Democratic and Popular Republic of Algeria

Ministry of Higher Education and Scientific Research

University of Kasdi Merbah, Ouargla

Faculty of new information and communication technologies

Department of computer science and Information technology

Thesis Submitted to the Department of computer science and Information in Candidacy for the Degree of Master in computer science, option “fundamental computing”.

Presented by:

FASSOULI FAYCAL

COVID-19 SCREENING BASED ON SUPERVIED DEEP LEARNING

Committee members

Dr. AAZAOUI N	President
Dr. Belhadj M	Examiner
Dr. AIADI Oussama	Supervisor

Academic Year 2021/2022

Acknowledgments

First and foremost, praise and thanks to ALLAH the sacred and the mighty for giving me the ability and helping me in accomplishing this thesis.

Then, I would like to express my sincerest gratitude to my supervisor Dr. AIADI Oussama for his continuous support, help, guidance, motivation and patience throughout this thesis.

A very special thanks go to the honorable members of my dissertation committee jury who accepted to examine and evaluate this thesis namely, Dr. Belhadj M and Dr. Aazaoui N.

I gratefully acknowledge all the professors within the Department of computer science and information technology.

And lastly, I want to express my gratitude to my family members. my mother, Samia. My father, Fassouli Abderrahmane (رحمه الله و غفرله) and my brother and sister. Also, my aunt who supported me throughout the year.

Abstract

CT-scans images are helpful for detecting COVID-19. In this thesis, we are interested by investigating the performance of deep networks with CT-scan images and with detecting COVID-19. We used Convolution deep networks and we will provide insights on the performance of this deep networks. also, compare the performance of this networks and using ensemble learning we try to combine different models. The proposed approach is three-fold and comprises three stages which are training, training with data augmentation then ensemble learning. The first stage is done to train our models with CT-scan images and see their performance. While the second stage is dedicated to see how models will perform using data augmentation technique on CT-scan images. The last stage is using ensemble learning and combine the results of models to increase the accuracy. We conduct experiment on COVID-19-CT dataset. With accuracy as performance metric. Experimental results reveal that model without data augmentation are more performing than models with it. VGG16 model is better performing than ResNet50.

KEYWORDS: COVID-19, Deep Learning, Neural Network, Data Augmentation, Ensemble Learning.

ملخص

تعد صور التصوير المقطعي المحوسب مفيدة في اكشاف COVID-19. في هذه الأطروحة ، نحن مهتمون بالتحقيق في أداء الشبكات العميقة باستخدام صور الأشعة المقطعية و الكشف عن COVID-19. استخدمنا شبكات العميقة و سنقدم رؤى حول أداء هذه الشبكات العميقة. أيضًا ، نقارن أداء هذه الشبكات وباستخدام التعلم الجماعي نحاول الجمع بين نماذج مختلفة. النهج المقترح ثلاثي الأبعاد ويتألف من ثلاث مراحل وهي التدريب ، والتدريب مع زيادة البيانات ، ثم التعلم الجماعي. يتم الانتهاء من المرحلة الأولى لتدريب نماذجنا على صور الأشعة المقطعية والاطلاع على أدائها. بينما تم تخصيص المرحلة الثانية لمعرفة كيفية أداء النماذج باستخدام تقنية زيادة البيانات على صور التصوير المقطعي المحوسب. المرحلة الأخيرة هي استخدام التعلم الجماعي والجمع بين نتائج النماذج لزيادة الدقة. نجري تجربة على مجموعة بيانات COVID-19-CT. مع الدقة كمقياس للأداء. تكشف النتائج التجريبية أن النموذج بدون زيادة البيانات يكون أكثر أداءً من النماذج الموجودة به. نموذج VGG16 هو أفضل أداء من ResNet50.

كلمات مفتاحية: كوفيد-19، التعلم العميق، الشبكة العصبية، زيادة البيانات، التعلم الجماعي.

Contents

List of figures

List of tables

CHAPTER1: GENERAL INTRODUCTION

1- Introduction

Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data.

Techniques based on machine learning have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical and medical applications.

Computer vision is a field which aims to create methods which allows the machine to "recognize", meaning to extract meaningful information in digital image or video. Because the shapes represented in digital matter can be seen from any angle, this problem equates to attempting to replicate the mechanism of human vision, and the road, like any study emulating the human brain, is likely to be excessively long.

Humans have an advantage. Human vision benefits from lifetimes of context to teach it how to distinguish objects apart, how far away they are, whether they are moving, and whether something is incorrect with an image.

Computer vision teaches computers to execute similar tasks, but using cameras, data, and algorithms rather than retinas, optic nerves, and a visual cortex, it must do it in a fraction of the time.

Computer vision often appear in our daily lives, such as phones since you can unlock your phone using facial recognition or fingerprint recognition. In addition, the medical field such as radiographic diagnostics, X-Ray analysis and CT and MRI analysis, Cancer detection. And many other applications that simplify various tasks e.g., Pedestrian detection, Self-driving cars, Reading text and barcodes …. etc.

Computer vision is also commonly used to analyze CT and MRI data. It seems to be the key to improving patient outcomes, from building AI systems to evaluate radiological images with the same levels of accuracy as human doctors (while minimizing disease detection time) to deep learning algorithms to boost the resolution of MRI scans.

2- Problematic

In this thesis, we will try handling the research question concerned with determining for a single CT-Scan image the probability of having COVID-19. Noting Coronavirus disease (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. The virus can spread from an infected person's mouth or nose in small liquid particles when they cough, sneeze, speak, sing or breathe.

A CT scan contains hundreds of slides, and using such scans to diagnose COVID-19 can cause delays in hospitals. Artificial intelligence algorithms could help radiologists diagnose COVID-19 infection in these images more quickly and effectively.

3- Motivations

The "wisdom of the crowd" principle states that a large number of people with average expertise on a subject can provide credible answers to issues like quantity prediction, spatial reasoning, and general knowledge. The combined results balance out the noise and, in many cases, outperform those of highly qualified specialists. The same rule can be used to artificial intelligence systems that use machine learning, a type of AI that uses mathematical models to anticipate outcomes.

Ensemble learning is used in machine learning to achieve crowd wisdom. The outcome of an ensemble, which is a collection of machine learning models, can be more accurate than any one member of the group for many tasks.

Recently convolutional networks have enjoyed a great success in large scale image recognition. In particular, an important role in the advance of deep visual recognition architectures has been played by the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al., 2014).[9].

This work is motivated by the above reasons, where we will try to make a multi modal of deep convolutional neural networks, and ensemble their outcome.

4- Contributions

The contributions of this thesis can be summarized as follows:

- Investigation of performance of different pre-trained deep networks namely VGG16 and ResNet50 in COVID-19 detection from CT-scan images.
- Investigation of the effect of data augmentation on the performance of those networks.
- Proposing a multi-modal-based ensemble learning scheme based on different deep networks, where the wavelet normalization technique is used to generate the other modality from the original dataset images.

5- Thesis structure

This thesis is organized as follow:

- First chapter state the general background of the work involving general scheme of recognition systems, classification methods …etc.
- Second chapter is dedicated to introduce different concepts and methods used in this work.
- Third chapter describe the process of work we conducted and the steps we took to finalize this work.
- At last, a general conclusion is given and results of the different methods are given.

CHAPTER2: BACKGROUND

1- Introduction

In the past few years there were huge advancement in technology and computers capabilities reached a new level. Which allowed scientist to try what they could not before, especially in artificial intelligence which demands huge computing powers.

In this chapter, we will present the main background of our work, including details about image processing, computer vision applications. Also, introduce neural network and its components and an overview on deep learning.

2- What is Image

By referring to Gonzelez and Woods book [1], we can quote the following definition for the digital image:

"An image may be defined as a two-dimensional function, $f(x, y)$, where x and y are spatial (plane) coordinates, and the amplitude of f at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y and the amplitude values of f are all finite, discrete quantities, we call the image a digital image".

Common image file formats online include:

- JPEG (pronounced JAY-peg) is a graphic image file produced according to a standard from the Joint Photographic Experts Group, an ISO /IEC group of experts that develops and maintains standards for a suite of compression algorithms for computer image files. JPEGs usually have a .jpg file extension.
- GIF (pronounced JIF by many, including its designer; pronounced GIF with a hard G by many others) stands for Graphics Interchange Format. The GIF uses the 2D raster data type and is encoded in binary. GIF files ordinarily have the .gif extension.

- GIF89a is an animated GIF image, formatted according to GIF Version 89a. One of the chief advantage formats is the ability to create an animated image that can be played after transmitting to a viewer page that moves - for example, a twirling icon or a banner with a hand that waves or letters that magically get larger. A GIF89a can also be specified for interlaced GIF presentation.
- PNG (pronounced *ping*) is a Portable Network Graphics) is a file format for image compression that was designed to provide a number of improvements over the GIF format. Like a GIF, a PNG file is compressed in lossless fashion (meaning all image information is restored when the file is decompressed during viewing). Files typically have a .png extension.
- SVG is Scalable Vector Graphics, the description of an image as an application of XML. Any program such as a browser that recognizes XML can display the image using the information provided in the SVG format. Scalability means that the file can be viewed on a computer display of any size and resolution, whether the small screen of a smartphone or a large widescreen display in a PC. Files usually have .svg extension.

3- Image processing

Image processing is performing operations to an image in order to improve it or extract relevant information from it. It's a sort of signal processing in which the input is an image and the output is either that image or its characteristics/features. Image processing is one of the most quickly evolving technology today. It is also a critical research field in engineering and computer science.

Image processing basically includes the following three steps:

- Importing the image via image acquisition tools.
- Analyzing and manipulating the image.
- Output in which result can be altered image or report that is based on image analysis.

We can manipulate and analyze images by using programmable algorithms on computer, such as enhancing image quality (e.g., increase image contrast), detecting edges (figure1), removing noise…etc. also we can convert image to set of numerical values that can further be used for other tasks e.g., image classification.

Figure 1 Edge detection

4- Computer vision

Computer vision is the science of acquiring, processing, and analyzing digital images. Computer vision aims to automatically and ideally imitate the human visual system in an accurate and efficient manner. The following are the three levels of computer vision:

- **Low-level vision:** process image for feature extraction (edge, corner, or optical flow). see figure (1).
- **Middle-level vision:** object recognition, motion analysis, and 3D reconstruction using features obtained from the low-level vision.
- **High-level vision:** interpretation of the evolving information provided by the middle level vision as well as directing what middle and low-level vision tasks should be performed. Interpretation may include conceptual description of a scene like activity, intention and behavior. See figure (2).

