Detection of COVID-Lung Cancer through X-ray data through Distributed Deep Convolutional VGGNet and ResNet Models

Syamala Bathula[1*], K Sreenivasa Rao[2]

Abstract

When compared to the general population, lung cancer patients have a higher incidence of COVID-19 infection, pulmonary problems, and poorer survival results. As a reference for prioritising cancer care issues during the epidemic, the world's main professional organisations issued new recommendations for the diagnosis, treatment, and follow-up of lung cancer patients. In the modern world, we are battling COVID-2019, a coronavirus-driven pandemic that is among the worst in human history. If the infection is discovered early, the patient can receive treatment right away (before it enters the lower respiratory tract). once the infection has reached the lungs, to look for ground-glass opacity on the chest X-ray caused by fibrosis. Based on the significant differences in the X-ray images of an infected and non-infected person, artificial intelligence systems can be used to determine the presence and severity of illness. For this study, I employed feature extraction from Transfer Learning, which entails importing a pre-trained CNN model, such as Distributed Deep Convolutional VGGNet or Distributed Deep Convolutional with ResNet Model, and changing the last layer to meet my needs.Using Distributed Deep ConvolutinalVGGNet, the model can get an F1-score of 0.88, which is the best among all pretrained models. Furthermore, the X-rays suggesting COVID-19 are divided into three categories based on severity: mild, medium, and severe. Because precision and recall are both essential factors in this study, the data are analysed using the F1-Score. The confusion matrix and the results for the F1-Score, Precision, Recall, and overall Accuracy are also supplied to give a full analysis of the Model performance.As a warning to society, the proposed approaches have had a considerable impact upon the country.

KeyWords:covid x rays, ,Resnet50,VGG16 Covid-19,x-rays,ML,DL,CT hadoop
DOI Number: 10.14704/NQ.2022.20.12.NQ77143 **NeuroQuantology 2022; 20(12):1629-1641**

Introduction

A lot is still unknown about COVID-19 because the data and study are still in their early phases, especially in terms of recovery, immunity, transmission, and connections to other underlying diseases. It is true that some people are more likely to contract a virus with a more severe strain, according to experts. Doctors can only speculate about the appropriate course of action because there isn't a formal study that ties the coronavirus to lung cancer. The new coronavirus has not yet been proven to be a direct cause of lung cancer. An early Wuhan investigation found that the risk of COVID-19 infection was nearly twice as high in lung cancer patients as in the general population.

It's unclear whether this is the effect of these patients' therapies, which, like those for many cancer patients, immunocompromised them. According to Dr. Adil Akhtar, a specialist in oncology and palliative care and an associate professor of medical oncology and haematology at the University of Auckland-William Beaumont School of Medicine, there may be a higher incidence of COVID-19 in patients with lung cancer. Lung cancer is the most common cancer, claims a Chinese study.According to Dr. Wasif M. Saif, chief medical officer and medical

Corresponding author: SyamalaBathula
Address: [1]Research Scholar, Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation,Hyderabad - 500075, India & Assistant Professor Department of Computer Science and Engineering, Vasavi College of Engineering Hyderabad, India, [2]Department of Computer Science and Engineering, KoneruLakshmaiah Education Foundation, Hyderabad, Telangana- 500075

E-mail: syamala.b@staff.vce.ac.in, ksrao517@gmail.com

director of Northwell Health Cancer Institute, "COVID-19 disproportionately harms frail people, notably the elderly and persons with comorbidities, including immunocompromised people like cancer patients." Dr. Brendan Stiles, a thoracic surgeon at New York Presbyterian Hospital and an associate professor of cardiothoracic surgery at Weill Cornell Medicine, noted that many lung cancer patients are older, have severe lung disease, and have diminished lung capacity, even though lung cancer does not increase the risk of coronavirus complications. These patients almost certainly develop lung sickness if the coronavirus creates issues.

Additionally, it's likely that chemotherapy-induced immunosuppression in lung cancer patients makes them more vulnerable to catching the coronavirus and suffering catastrophic consequences as a result of infection. The respiratory system as well as other organs in the lungs are both affected by lung diseases, also referred to as respiratory disorders [1]. Pneumonia, tuberculosis, and coronavirus infections are all lung illnesses (COVID-19). The International Respiratory Society Forum estimates that 334 million people worldwide have asthma, 1.4 million die from tuberculosis annually, 1.6 million from lung cancer, and 1.6 million from pneumonia. As a result of the bloodshed, thousands of people have died. Millions of people have been infected by COVID-19, which has put strain on healthcare systems worldwide [3, 4].

One of the leading causes of mortality and disability worldwide is lung disease. Early detection is essential for better long-term survival and recovery [5, 6]. In the past, lung issues may be identified via skin testing, blood tests, sputum samples [7], chest X-rays, and computed tomography (CT) [8]. In-depth instruction has lately demonstrated potential in the diagnosis of ailments, notably lung diseases, using medical imaging. A subset of machine learning called "deep learning" focuses on procedures that are modelled after the structure and operation of the brain. Recent advances in machine learning, particularly deep learning, have made it easier to recognise, quantify, and categorise medical visual patterns [9].

This development has been made possible by in-depth training, which has allowed personnel to learn features only from data instead of manually developing characteristics based on specific field knowledge. The use of deep training as the industry standard for enhancing performance in a variety of medical settings is gaining ground quickly. These developments therefore assist doctors in accurately diagnosing and categorising illnesses [10] [11].

