hard to train. Also, testing set has images which are used for the augmentation. To overcome this problem, the models like Resnet50, VGG16 and VGG19 are used. Another reason is the impact of dropout function, which sets a subset of features to 0 during training process, while it use all features during testing phase and are scaled appropriately. As a result, the model is proven as more robust during testing, which can lead to higher testing accuracy. Fig. 10 illustrates a comparison graph that was constructed using all six trained models.

Figure 11 shows the graph plots between training and testing accuracy and data loss for every model. It is seen that there are many spikes in the testing accuracy and testing data loss. The spikes are an expected side effect of Adam's Mini-Batch Gradient Descent. Some mini batches contain unexpectedly some unfavorable data for the optimization and hence generates the spikes. The proposed model has only two classes, which makes it slightly biased with the testing data.

5 CONCLUSION

In this work, the data has been augmented to increase the size of the dataset followed by feature extraction and feature selection using predefined CNN models and to avoid over-fitting. The models including ResNet50, DenseNet121, DenseNet169, DenseNet201, VGG16 and VGG19 are trained on the selected trainable features on the augmented dataset. The best accuracy is observed by ResNet50 as 97.19% in training and 98.05% in testing. The proposed model by no means is a commercially ready solution as the privacy concerns. The non-availability of medical dataset is the major issue faced while development of reliable models in the field of medical science. However, the collection of new data will help to increase the variability in the datasets and to increase the accuracy. Researchers and data scientists may utilize this work to facilitate the construction of highly accurate and trustworthy models.

REFERENCES

[1] Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong, and Marzyeh Ghassemi. 2020. Covid-19 image data collection: Prospective predictions are the future. *arXiv preprint arXiv:2006.11988* (2020).

[2] Kaggle Dataset. 2020. COVID-19 radiography database. https://www.kaggle.com/tawsifurrahman/COVID19-radiography-database/.

[3] Nilanjan Dey, Yu-Dong Zhang, V Rajinikanth, R Pugalenthi, and N Sri Madhava Raja. 2021. Customized VGG19 architecture for pneumonia detection in chest X-rays. *Pattern Recognition Letters* 143 (2021), 67–74.

[4] Lilit Garibyan and Nidhi Avashia. 2013. Research techniques made simple: polymerase chain reaction (PCR). *The Journal of investigative dermatology* 133, 3 (2013), e6.

[5] Ghulam Gilanie, Usama Ijaz Bajwa, Mustansar Mahmood Waraich, Mutyyba Asghar, Rehana Kousar, Adnan Kashif, Rabab Shereen Aslam, Muhammad Mohsin Qasim, and Hamza Rafique. 2021. Coronavirus (COVID-19) detection from chest radiology images using convolutional neural networks. *Biomedical Signal Processing and Control* 66 (2021), 102490.

[6] Abhishek Gupta, Alagan Anpalagan, Ling Guan, and Ahmed Shaharyar Khwaja. 2021. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array* 10 (2021), 100057.

[7] Joseph Hadaya, Max Schumm, and Edward H Livingston. 2020. Testing individuals for coronavirus disease 2019 (COVID-19). *Jama* 323, 19 (2020), 1981–1981.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 770–778.

[9] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 4700–4708.

[10] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. 2018. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell* 172, 5 (2018), 1122–1131.

[11] HY Li, CM Zhang, YY Lv, WQ Dai, B Xu, and XM Qi. 2020. Value of chest imaging in the newborn with suspected COVID-19. *Eur Rev Med Pharmacol Sci* 24, 22 (2020), 11971–11976.

[12] Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson. 2019. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing* 57, 9 (2019), 6690–6709.

[13] Daniel W Otter, Julian R Medina, and Jugal K Kalita. 2020. A survey of the usages of deep learning for natural language processing. *IEEE transactions on neural networks and learning systems* 32, 2 (2020), 604–624.

[14] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[15] DY Wan, XY Luo, W Dong, and ZW Zhang. 2020. Current practice and potential strategy in diagnosing COVID-19. *Eur Rev Med Pharmacol Sci* 24, 8 (2020), 4548–4553.

[16] Shui-Hua Wang and Yu-Dong Zhang. 2020. DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16, 2s (2020), 1–19.

[17] Myounggyu Won, Sayan Sahu, and Kyung-Joon Park. 2019. DeepWiTraffic: Low cost WiFi-based traffic monitoring system using deep learning. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS).* IEEE, 476–484.

[18] Pavinder Yadav, Nidhi Gupta, and Pawan Kumar Sharma. 2022. A comprehensive study towards high-level approaches for weapon detection using classical machine learning and deep learning methods. *Expert Systems with Applications* (2022), 118698.

[19] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. 2018. Convolutional neural networks: an overview and application in radiology. *Insights into imaging* 9, 4 (2018), 611–629.