Figure 2 image captioned as 'a child is sitting on the ground and holding an umbrella'

5- Computer vision applications

Computer vision is one of the areas in Machine Learning where core concepts are already being integrated into major products that we use every day.

- **Self-driving cars:** Computer vision enables self-driving cars to make sense of their surroundings. Cameras capture video from different angles around the car and feed it to computer vision software, which then processes the images in real-time to find the extremities of roads, read traffic signs, detect other cars, objects and pedestrians. The self-driving car can then steer its way on streets and highways, avoid hitting obstacles, and (hopefully) safely drive its passengers to their destination. Figure (3).

Figure 3 the next generations cars

- **Facial recognition:** Computer vision is also important in facial recognition applications, which allow computers to match images of people's faces to their identities. Computer vision algorithms detect facial features in images and compare them to face profile databases. Consumer devices use facial recognition to verify their owners' identities. Facial recognition is used in social media apps to detect and tag users. Facial recognition technology is also used by law enforcement to identify criminals in video feeds.

Figure 4 face recognition

- **Augmented reality & Mixed reality:** Computer vision is also important in augmented and mixed reality, which allows computing devices such as smartphones, tablets, and smart glasses to overlay and embed virtual objects on real-world imagery. AR gear detects objects in the real-world using computer vision to determine where to place a virtual object on a device's display. Computer vision algorithms, for example, can assist AR applications in detecting planes such as tabletops, walls, and floors, which is critical in establishing depth and dimensions and placing virtual objects in the physical world.

Figure 5 use of augmented reality and mixed reality

- **Healthcare:** Computer vision has also played an important role in medical technology advancements. Computer vision algorithms can aid in the automation of tasks such as detecting cancerous moles in skin images and detecting symptoms in x-ray and MRI scans.

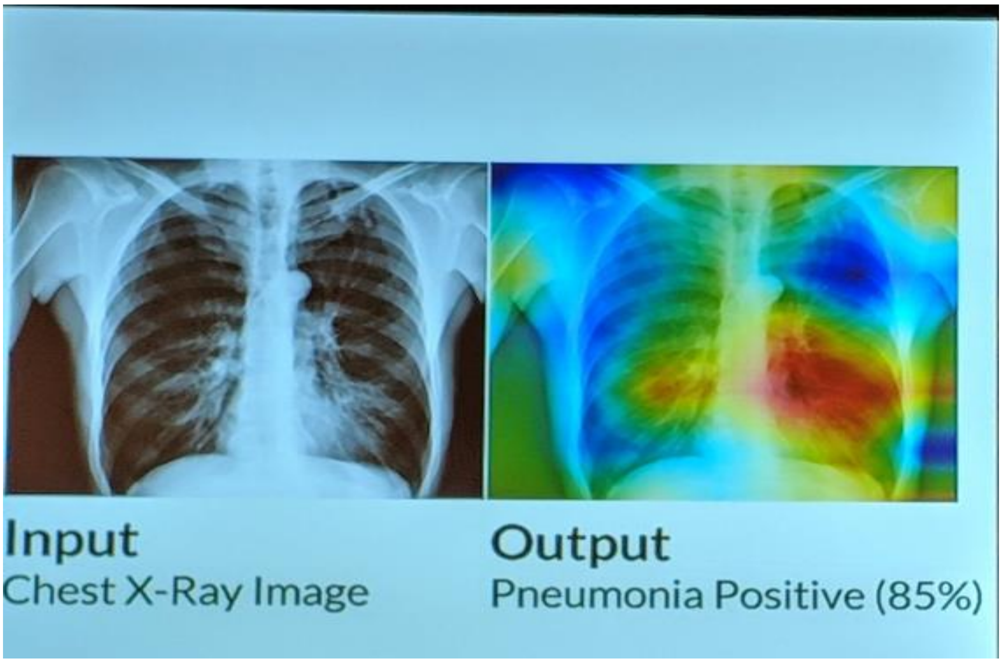

Figure 6 detecting disease in X-ray image

Another examples of computer vision applications:

- Robotic
- Medicine
- Industrial automation
- Action recognition
- …etc.,

6- Neural Network

A neural network is an artificial intelligence method for teaching computers to process data in a manner inspired by the human brain. Deep learning is a type of machine learning process that employs interconnected nodes or neurons in a layered structure that resembles the human brain. It generates an adaptive system that computers can use to learn from their mistakes and continuously improve. Thus, artificial neural networks attempt to solve complex problems with greater accuracy, such as summarizing documents or recognizing faces.

The human brain is the source of inspiration for neural network architecture. Neurons in the human brain form a complex, highly interconnected network and send electrical signals to one another to help humans process information. An artificial neural network, on the other hand, is made up of artificial neurons that collaborate to solve a problem. Artificial neurons are software modules known as nodes, and artificial neural networks are software programs or algorithms that solve mathematical calculations using computing systems.

Neural networks can assist computers in making intelligent decisions with minimal human intervention. This is because they can learn and model nonlinear and complex relationships between input and output data.

We give some examples where NN are used:

- Medical diagnosis by medical image classification.
- Targeted marketing by social network filtering.
- Financial predictions by processing historical data of financial instruments.

- Electrical load and energy demand forecasting
- Process and quality control.
- Chemical compound identification.

Four of the most important applications of neural networks:

- **Computer vision:** The ability of computers to extract information and insights from images and videos is known as computer vision. Computers can distinguish and recognize images in the same way that humans do.

- **Speech recognition:** Despite differences in speech patterns, pitch, tone, language, and accent, neural networks can analyze human speech. Speech recognition is used by virtual assistants such as Amazon Alexa and automatic transcription software to perform tasks such as these:
    - Assist call center agents and automatically classify calls.
    - Convert clinical conversations into documentation in real time.
    - Accurately subtitle videos and meeting recordings for wider content reach.

- **Natural language processing:** The ability to process natural, human-created text is referred to as natural language processing (NLP). Neural networks assist computers in extracting meaning and insights from text data and documents. NLP has several applications, including the following:
    - Automated virtual agents and chatbots.
    - Automatic organization and classification of written data.
    - Business intelligence analysis of long-form documents like emails and forms.
    - Indexing of key phrases that indicates sentiment, like positive and negative comments on social media.
    - Document summarization and article generation for a given topic

- **Recommendation engines:** To develop personalized recommendations, neural networks can track user activity. They can also analyze all user behavior to discover new products or services that may be of interest to a particular user. Curalate, a

Philadelphia-based startup, for example, assists brands in converting social media posts into sales. Curalate's intelligent product tagging (IPT) service is used by brands to automate the collection and curation of user-generated social content. IPT employs neural networks to find and recommend products based on the user's social media activity. Consumers no longer need to sift through online catalogs to find a specific product based on a social media image. Instead, they can use Curalate's auto product tagging to quickly purchase the product.

7- Architecture of neural network

Neural network architecture:

- **Simple neural network:** A basic neural network has interconnected artificial neurons in three layers:
    - **Input layer**: The input layer is where information from the outside world enters the artificial neural network. Input nodes process data, analyze or categorize it, and then forward it to the next layer.
    - **Hidden layer:** The input layer or other hidden layers provide input to hidden layers. A large number of hidden layers can exist in artificial neural networks. Each hidden layer analyzes the previous layer's output, processes it further, and passes it on to the next layer.
    - **Output layer**: The output layer displays the final result of the artificial neural network's data processing. It can have one or more nodes. For example, if we have a binary (yes/no) classification problem, the output layer will have one output node that returns a value of 1 or 0. In the case of a multi-class classification problem, the output layer may include more than one output node.

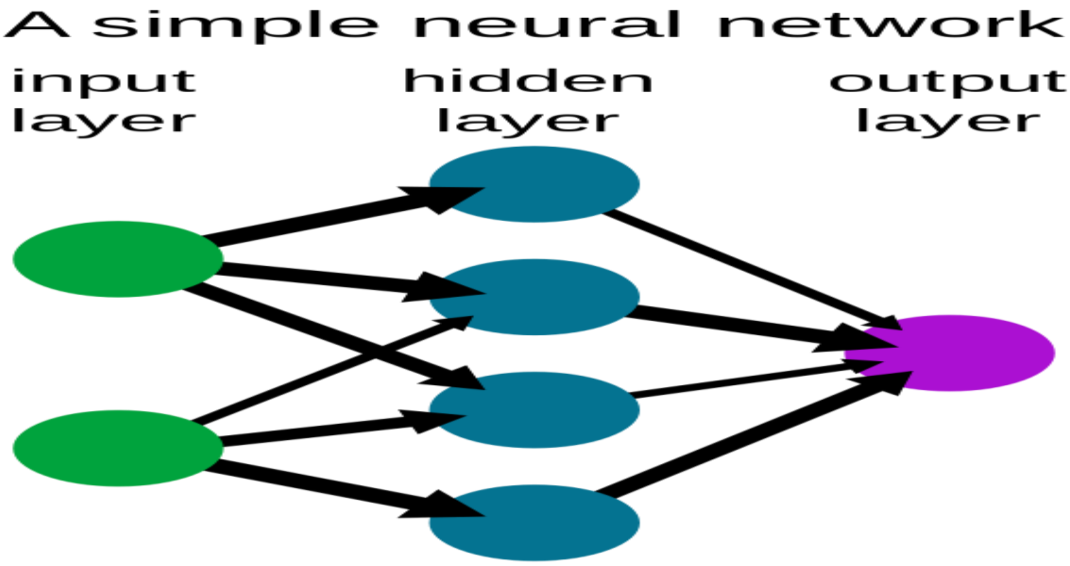

Figure 7 simple neural network

- **Deep neural network architecture:** Deep neural networks, also known as deep learning networks, are made up of several hidden layers that are linked together by millions of artificial neurons. The connections between nodes are represented by a number called weight. If one node excites another, the weight is positive; if one node suppresses another, the weight is negative. Nodes with higher weight values have a greater influence on other nodes.

  Deep neural networks can theoretically map any input type to any output type. However, they require significantly more training than other machine learning methods. They require millions of examples of training data rather than the hundreds or thousands required by a simpler network.