Related Works

Transfer learning has gained popularity in computer vision because it makes it possible to build correct models [12]. A model that has been learned in one domain can be applied to another using transfer learning. It is possible to do transfer learning with or without the use of a learned model [13]. ImageNet is used extensively in CNN architecture training [14]. "The pictures were taken from the internet and given human labels using Amazon Mechanical Turk crowdsourcing. A subset of ImageNet, with roughly 1000 photos in each of its 1000classifications, is used by ILSVRC.

A total of 150,000 testing photographs, 50,000 validation photos, and 1.2 million training photos are accessible. [15]. The weights of the model's upper layers are often frozen, whereas the weights of its lower layers are only retrained [16] Ensemble classification is the practise of combining different classifiers to make a prediction. In comparison to using a single model, ensemble forecasting reduces prediction variance and produces predictions that are more accurate. According to the literature, popular polling, probability score averaging, and stacking are the ensemble approaches employed. [17].

In this part, thirteen datasets from COVID-19 are made available to the general audience. As a result of the COVID-19 outbreak, certain datasets have been made available to the public. In terms of the number of photographs they contain, several of these collections are still expanding. As a result, the datasets may contain less or more images than what is mentioned in this work [31–32]. It should be mentioned that some of the images are accessible through different databases. Future research should therefore search for duplicate images [18].

The severe acute respiratory syndrome coronavirus is the source of the novel and incredibly contagious respiratory disease known as Covid. The COVID-19 virus we currently have has reached every continent. 6,663,304 confirmed cases and 392,802 confirmed deaths are present as of June 7, 2020, globally. It spreads swiftly, and some victims pass away from respiratory failure before the illness fully manifests itself. Research and evaluation of the disease's critical phase are necessary to determine the clinical symptoms,

pathogenesis, and even the course of treatment for COVID-19. The great majority of patients have successfully been released in China. No studies to date have offered an early prognosis of the severity of lung damage and recovery in COVID-19 patients.

The chicken patients were released, many had radiological issues, and some even had lung fibrosis, according to a retrospective study. Low lung function in COVID-19 patients need special monitoring throughout the early stages of recovery. To better comprehend the anticipated clinical outcomes, a retrospective analysis of 57 COVID-19 patients who had been approved but had recovered was carried out. Serial evaluations of lung function, lung imaging, and cardiovascular capacity were performed 30 days following discharge. Additionally, we compared critically ill and non-critically ill individuals using performance metrics.

This section outlines the procedures followed in a recent study that included intensive training to identify lung disease. The technique is shown in a block diagram. For the first time, a relevant database is mentioned as being the main source of article links. Since Scopus is one of the most well-known databases for peer-reviewed scientific publications, it was selected. According to the quantity of citations, many important papers are included; nevertheless, Scopus is not indexed by Google Scholar. A number of pre-printed papers about COVID-19 have been distributed as a result of the recent disease epidemic.

To ensure that it only contains the most current occupations, "this study only includes the most recently publicised positions (2016-2020)." However, there are some older but essential parts. Using pertinent search phrases, all essential documents for the detection of deep lung disorders were located. In-depth training, screening, classification, CNN, lung sickness, tuberculosis, pneumonia, lung cancer, COVID-19, and coronavirus were a few of the terms mentioned. There are only English-language articles in the survey. By the time this phase was over, we had located 366 items. Second, only pertinent works were considered for selection. During the screening, just the right and abstract were considered.

This study's primary selection criterion is articles that use in-depth training techniques to identify pertinent illnesses. Unsuitable members were kicked out. According to the inspection, the short list only has 98 goods. To categorise a picture into healthy lungs and diseased lungs is the main goal of lung disease diagnosis. Training is necessary to obtain the lung disease classifier, also known as a model. Learning is the process of teaching a neural network to recognise a series of images. Learn how to create a model that can categorise obsessed images in accordance with their class labels by attending in-depth training.

As a result, imaging the lungs with the disease to be discovered is the first step in the diagnosis of lung disease employing in-depth training. In the second step, the neural network is trained until it can identify the disease. The fresh pictures need to be categorised as the last stage. The prototype determines the category of person photographs when it gets new, previously unseen images. A high-level explanation of the method is shown in Figure 1. According to the justification in section 2, the study was constrained because the majority of the works examined came from authors who had their works included in the Scopus database.

COVID-19-related posts were allowed as an exception, as most comments made it past the prepress stage at the time of this study. “In terms of publishing years, the most current papers that have been examined are those that were published before October 2020. Something is being done as a result of the conclusions reported in this evaluation report[19]”.

Methods And Materials

Dataset

The data sets utilised in the measured research are described in this section. Compilation of data sets for the diagnosis of COVID-19, lung cancer, pneumonia, and tuberculosis. This is done to give readers useful information about the data sets. Tables only include openly accessible public data sets; private data sets require access authorization. Several organisations have recently acknowledged the use of X-rays in deep machine learning approaches to identify COVID-19 pneumonia. However, the majority of these studies were based on a very tiny data set that had only a few COVID-19 samples

Because of this, it is difficult to summarise the findings from these research, and there is no assurance that the observed performance will hold true when these models are tested on a larger data set [13]