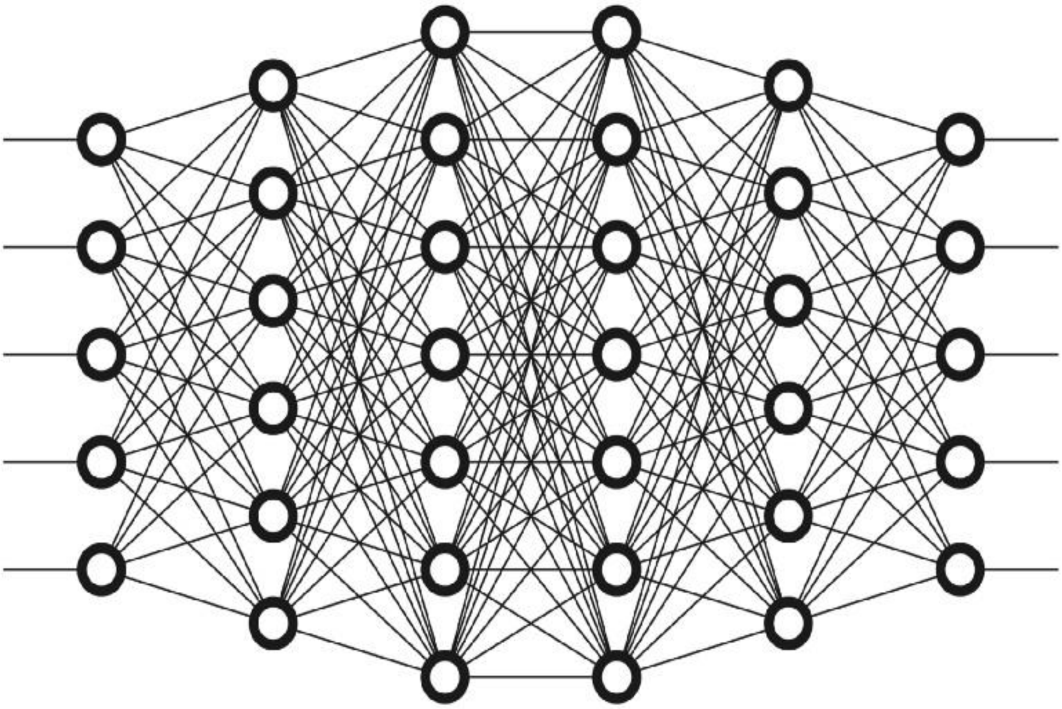

Figure 8 Deep neural network

8- Deep learning

Artificial intelligence is a branch of computer science that studies how to give machines the ability to perform tasks that require human intelligence. Machine learning is an artificial intelligence technique that allows computers to access massive datasets and train them to learn from them. Machine learning software discovers patterns in existing data and applies them to new data to make intelligent decisions. Deep learning is a subset of machine learning that processes data using deep learning networks.

Traditional machine learning methods necessitate human input in order for the machine learning software to function properly. A data scientist manually selects the set of relevant features to be analyzed by the software. This limits the software's capabilities, making it difficult to create and manage.

In deep learning, on the other hand, the data scientist only provides raw data to the software. The deep learning network generates the features on its own and learns more autonomously. It is capable of analyzing unstructured datasets such as text documents, determining which data attributes to prioritize, and solving more complex problems.

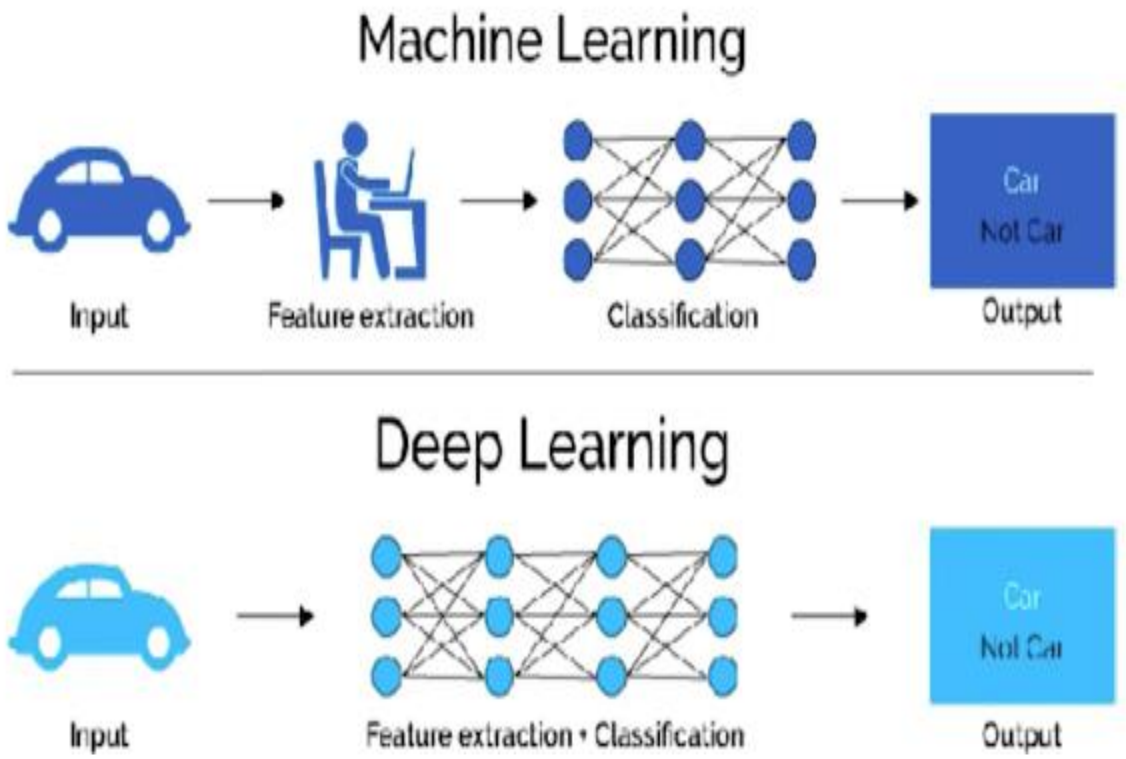

Figure 9 Machine learning vs Deep learning

9- Convolutional Neural Network (CNN)

The convolutional neural network (CNN) is a sort of neural network model created for working with two-dimensional picture data, while it can also be utilized with one-dimensional and three-dimensional data.

The higher performance of convolutional neural networks with picture, speech, or audio signal inputs sets them apart from other neural networks. They are divided into three sorts of layers:

- Convolutional layer
- Pooling layer
- Fully-connected (FC) layer

Some of the most well-known CNN models:

- **VGG16**: composed of 16 layers.
- **ResNet50**: composed of 50 layers.
- **LeNet**: composed of 7 layers.
- **GoogleNet**: composed of 22 Layers.

10- Architecture of Convolutional Neural Network

CNN is made up of numerous stacked layers, the first of which are convolution and pooling layers, followed by fully connected and soft-max layers. Each of these layers is described in detail in the sections that follow.

10.1 Convolution Layer

The convolutional layer, which gives the network its name, is at the heart of the convolutional neural network. This layer performs a process known as "convolution."

A convolution is a linear process in a convolutional neural network that involves the multiplication of a set of weights with the input, similar to a standard neural network. The multiplication is made between an array of input data and a two-dimensional array of weights, called a filter or a kernel. Figure (10).

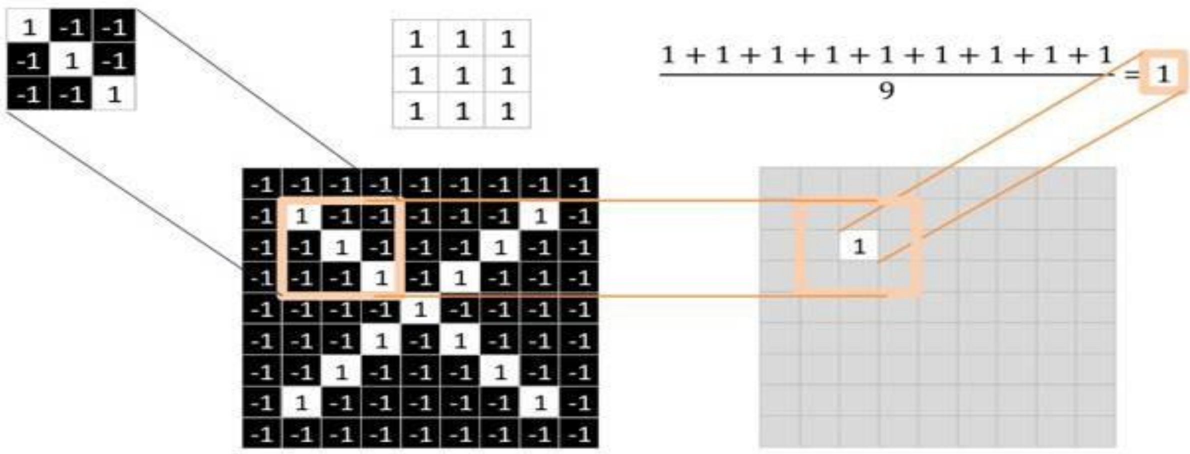

Figure 10 convolution operation

The filter is smaller than the input data and the type of multiplication applied between a filter-sized patch of the input and the filter is called a dot product. A dot product is the element-wise multiplication of the input and filter's filter-sized patch, which is then summed, always yielding a single value. The operation is often referred to as the "scalar product" because it produces a single value.

It is intentional to use a filter that is smaller than the input because it allows the same filter (set of weights) to be multiplied by the input array several times at different points on the input. From left to right, top to bottom, the filter is applied systematically to each overlapping section or filter-sized patch of the incoming data. Figure (11).

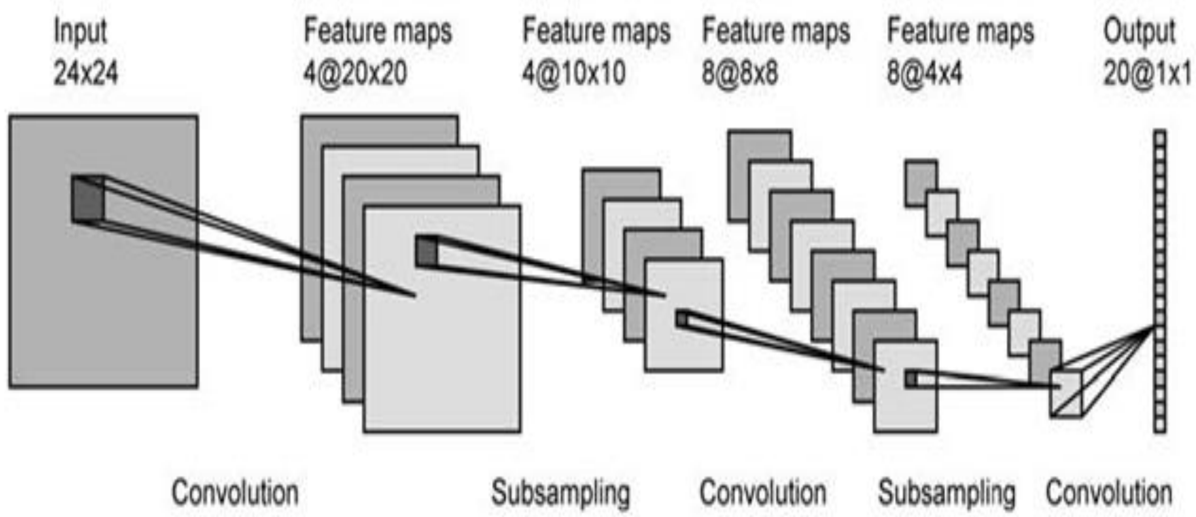

Figure 11 process of making feature map

The result of multiplying the filter once with the input array is a single value. As the filter is applied to the input array multiple times, the result is a two-dimensional array of output values representing input filtering. As a result, the two-dimensional output array of this operation is known as a "feature map.". figure (12).

This layer has certain parameters to adjust, which are:

- The number of filters affects the depth of the output. For example, three distinct filters would yield three different feature maps, creating a depth of three.
- Zero-padding is usually used when the filters do not fit the input image. This sets all elements that fall outside of the input matrix to zero, producing a larger or equally sized output. There are three types of padding:

- Valid padding: This is also known as no padding. In this case, the last convolution is dropped if dimensions do not align.
- Same padding: This padding ensures that the output layer has the same size as the input layer
- Full padding: This type of padding increases the size of the output by adding zeros to the border of the input

- Stride is the distance, or number of pixels, that the kernel moves over the input matrix. While stride values of two or greater is rare, a larger stride yields a smaller output.

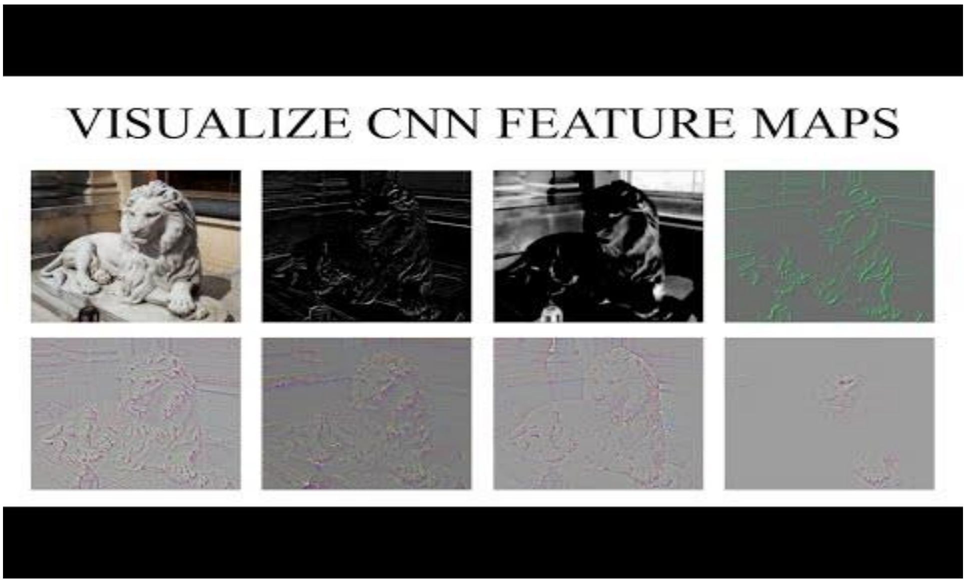

Figure 12 feature map

Once a feature map is created, we can pass each value in the feature map through a nonlinearity, such as a ReLU, much like we do for the outputs of a fully connected layer.

10.2Pooling Layer

Down sampling, also known as pooling layers, is a dimensionality reduction technique that reduces the number of factors in the input. The pooling process sweeps a filter across the entire input, similar to the convolutional layer, however this filter does not have any weights. Instead, the kernel uses an aggregation function to populate the output array from the values in the receptive field. Pooling can be divided into two categories:

- **Max pooling:** As the filter moves across the input, it selects the pixel with the maximum value to send to the output array. As an aside, this approach tends to be used more often compared to average pooling. Figure (13).

- **Average pooling:** As the filter moves across the input, it calculates the average value within the receptive field to send to the output array.

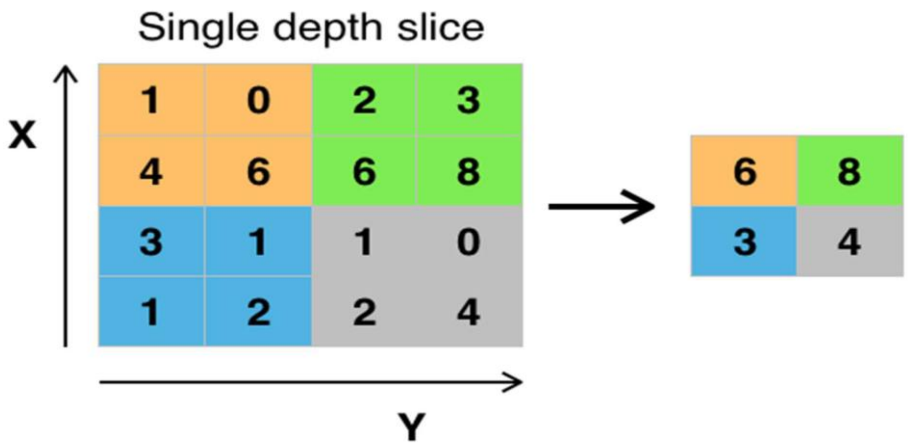

Figure 13 max pooling operation

While a lot of information is lost in the pooling layer, it also has a number of benefits to the CNN. They help to reduce complexity, improve efficiency, and limit risk of overfitting.

10.3 Rectified Linear Units layer (ReLu)

In this layer, a zero-based element-wise function (max(0, x)) is applied to threshold values, preserving positive values while converting negative values to zero. figure (14).

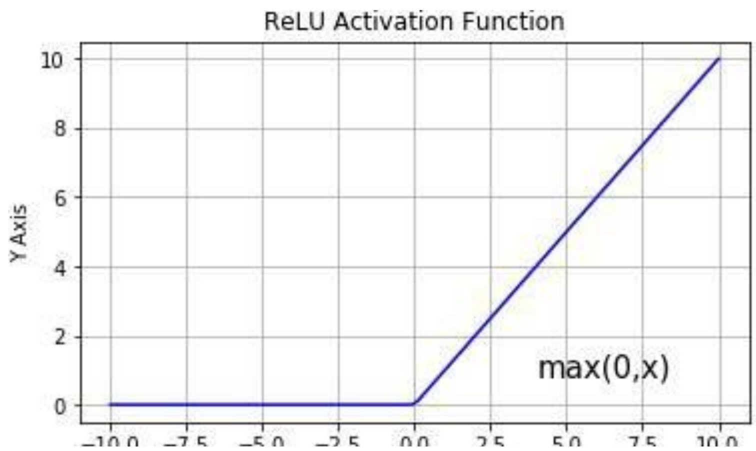

Figure 14 ReLU function

10.4 Fully-connected Layer (FC Layer)

This layer performs classification tasks based on the characteristics retrieved by the preceding layers and their various filters. While convolutional and pooling layers often utilize ReLu functions to categorize inputs, FC layers typically use a SoftMax activation function to provide a probability from 0 to 1. Figure (15).

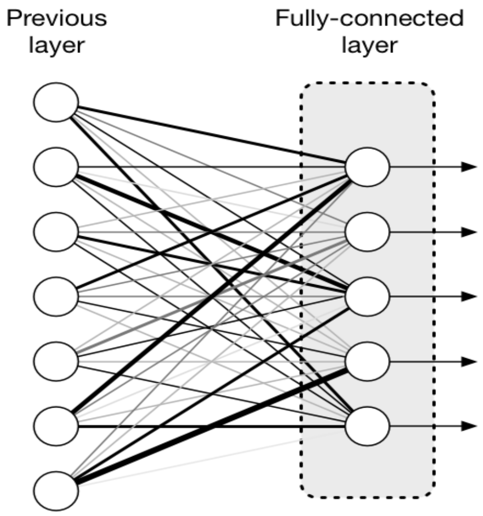

Figure 15 Fully Connected layer

The whole CNN architecture looks as follows. figure (16).

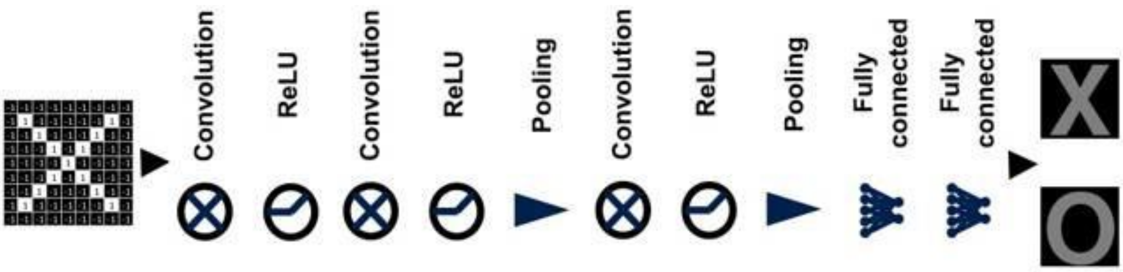

Figure 16 architecture of CNN

11- Conclusion

In this chapter, we have introduced some important notion, starting from image and image processing, passing by Neural network and its different architectures to computer vision and applications of computer vision, and finalizing by the convolutional neural network. We tried to provide some information to make it easy to understand the work.

CHAPTER3: MATERIAL & METHOD

1- Introduction

Convolution neural networks (CNN) has shown their powerful ability in image recognition and classification.

The purpose of this chapter is to provide an overview on our proposed method then we will explain each component separately.

2- Proposed Method

The proposed method can be divided into five phases which are training, data augmentation and ensemble learning. Then, manipulating the dataset, finally, ensemble learning of the new generated images. In the following we present each step.

2.1 Training

This step can be summarized as follow:

- Using pre-trained VGG16 model with "imagenet" weights and train it with the COVID-19-CT dataset images.
- Using pre-trained ResNet50 model with "imagenet" weights and train it with the COVID-19-CT dataset images.

2.2 Training with Data Augmentation

This step can summarize as follows:

- Using keras preprocessing module we perform four functions on our original images. See figure (17).
- Re-train our models with the new data.