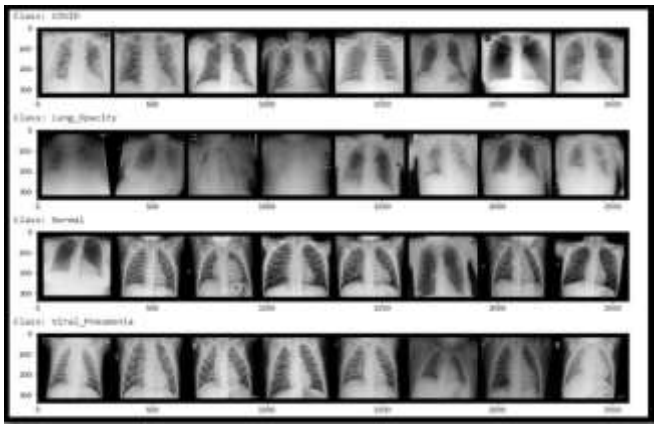


Fig:1 sample data set

COVID-19 comprises thirteen publicly available data sets, as indicated in Table 8. Several data sets were made available as a result of the COVID-19 outbreak. In terms of the amount of images in each of these collections, several of them are still growing. As a result, the number of pictures in the data sets may differ from what is stated in this text. It's worth noting that some of the photographs are available in a variety of databases. As a result, in the future, researchers will need to check for duplicate photographs. Summarize the research you've done using taxonomy.Customers can quickly obtain content that is relevant to their interests as a result of this. The next section looks at the distribution of functions based on taxonomic features that have been identified. There are atotal of 21,165 samples, which are divided into four categories: Covid-19, opacity of the lungs, normal, and viral pneumonia are all examples of viral pneumonia.

All of the photographs are saved as Portable Network Graphics (PNG) files and are 299x299 pixels in size. According to the most recent update, the Hadoop data security ecosystem now comprises 3,616 positive COVID-19 cases, 10,192 normal, 6,012 lung opacity (non-COVID lung infection), and 1,345 viral pneumonia images.

System Architecture

The input is a CXR image of a dataset with two subsections: COVID-19 patients and normal patients, as shown in Figure 1. This system underwent preprocessing, which includes loading photographs of a specific size, partitioning the dataset, and data augmentation methods, prior to training the model. The accuracy increased once the model was set up and adjusted.

The confusion matrix, model loss, and model accuracy were exhibited to show how loss and accuracy changed over time. Finally, we can determine whether an image in the output component is of a COVID-19patient if a user provides a photo as an input model.

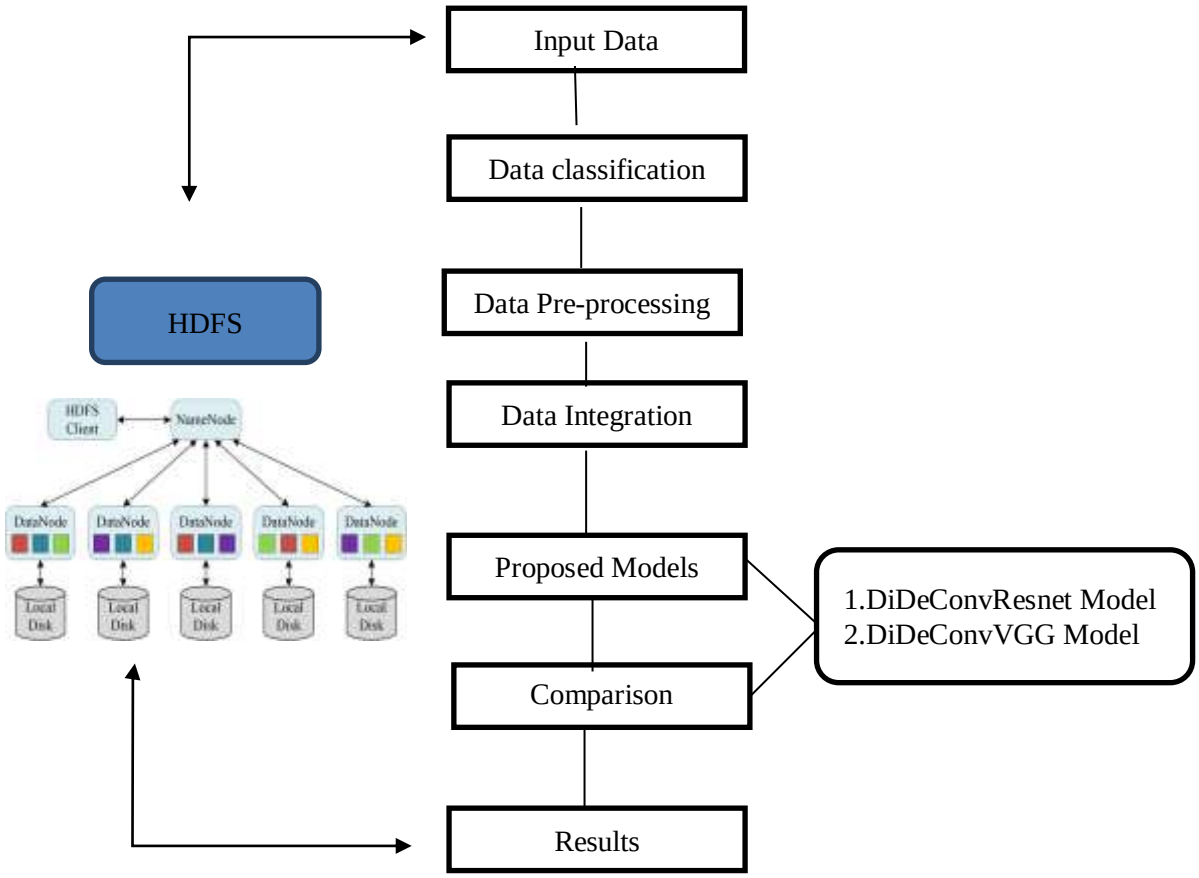


Fig2:System Design

The system is shown in the block diagram in the simplest way feasible. Our system's decision-making component is crucial, and it has a significant role in this investigation. The choice is based on a model for managing and controlling data transmission between nodes via name node, data node, and resource manager that was created using a substantial amount of image data and kept in the Hadoop eco system.