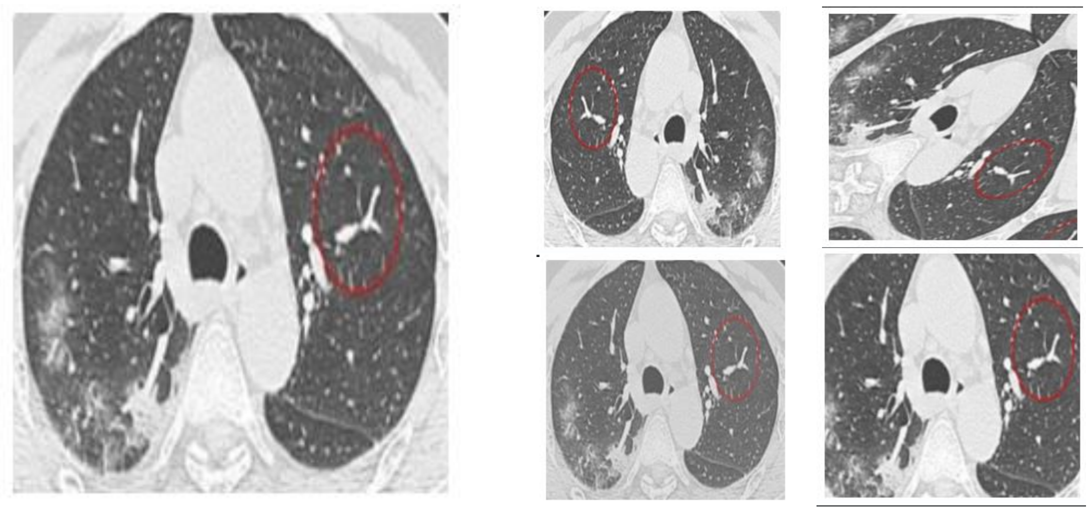

Figure 17 Data augmentation samples

2.3 Ensemble Learning

In this step we will do as follow:

- Ensemble technique uses the same training set for training the two pretrained models (VGG16 and ResNet50).
- Predicts the probabilities of the samples in the test set by the fine-tuned models to calculate the average probability score, thus giving equal weightage to all the three classifiers.

$$P^{(i)} = \left[P_1^{(i)} P_2^{(i)} \cdots P_n^{(i)}\right]$$

- The final prediction score using the ensemble technique is given by.

$$\begin{aligned} P^e &= \frac{\sum_{i=1}^{m} P^{(i)}}{m} \\ &= \left[\frac{\sum_{i=1}^{m} P_1^{(i)}}{m} \frac{\sum_{i=1}^{m} P_2^{(i)}}{m} \dots \frac{\sum_{i=1}^{m} P_n^{(i)}}{m}\right] \\ &= [P_1' P_2' \dots P_n'] \end{aligned}$$

2.4 Manipulating The Dataset

This step goes as follow:

- By using MATLAB Inface toolbox, we perform the wavelet denoising function on all our dataset images
- Save the new generated images in our hard drive. See figure (18).

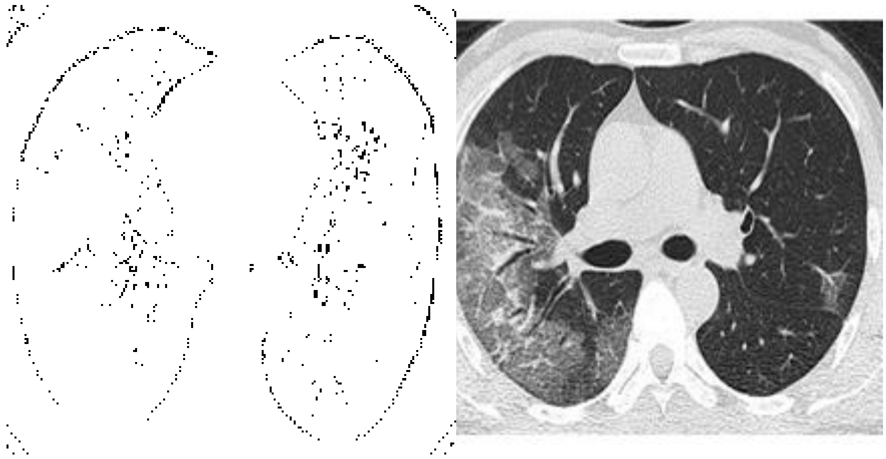

Figure 18Wavelet normalization generated images sample

2.5 Ensemble learning of the new generated images

And this goes as follow:

- Perform ensemble learning technique on the generated images.
- Predicts the probabilities of the samples in the test set by the fine-tuned models to calculate the average probability score, once giving equal weightage to the two classifiers. And once with giving different weightage.
- Calculate The final prediction score using the ensemble technique.

3- VGG16 Model

VGG stands for Visual Geometry Group, and it is a multilayer deep Convolutional Neural Network (CNN) architecture. The term "deep" refers to the number of layers in VGG-16 or VGG-19, which have 16 or 19 convolutional layers respectively, VGG16 architecture see figure (19).

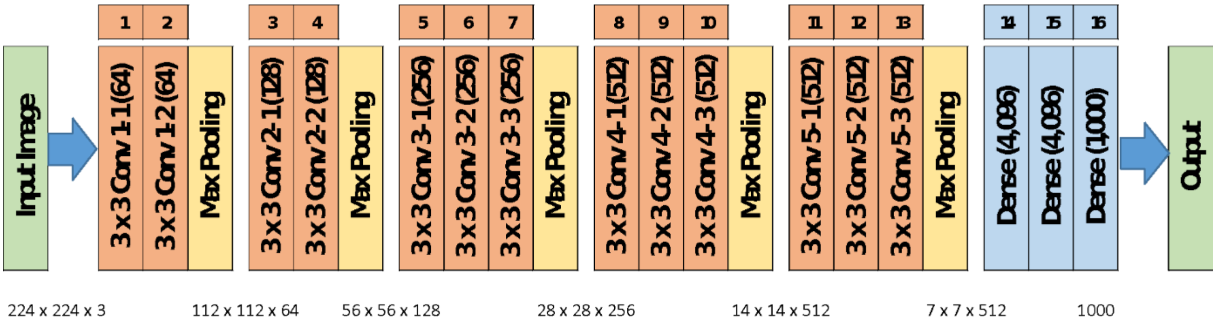

Figure 19VGG16 architecture

The VGG architecture serves as the foundation for cutting-edge object recognition models. The VGGNet, which was created as a deep neural network, outperforms baselines on a variety of tasks and datasets in addition to ImageNet. Furthermore, it is still one of the most widely used image recognition architectures today. Figure (20).

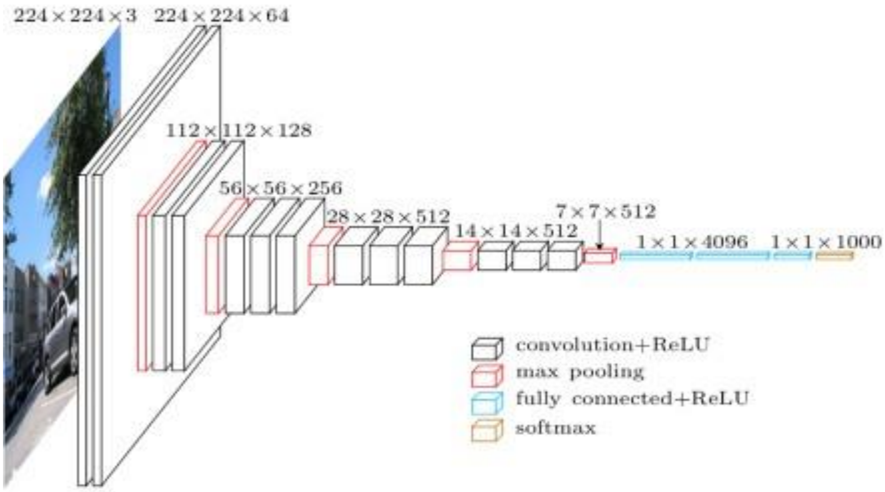

Figure 20VGG16 process explained

The VGG model, or VGGNet, that supports 16 layers is also referred to as VGG16, which is a convolutional neural network model proposed by A. Zisserman and K. Simonyan from the University of Oxford. Reference [2].
The VGG16 model achieves almost 92.7% top-5 test accuracy in ImageNet. ImageNet is a dataset consisting of more than 14 million images belonging to nearly 1000 classes.

Moreover, it is one of the most popular models. It replaces the large kernel-sized filters with several 3*3 kernel-sized filters one after the other.

As previously stated, the VGGNet-16 has 16 layers and can categorize photos into 1000 different item categories, such as keyboards, animals, pencils, and mice. In addition, the model features a 224-by-224 picture input size.

4- ResNet50 Model

ResNet-50 is a 50-layer deep convolutional neural network. ResNet, short for Residual Networks, is a well-known neural network that serves as the foundation for many computer vision applications. ResNet was a game-changer because it allowed us to train extraordinarily deep neural networks with 150+ layers.

It was first introduced in 2015. Reference (3).

This model was immensely successful, as can be ascertained from the fact that its ensemble won the top position at the ILSVRC 2015 classification competition with an error of only 3.57%.

Convolutional Neural Networks have a major disadvantage — 'Vanishing Gradient Problem'. During backpropagation, the value of gradient decreases significantly, thus hardly any change comes to weights. To overcome this, ResNet is used. It makes use of "SKIP CONNECTION". See figure (21).

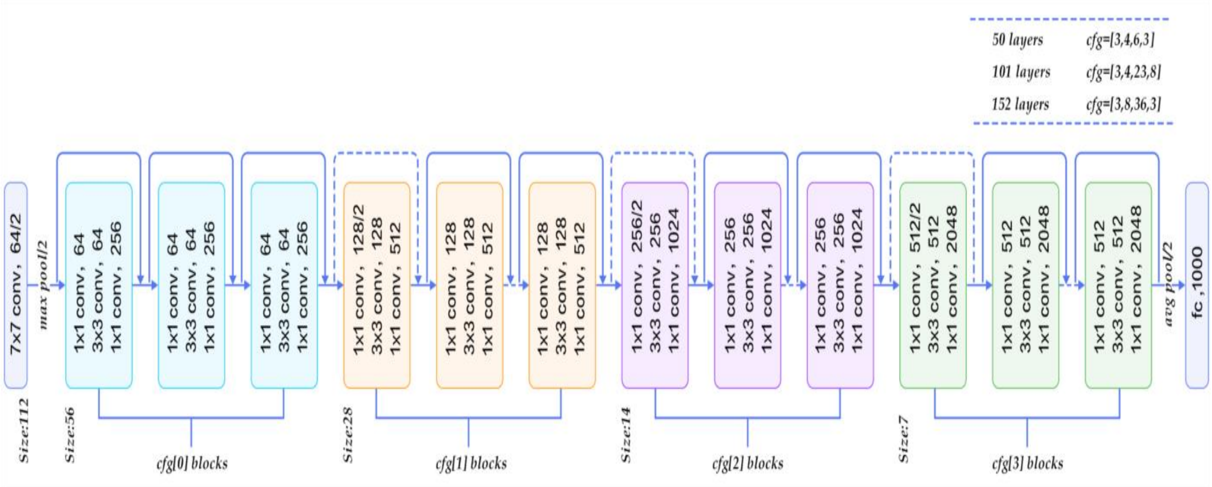

Figure 21ResNet50 architecture

SKIP CONNECTION is a direct connection that skips over some layers of the model. See figure (24). The output is not the same due to this skip connection. Without the skip connection, input ‘X gets multiplied by the weights of the layer followed by adding a bias term.

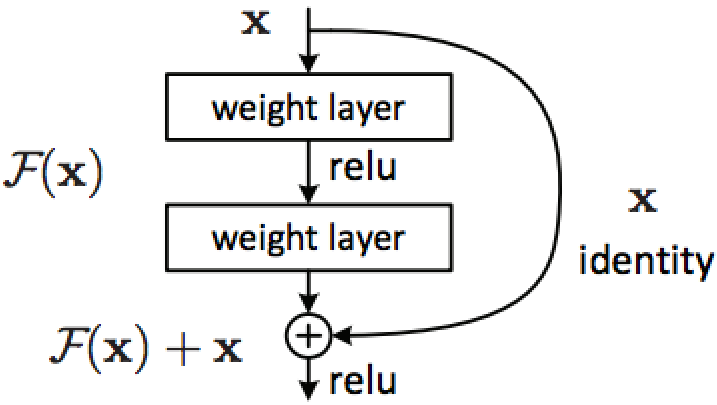

Figure 22SKIP CONNECTION

5- VGG16 and ResNet50

The ResNet-34 was the first ResNet design, which included inserting shortcut connections within a plain network to transform it into its residual network counterpart. The plain network was influenced by VGG neural networks (VGG-16, VGG-19) in this case, with 3 filters in the convolutional networks. ResNets, on the other hand, have fewer filters and are less sophisticated than VGGNets.

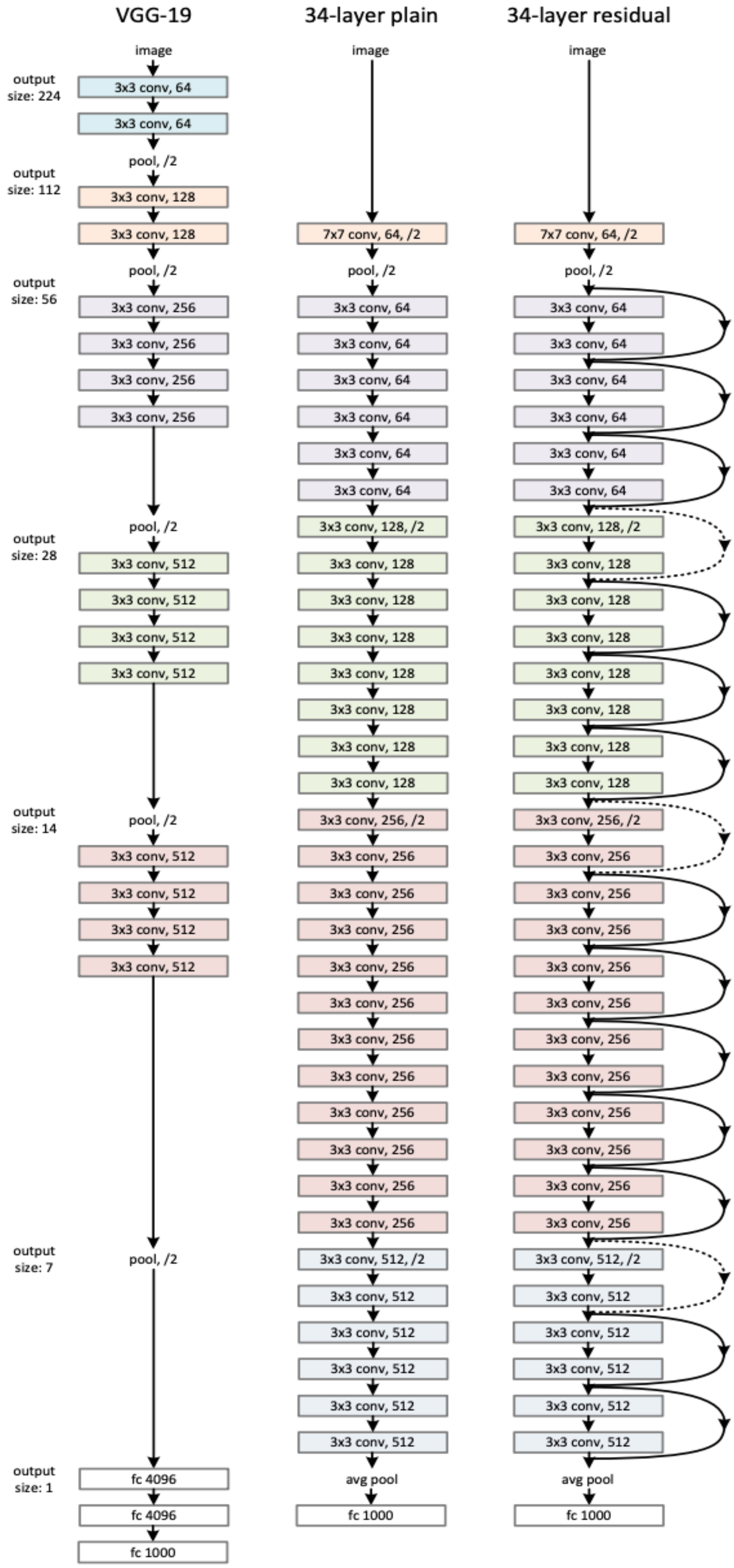

Figure 23ResNet50 and VGG16 and ResNet34

While the Resnet50 architecture is based on the above model figure (23), there is one major difference. In this case, the building block was modified into a bottleneck design due to concerns over the time taken to train the layers. Therefore, each of the 2-layer blocks in Resnet34 was replaced with a 3-layer bottleneck block, forming the Resnet 50 architecture. This has much higher accuracy than the 34-layer ResNet model.

6- Ensemble learning

Ensemble learning is a generic machine learning meta method that aims to improve predictive performance by combining predictions from many models.

Ensemble learning combines the mapping functions learned by different classifiers to generate an aggregated mapping function. Different solutions for computing this combination have been proposed over the years by various methodologies.

The most prevalent strategies that are commonly utilized in the papers are described below:

- **Bagging:** this technique is one of the earliest ensemble methods proposed. For this method subsamples from a dataset are created randomly. Meaning the same data can appear in different subsamples. These subsets are treated as independent datasets. On which different models will be fit. During test time the predictions from all models trained on different subsets of the same data are accounted. See figure (24).

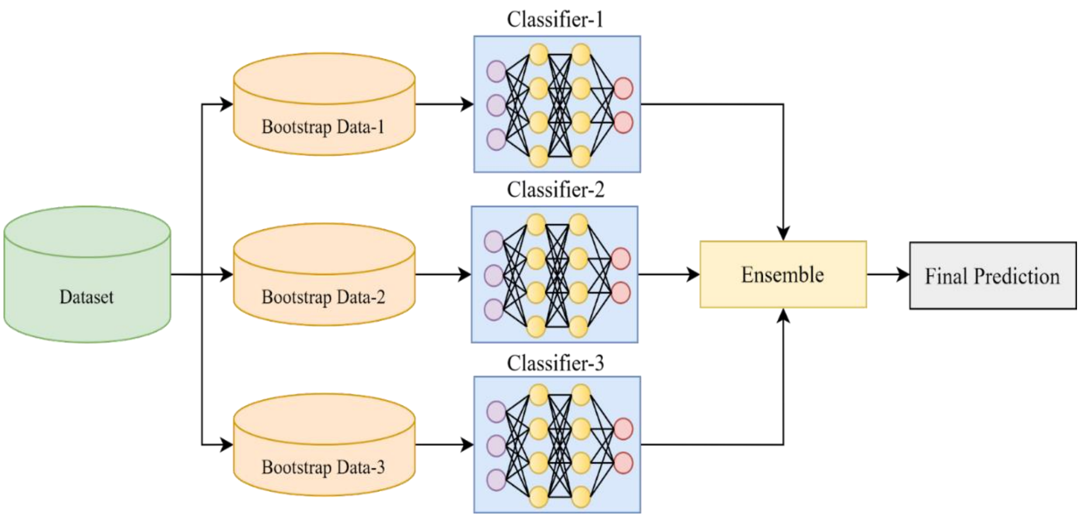

Figure 24bagging example

- **Boosting:** in this technique the dataset is loaded into the first classifier, and the predictions are examined. The misclassified data which classifier 1 fails to make valid prediction are sent to classifier 2 to concentrate on the problematic and learn a suitable decision boundary, The same step is taken if we have more classifiers. And the ensemble of all this these previous classifiers is then computed to generate the final prediction on the test data. See figure (25).