Methods

ConvolutionalNeuralNetworks(CNN)

The proposed approach addresses the issue of early lung diagnostic identification using deep learning methods based on a Hadoop Distributed Deep Convolutional Neural Network (HdiDConvNNet). Our study's objective is to use a number of distributed deep learning techniques to distinguish between CT scan images that are COVID-19 and those that are not. Clinicians can quickly and effectively detect COVID-19 by using automated COVID-19 diagnoses based on CT scan images. The data are analysed using the F1-Score because precision and recall are both crucial elements in this investigation.

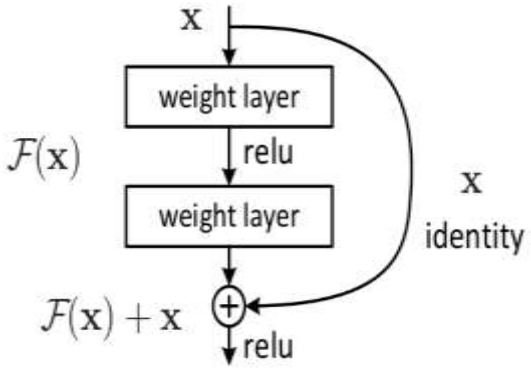
The confusion matrix and the results for the F1-Score, Precision, Recall, and overall Accuracy are also supplied to give a full analysis of the Model performance. A deep neural network (DNN) known as the Hadoop Dis-tributed Deep Convolutional Neural Network (HdiDConvNNet), which was first introduced by Fukushima in 1988 [8], is frequently used in the Hadoop ecosystem for image classification. due to limitations in computational hardware. CNN is capable of properly performing feature extraction, one of the most crucial components of the networks

Distributed Deep Convolutional with ResNetModel (DiDCResNet)

'ResNet50 is a ResNet model version having 48 Convolution layers, 1 Max-Pool layer, and 1 Average Pool layer. Its floating point operations are 3.8 x 109. We have carefully examined the design of ResNet50, a commonly used ResNet model [8]. The question of whether the pattern works arises because "we know that deep convolutional neural networks are particularly great at identifying low,

mid, and high level information from images, and arranging additional layers generally gives us better accuracy." better. added layers

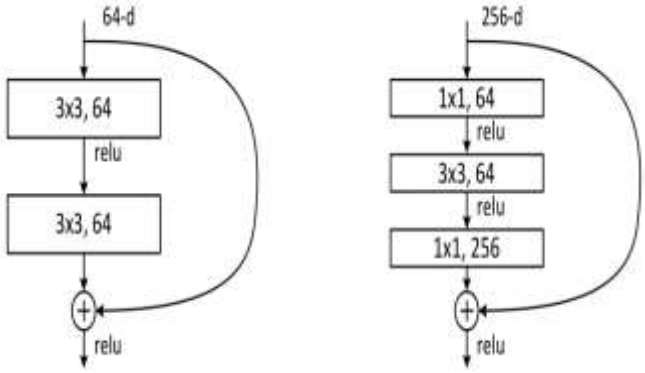
In order to allow networks with dozens of layers to converge, this problem raises the gradient vanishing/exploding problem. However, as deep neural networks begin to converge, we see a second issue in which accuracy saturates and subsequently degrades. Unlike what would be anticipated, over-assembly is not the cause of this issue, because adding more layers to a suitable deep pattern just makes the learning error worse. The authors have developed a sophisticated framework for residual learning to address this issue, and as a result, they are proposing quick access links that just carry out identification matching



layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

The non-linear layers were allowed to fit a different mapping, "F(x): =H(x)x," so that the original mapping became "H(x): =F(x)pressly allowed the layers to fit a residual mapping and labelled that as H(x). Additionally, the model did not receive any additional parameters, and the computational time was controlled thanks to these shortcut identity mappings

“Now we'll talk about Resnet 50, “as well as the architecture for the above-mentioned 18- and 34-layer ResNets, which is given residual mapping but not displayed for clarity. ‘For the ResNet 50 and higher, a little adjustment was made in that previously, shortcut connections skipped two levels, but now they skip three layers, and there were also 1 * 1 convolution layers added, which we will go over in depth with the ResNet 50 Architecture



64 distinct cores with a step size of 2 and a 7 * 7 core size give us 1 layer. The maximum pool with step size 2 is then visible. The next convolution consists of three layers: a 1*1.64 core, a 3*3.64 core, and a 1*1.256 core. These three levels are repeated a total of three times, giving us nine layers in this phase. The following phase is performed four times, giving us 12 layers: a 1*1,128 core, a 3*3,128 core, and eventually a 1*1,512 core.

The following cores are 1*1.256, 3*3.256, and 1*1.1024; these are repeated six times for a total of 18 layers. We now have a total of 9 layers after I core 1*1.512 twice more with two more pieces of 3*3.512 and 1*1.2048. Then we create an intermediate group, finish it with a layer with 1000 fullylinked nodes, and then use a SoftMax function to give us one layer. The activation functions and the maximum/mean unifying layers are actually not taken into account. This results in 1 + 9 + 12 + 18 + 9 + 1 = 50 deep convolutional network layers, to put it briefly

Distributed Deep Concolutional with VGG Model(DiDeConvVGG)

Despite the fact that AlexNet has demonstrated the viability of deep convolutional neural networks, it does not provide a general framework for future researchers to use when creating new networks. We will introduce some of the popular heuristic design principles for deep networks in the parts that follow. Similar advancements in chip design, when engineers switched from using transistors to using logic components and logic blocks, can be seen in this field. Similar to how the design of neural network architectures has evolved, researchers are now thinking in terms of blocky, repetitive layer patterns rather than individual neurons or even entire layers.

The Visual Geometry Group (VGG) at the University of Oxford initially proposed the use of blocks in their eponymous VGG network. Any contemporary deep learning framework can be used to incorporate these repeated structures into code simply by using loops and functions.