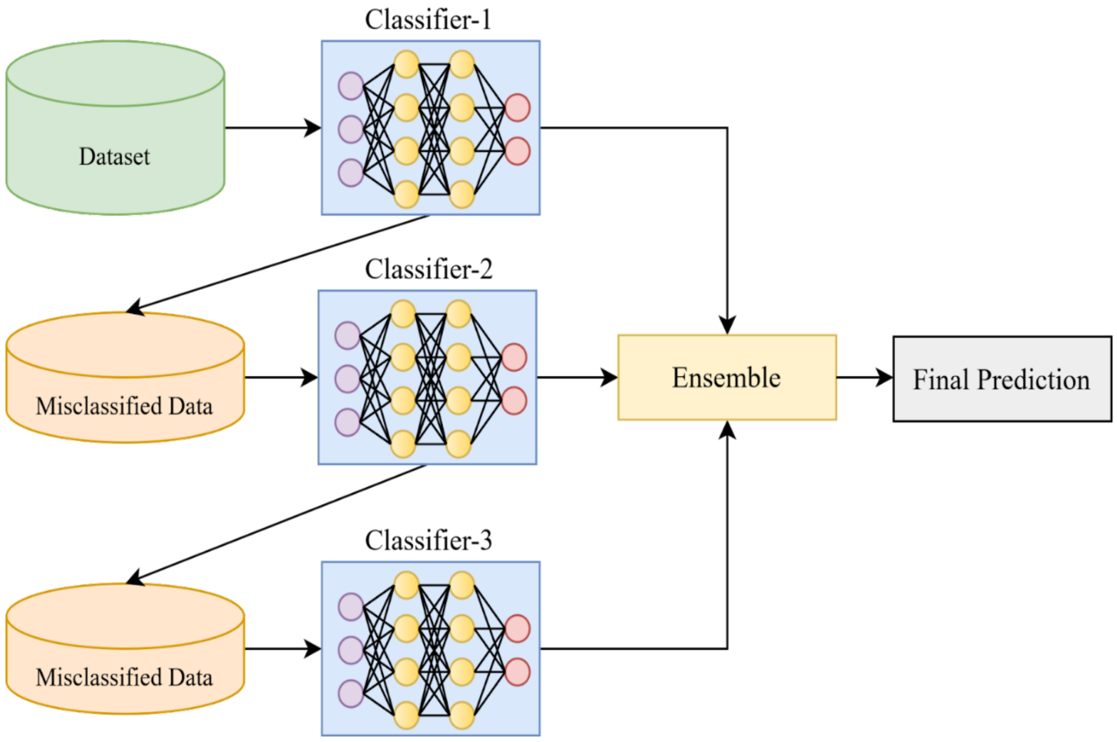

Figure 25Boosting method

- **Max rule:** The probability distributions produced by each classifier are what the "Max Rule" ensemble approach is based on. This approach uses the idea of the classifiers' "confidence in prediction". Here, for a predicted class by a classifier, the corresponding confidence score is checked. The class prediction of the classifier that predicts with the highest confidence score is deemed the prediction of the ensemble framework.
- **Weighted Probability averaging:** In this ensemble technique, the probability scores for multiple models are first computed. Then, the scores are averaged over all the models for all the classes in the dataset. Then we calculate a weighted average of the probability. The weights in this approach refer to the importance of each classifier, i.e., a classifier whose overall performance on the dataset is better than another classifier is given more importance while computing the ensemble, which leads to a better predictive ability of the ensemble framework.

CHAPTER4: EXPERIMENTAL RESULTS

1- Introduction

This chapter is dedicated to report our findings after performing all the steps mentioned above. We start by introducing the dataset on which experiments are carried out, afterwards, experimental results are reported.

2- Experimental protocol

2.1 Dataset presentation

We conduct our experiments on COVID-CT-Dataset [4]. This is made up of 349 positive COVID-19 CT-scan images from 216 patients and 397 CT-scan images that are negative to COVID-19. This dataset is open-sourced to the public.

Figure (26) depicts samples from this dataset with negative COVID-19. And figure (27) depicts samples with positive COVID-19.

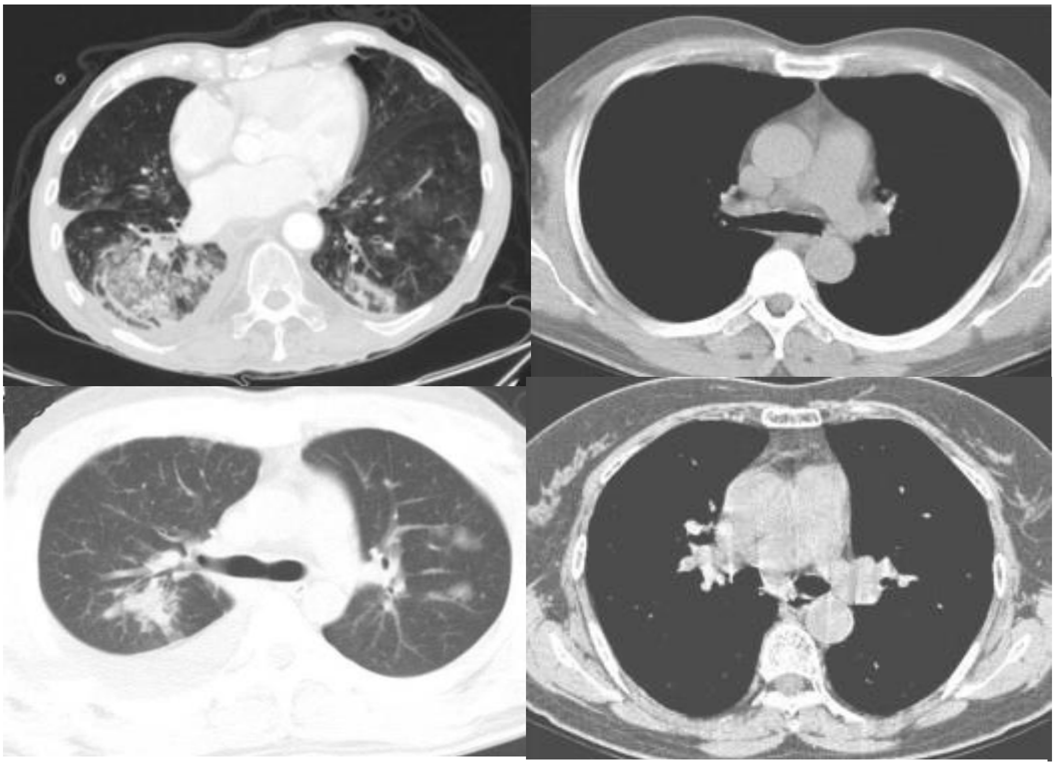

Figure 26Negative COVID-19 CT

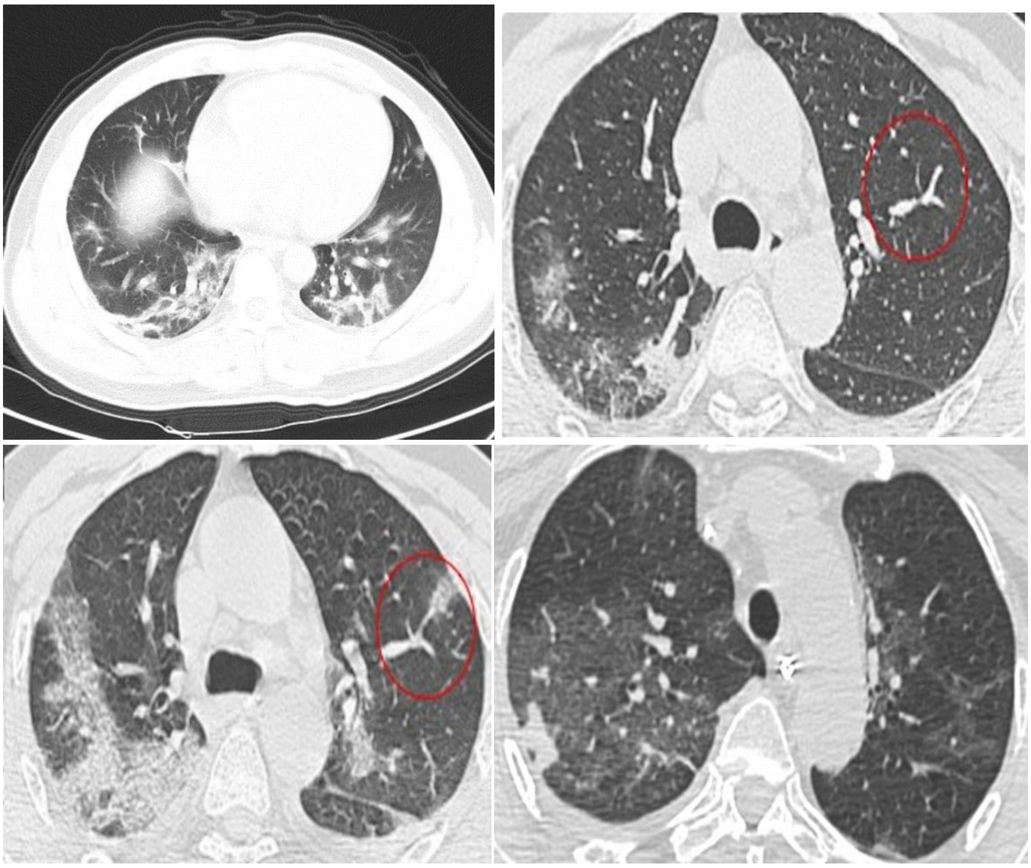

Figure 27Positive COVID-19 CT

2.2 Performance metrics

To measure the performance of the proposed system, Accuracy metric is adopted. Accuracy is defined as

$$\frac{\text{Number of correctly classified samples}}{\text{Number of images}}$$

3- Experimental results

3.1 First experiment: reporting the performance of pre-trained models

In this experiment, we report the obtained experimental results using fine-tunned VGG16 and ResNet50 models with the original dataset images. Table 1 shows the accuracy of each model. figure (28) and figure (29) show the changes throughout the training. For ResNet50 and VGG16 respectively.

	VGG16	ResNet50
accuracy	1	1
Validation accuracy	0.71	0.71
Train loss	0.0	0.0
Validation loss	6.95	2.82

Table 1Results of the pretrained models

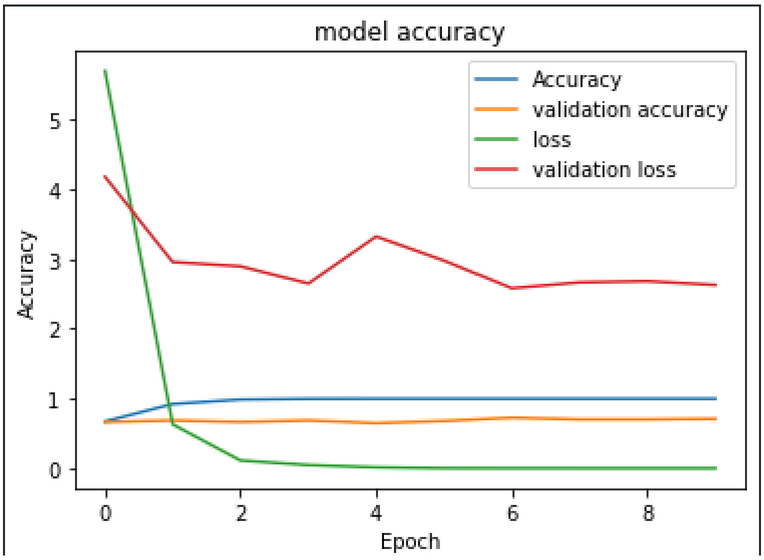

Figure 28ResNet50 performance

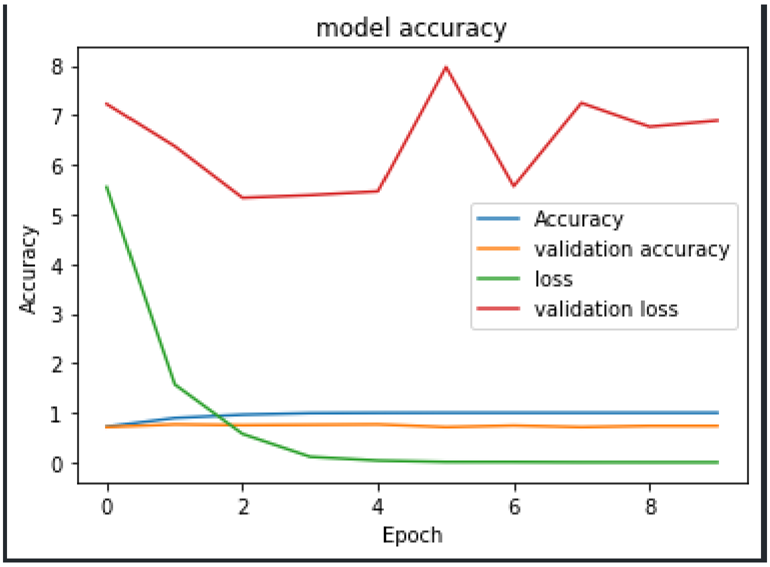

Figure 29VGG16 performance

From the table and figures above, we observe the following:

- We can see the accuracy and loss have perfect values which means our models are classifying the data perfectly
- The models achieved 71% on validation data for 10 epochs
- We can see the power of VGG16 and ResNet50 in classifying and generalization of data.