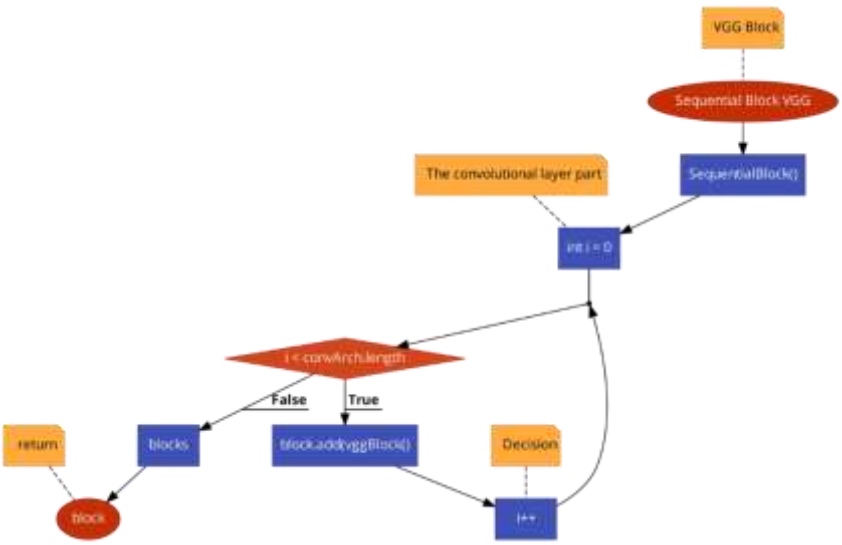


Fig:3 Control Flow GraphDistributed Deep Concolutional with VGG Model

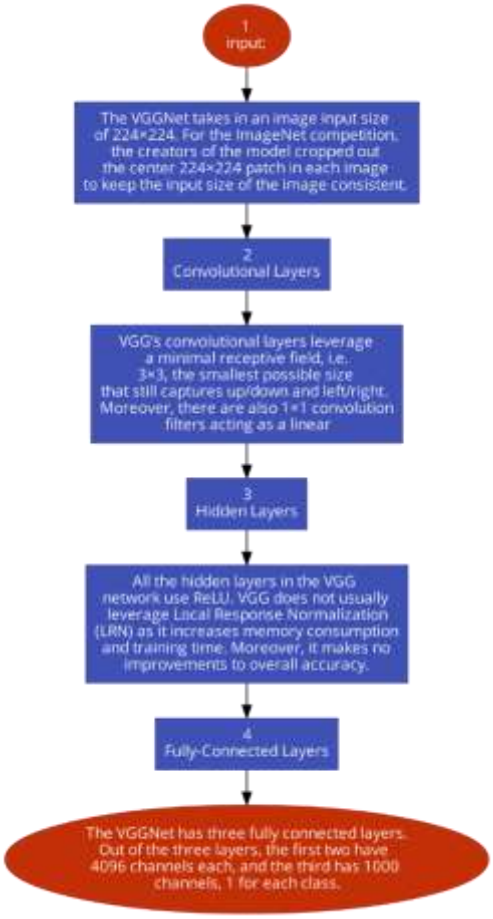


Fig:4 Design steps for VGG

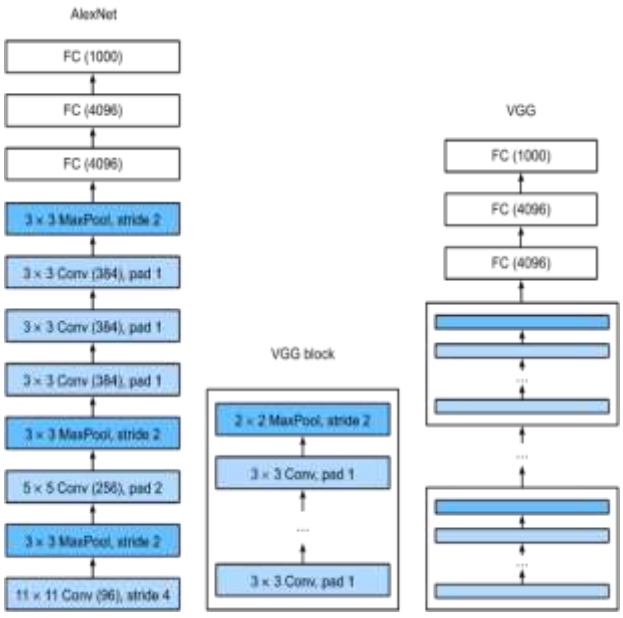


Fig:5 VGG Network

Five convolutional blocks made up the original VGG network; the first two of these blocks each have one convolutional layer, while the latter three each have two. The first block has 64 output channels, and the number of output channels increases by a factor of two with each succeeding block until there are 512 output channels altogether. This network is known as VGG-11 since it has three completely connected layers and eight convolutional layers.

Complexity and challenges

Five convolutional blocks made up the "initial VGG network, the latter three of which each had two convolutional layers and the first two of which each had one. The first block has 64 output channels, and the number of output channels increases by a factor of two with each succeeding block until there are 512 output channels altogether. This network is known as VGG-11 since it has three completely connected layers and eight convolutional layers.

Experimental Results And Discussions

As a starting point, this section presents an architecture that consistently gives me with reasonable outcomes. There are four different sorts of X-ray images available in the model. For concluding comments, the 'confusion matrix, accuracy, precision, recall, and F-score' are examined. The Covid class has an accuracy rate of 89% and a recall rate of 0.895%. As a result, the model is unable to correctly categorise all Covid-19 samples. It is, nonetheless, usually correct to do so (high precision - low FP). Precision and Recall (both better than 88 percent) are similar in the Lung Opacity and Normal classes, indicating that the model is good at recognising and categorising these samples. Here, we have to consider training models resnet50 with 1600 samples and maximum of 20 epochs and validating with 200 samples from the corpus data set.

(i): Training model resnet50 with 1600 samples and max of 20 epochs, and validating with 200 samples

Epoch	train loss	val loss	time(sec)
1	1.10196	0.91863	26.36
2	0.77953	0.64788	26.35
3	0.62232	0.58128	26.3
4	0.57838	0.64373	26.44
5	0.52742	0.52919	26.33
6	0.48673	0.53723	26.3
7	0.47148	0.5592	26.48
8	0.45234	0.46929	26.39
9	0.4299	0.54057	26.5
10	0.40807	0.45246	26.43
11	0.39559	0.50645	26.34
12	0.39706	0.45655	26.58
13	0.37875	0.43944	26.5
14	0.36924	0.42109	26.32
15	0.35774	0.42202	26.35

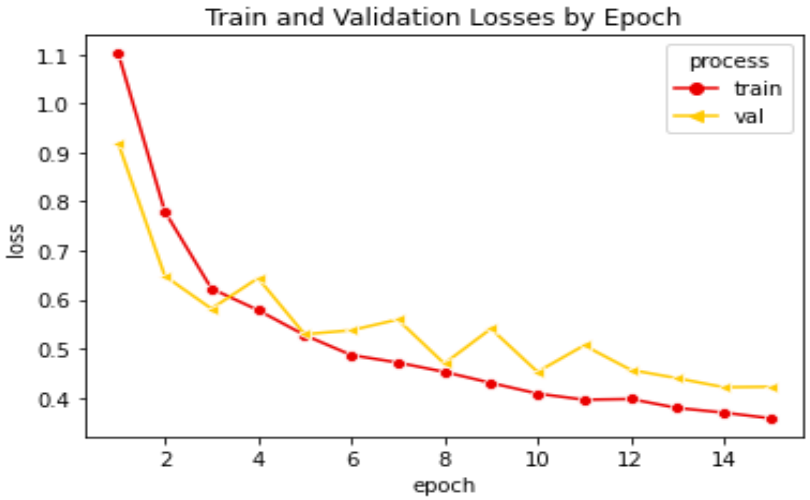


Fig:6 train and validation losses by epoch

In Fig(6) ,we have to consider training models resnet50 with 1600 samples and maximum of 20 epochs and validating with 200 samples from the corpus data set. The relation between loss and epoch determination as show in fig (6).

Training model vgg16 with 1600 samples and max of 20 epochs, and validating with 200 samples

epoch	train loss	val loss	time(Sec)
1	0.97049	0.77628	51.61
2	0.66439	0.6018	51.9
3	0.56597	0.54785	51.73
4	0.52116	0.55944	51.66
5	0.48218	0.49195	51.49
6	0.47613	0.47877	51.81
7	0.47394	0.46751	51.8
8	0.44823	0.44091	51.79
9	0.43496	0.43908	51.78
10	0.42257	0.43814	51.61
11	0.40768	0.42089	51.75
12	0.39885	0.43289	51.76
13	0.39214	0.42201	51.63
14	0.39148	0.39706	51.6
15	0.38929	0.41296	51.72

Here, we have to consider the Training model vgg16 with 1600 samples and max of 20 epochs, and validating with 200 samples and determines each epochs train loss and validation loss with respective time.'In a nutshell, the purpose of our primary study, which was pre-planned, was to assess the influence of baseline parameters of interest on the severity and recovery of COVID-19 in lung cancer patients. We looked at variables that may directly affect the severity and recovery of COVID-19 infections using a survey of the literature on COVID-19 publications and directed acyclic graphs to uncover correlations between baseline features of interest.

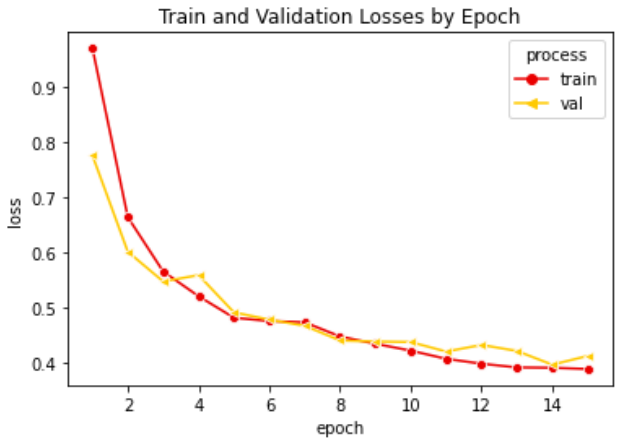


Fig:7Train and validation losses by epoch

Here, we have to consider the Training model vgg16 with 1600 samples and max of 20 epochs, and validating with 200 samples and determines each epochs train loss and validation loss with respective time.the relation between train and validation losses by epoch values shown in fig (7)