3.2 Second experiment: using data augmentation

In this experiment we report the obtained results using the same models. Adding data augmentation technique to our dataset images. Table 2 shows the accuracy of each model. figure (30) and figure (31) show the performance during training for VGG16 and ResNet50 respectively.

	VGG16	ResNet50
accuracy	0.94	0.93

Validation accuracy	0.64	0.68
Train loss	0.146	0.55
Validation loss	1.399	5.63

Table 2Results using data augmentation

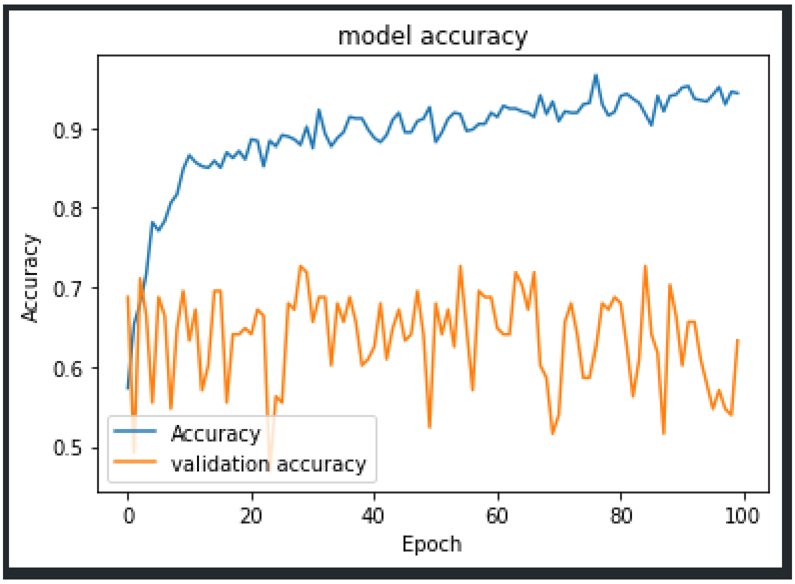

Figure 30VGG16 performance with data augmentation

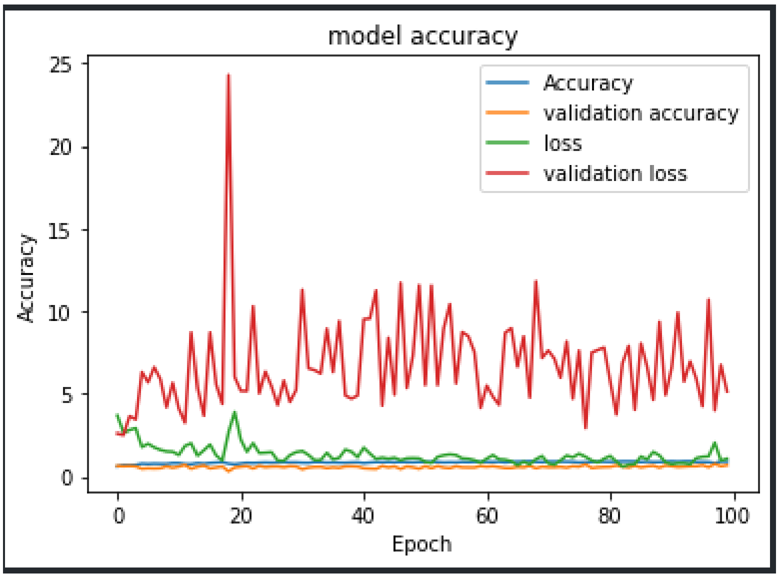

Figure 31ResNet50 performance with data augmentation

From the table and figures above, we observe the following:

- Data augmentation technique performed badly with CT-scan since the results for models without data augmentation performed better.
- Data augmentation is a well-known technique to help decrease the loss value but, in our case, it was not of a good use.

3.3 Third experiment: ensemble learning

In this experiment, we try to make an ensemble of the two trained models. In attempt to increase the accuracy. We do that by combining the prediction values of both models for the same test set images. Then, we chose the position of the maximum value and convert that probability into a prediction.

Table 3 shows the accuracy achieved using ensemble learning for VGG16 ResNet50 with data augmentation:

VGG16	ResNet50	Ensemble
0.88	0.87	0.87

Table 3Results of accuracy using ensemble learning and data augmentation

From the table above we observe the following:

- The accuracy is the average of both models.

Now using weighted probability averaging we do the same thing again with prioritizing VGG16 since the models achieved better results.

Table 4 shows the accuracy achieved using weighted probability averaging.

Weighted probability averaging
0.89

Table 4Results of accuracy using weighted probability averaging and data augmentation

From the table above we observe:

- The accuracy increased with small percentage.
- Ensemble learning technique helped increasing our accuracy.

Table 5 shows the accuracy achieved using ensemble learning for VGG16 ResNet50 without data augmentation:

VGG16	ResNet50	Ensemble
0.94	0.93	0.94

Table 5Results of accuracy using ensemble learning

From the table above we observe the following:

- The accuracy of models without data augmentation scored better results.
- The ensemble accuracy is the average of both models.

Now using weighted probability averaging we do the same thing again with prioritizing VGG16 since the models achieved better results.

Table 6 shows the accuracy achieved using weighted probability averaging.

Weighted probability averaging
0.95

Table 6 Results of accuracy using weighted averaging probability

From the table above we conduct that:

- Using weighted probability averaging increases the accuracy minimally.
- Models with no data augmentation performed better and their accuracy was higher

3.4 Fourth experiment: ensemble learning of the generated images

In this step we generate new images using the wavelet normalization function and predict their classes using the ensemble technique. Table 7 shows the accuracy obtained using this technique:

VGG16	ResNet50	Ensemble
0.47	0.47	0.47

Table 7Results of accuracy obtained using images manipulated by wavelet

From the table above we observe the following:

- The accuracy for both models decreased
- The wavelet functions didn't help in showing the features of the images so the classifier could not predict correctly.
- Maybe because we did not train the models using these images, they could not classify them.

GENERAL CONCLUSION

In the last two decades computer vision was a very active area and has made significant improvements. Computer vision has played an important role in medical technology it can help in the automation of tasks such as detecting cancerous moles in skin images…etc.,

After the COVID-19 outbreak, there were huge number of people infected with virus and was spreading rapidly. So, there was a shortage of medical staff and doctors so in attempt to help the medical staff and make it easier and quicker to diagnose people researchers started using X-ray chest scans and CT-scans to detect COVI-19.

In this thesis, we investigated the performance of deep networks in detecting COVID-19 from CT-scan images, our aim was to investigate the performance of VGG16 and ResNet50 in detecting COVI-19. As a second aim we compared the performance of these models with data augmentation and without it. Besides, as perhaps the last aim was to see how ensemble learning and using two models help increasing the accuracy in detecting COVID-19.

The proposed method can be divided into three phases, training, training using data augmentation and ensemble learning. The first phase is devoted to train our models with COVID-19 CT-scan images. In the second phase, training our models and performing data augmentation on COVID-19 Ct-scan. As last phase, ensemble learning the predictions of both models to determine which CT-scan is positive with COVID-19.

Experiments were carried out on COVID-19-CT dataset, with accuracy as performance metric. Our findings were a bit surprising. First the models trained without data augmentation performed better than models trained using data augmentation. Second ensemble learning helped increasing the accuracy but with small value. Third among the deep networks VGG16 performed slightly better.

Many other tracks are possible for the future works, among which using different deep networks to train the dataset, use a larger dataset this is motivated because it is difficult to fine-tuned deep networks with small datasets.

REFERENCE

1. Gonzales, R. C., & Woods, R. E. (2002). Digital Image Processing . Prentice-Hall, Inc. *New Jersey*..
2. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
3. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning. *Image Recognition*, *7*.
4. Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P. (2020). COVID-CT-dataset: a CT scan dataset about COVID-19. *arXiv preprint arXiv:2003.13865*.
5. Yan, Q., Wang, B., Gong, D., Luo, C., Zhao, W., Shen, J., ... & You, Z. (2020). COVID-19 chest CT image segmentation--a deep convolutional neural network solution. *arXiv preprint arXiv:2004.10987*.
6. Shah, V., Keniya, R., Shridharani, A., Punjabi, M., Shah, J., & Mehendale, N. (2021). Diagnosis of COVID-19 using CT scan images and deep learning techniques. *Emergency radiology*, *28*, 497-505.
7. Struc, V. (2011). The inface toolbox v2. 0 the matlab toolbox for illumination invariant face recognition. *Toolbox description and user manual, Ljubljana*.
8. Du, S., & Ward, R. (2005, September). Wavelet-based illumination normalization for face recognition. In *IEEE International Conference on Image Processing 2005* (Vol. 2, pp. II-954). IEEE.
9. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*, 211-252.
10. Yang, D., Martinez, C., Visuña, L., Khandhar, H., Bhatt, C., & Carretero, J. (2021). Detection and analysis of COVID-19 in medical images using deep learning techniques. *Scientific Reports*, *11*(1), 19638.
11. Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... & Xu, B. (2021). A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19). *European radiology*, *31*, 6096-6104.