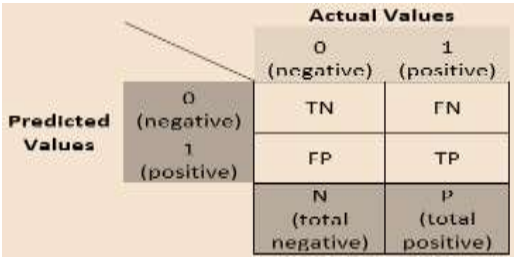
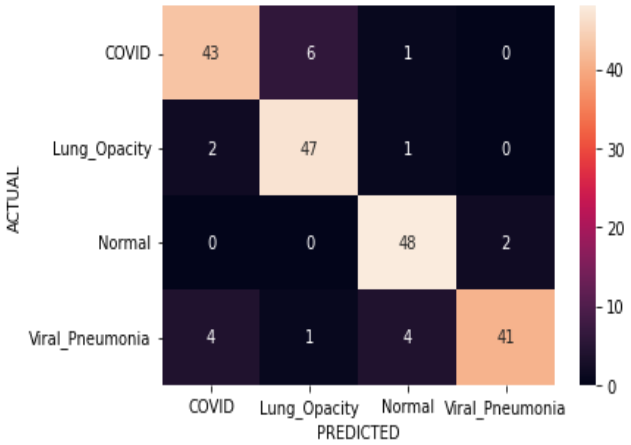
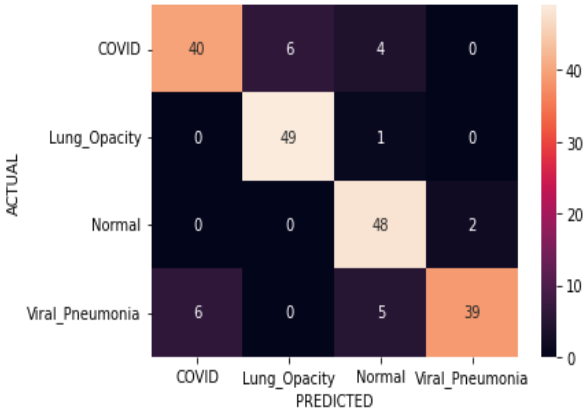


Fig 8: predicted and actual values



DiDeConResNet Model



DiDeConvVGG Model

"Overall, the model can recognise the samples, indicating that there is adequate TP," reads the table below. Samples of Normal or Lung Opacity can be anticipated if Covid-19 is incorrectly categorised. It is doubtful that viral pneumonitis is to blame. "Viral Pneumonia or Lung Opacity are misdiagnosed more frequently than Lung Opacity." It is common to mistake viral pneumonitis or lung opacity for normal samples. Covid-19 is less likely to be mistaken for it. Viral pneumonia is the category with the fewest misclassifications.

Model	Accuracy	Precision	Recall	F1-score
DiDeConResNet	0.895	0.897575	0.895	0.894333
DiDeConvVGG	0.88	0.88482	0.88	0.878175

the Covid class has an accuracy of 89% and a recall of 89%. The result indicates that the model is correctly categorise all Covid-19 samples. When it does, though, it is typically correct (high precision - low FP). Precision and Recall (above 80%) are similar in the Lung Opacity and Normal classes, indicating that the model is good at recognising and classifying these samples. In Covid-19, the Normal and Viral Pneumonia classes have the opposite outcome.

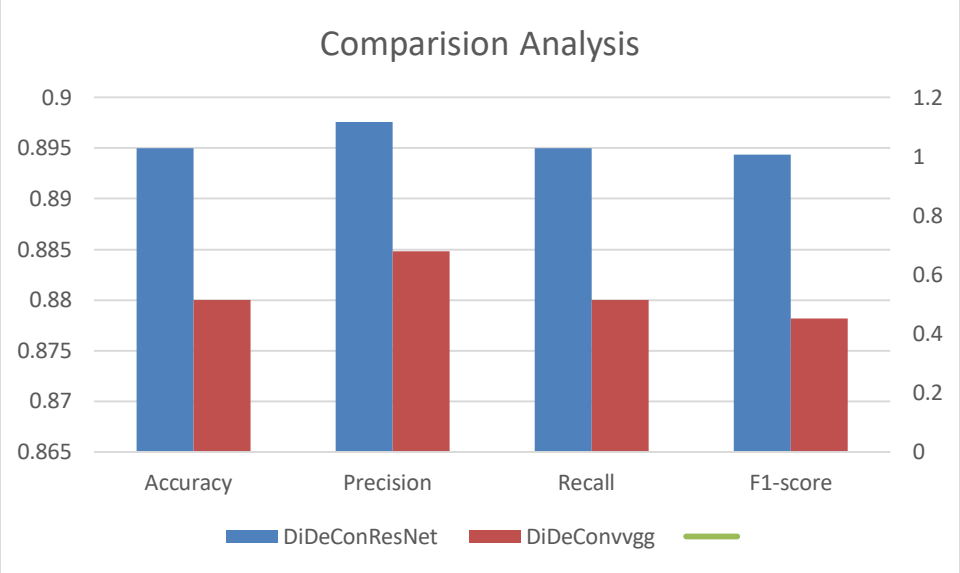


Fig: 9 Comparison Analysis

They have a recall that is higher than precision in Fig. 9. This suggests that the model can identify these samples, as seen by the drop in FN. The Normal class is frequently confused with viral pneumonia or lung opacity, as indicated in the Confusion Matrix, but it also results in FP. Precision and Recall are combined to create the F-Score. The Normal class scored highly since the Precision and Recall requirements are similar. For Covid, the Precision and Recall ratings deviated more sharply, leading to worse outcomes. Overall, it's a good sign if all of the general indications are above 70%.The quantity of COVID samples incorrectly labelled as NORMAL samples should be reduced, and the results per class should be improved.

In Fig (8) An evaluation tool for machine learning systems, frequently those that use supervised learning, is a confusion matrix, which is a matrix (table). The confusion matrix shows the instances of an actual class in each row, while the instances of a predicted class are shown in each column. The order can alternatively be reversed, with anticipated classes in rows and actual classes in columns, as is the case in this section of our course. The name confusion matrix alludes to the fact that it enables us to immediately identify the many kinds of flaws in our classification systems.

Conclusions And Future Scope

Healthcare professionals may detect COVID-19 with little to no processing of chest X-ray pictures thanks to deep learning algorithms. Over time, more articles on using deep learning to diagnose lung illness have been released. However, a thorough analysis of the condition of research and application was not conducted. In order to give a thorough study for the detection of lung diseases by deep learning, this document was created and will be available from September 2016 to September 2020. It will cover COVID-19, lung cancer, pneumonia, and tuberculosis. X-ray images of the lungs are used in current studies to detect Covid-19 influence lung cancer using machine learning and deep learning algorithms, with the essential methods addressed with less likely outcomes.

For academics, scientists, and healthcare workers, this is a difficult scenario. In their respective domains, researchers are continuously looking for potential answers to deal with this pandemic. "In order to create a forecast and evaluate the efficacy of the solution, we need to train each data set separately." The suggested techniques have served as a strong warning to society throughout the nation. Therefore, the existence of such a notion for novel study will be taken into consideration for earlier public advise and will greatly lower the rate of lung cancer fatalities influenced by Covid-19. For this application, precision and recall are key factors, hence the results are evaluated in terms of F1-Score.

To offer a thorough study of the Model performance, the confusion matrix and results for the F1-Score, Precision, Recall, and total Accuracy are also presented. Other difficulties were addressed and explained, such as the route deep learning and transfer learning would take in the future to screen for lung diseases using high-end

cloud data servers. The taxonomy could be used by other researchers to design their own studies and tasks. The suggested future course has the ability to increase productivity while also expanding the number of applications for comprehensive lung disease detection.

References

Kushagra, Rajneesh Kumar, Shaveta Jain. "Recent Advances in Machine Learning for Diagnosis of Lung Disease: A Broad View" , 2022 IEEE Delhi Section Conference (DELCON), 2022

"Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries" , Springer Science and Business Media LLC, 2019

"Big Data Analytics and Artificial Intelligence Against COVID-19: Innovation Vision and Approach" , Springer Science and Business Media LLC, 2020

World Health Organization. Coronavirus Disease 2019 (COVID-19) Situation Report; Technical Report March; World Health Organization: Geneva, Switzerland, 2020.

Rahaman, M.M.; Li, C.; Yao, Y.; Kulwa, F.; Rahman, M.A.; Wang, Q.; Qi, S.; Kong, F.; Zhu, X.; Zhao, X. Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. J. X-Ray Sci. Technol. 2020, 28, 821–839. [CrossRef]

HaitzSáez de OcárizBorde, David Sondak, PavlosProtopapas. "Convolutional neural network models and interpretability for the anisotropic reynolds stress tensor in turbulent one-dimensional flows" , Journal of Turbulence, 2021

Kangle Zhu, Lian Wang, Zanhao Chen, Jianing Ding, Jixuan Dong, Jinliang Chen. "Analysis and construction of an exosome derived competing endogenous RNA network for small cell lung cancer in the exoRbase database" , Research Square Platform LLC, 2022

Gordienko, Y.; Gang, P.; Hui, J.; Zeng, W.; Kochura, Y.; Alienin, O.; Rokovyi, O.; Stirenko, S. Deep Learning with Lung Segmentation and Bone Shadow Exclusion Techniques for Chest X-Ray Analysis of Lung Cancer. Adv. Intell. Syst. Comput. 2019, 638–647. [CrossRef]

Ankita Shelke, Madhura Inamdar, Vruddhi Shah, Amanshu Tiwari, Aafiya Hussain, Talha Chafekar, NinadMehendale. "Chest X-ray classification using Deep learning for automated COVID-19 screening" , Cold Spring Harbor Laboratory, 2020

Kieu, S.T.H.; Hijazi, M.H.A.; Bade, A.; Yaakob, R.; Jeffree, S. Ensemble deep learning for tuberculosis detection using chest X-Ray and canny edge detected images. IAES Int. J. Artif. Intell. 2019, 8, 429–435. [CrossRef]

Salman, F.M.; Abu-naser, S.S.; Alajrami, E.; Abu-nasser, B.S.; Ashqar, B.A.M. COVID-19 Detection using Artificial Intelligence. Int. J. Acad. Eng. Res. 2020, 4, 18–25.

Gao, X.W.; James-reynolds, C.; Currie, E. Analysis of tuberculosis severity levels from CT pulmonary images based on enhanced residual deep learning architecture. Neurocomputing 2019, 392, 233–244. [CrossRef]

O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep Learning vs . Traditional Computer Vision. Adv. Intell. Syst. Comput. 2020, 128–144. [CrossRef]

Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop, Swinoujscie, Poland, 9–12 May 2018; pp. 117–122. [CrossRef]

Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. J. Big Data 2019,6. [CrossRef]

Ker, J.; Wang, L. Deep Learning Applications in Medical Image Analysis. IEEE Access 2018, 6, 9375–9389. [CrossRef]

Wang, C.; Chen, D.; Hao, L.; Liu, X.; Zeng, Y.; Chen, J.; Zhang, G. Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model. IEEE Access 2019, 7, 146533–146541. [CrossRef]

Kabari, L.G.; Onwuka, U. Comparison of Bagging and Voting Ensemble Machine Learning Algorithm as a Classifier. Int. J. Adv. Res. Comput. Sci. Softw. Eng. 2019, 9, 1–6.

M.E.H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M.A. Kadir, Z.B. Mahbub, K.R. Islam, M.S. Khan, A. Iqbal, N. Al-Emadi, M.B.I. Reaz, M. T. Islam, "Can AI help in screening Viral and COVID-19 pneumonia?" IEEE Access, Vol. 8, 2020, pp. 132665 - 132676..

eISSN1303-5150