_____

Detection of Pulmonary Embolism: Workflow Architecture and Comparative Analysis of the CNN Models

Akhilesh Tawte[1], Sudhanshu Gonge[2], Rahul Joshi[3], Preeti Mulay[4], Deepali Vora[5], Ketan Kotecha[6]

[1]Dept. of Artificial Intelligence & Machine Learning
Symbiosis Institute of Technology, Pune, India
akhilesh.tawte.mtech2021@sitpune.edu.in

[2]Dept. of Computer Science & Engineering
Symbiosis Institute of Technology, Pune, India
sudhanshu.gonge@sitpune.edu.in

[3]Dept. of Computer Science & Engineering
Symbiosis Institute of Technology, Pune, India
rahulj@sitpune.edu.in

[4]Dept. of Computer Science & Engineering
Symbiosis Institute of Technology, Pune, India
preeti.mulay@sitpune.edu.in

[5]Dept. of Computer Science & Engineering
Symbiosis Institute of Technology, Pune, India
deepali.vora@sitpune.edu.in

[6]Dept. of Computer Science & Engineering
Symbiosis Institute of Technology Pune, India
director@sitpune.edu.in

Abstract—Machine learning has proven to be a practical medical image processing technique for pattern discovery in low-quality labelled and unlabeled datasets. Deep vein thrombosis and pulmonary embolism are both examples of venous thromboembolism, which is a key factor in patient mortality and necessitates prompt diagnosis by experts. An immediate diagnosis and course of treatment are necessary for the life-threatening cardiovascular condition known as pulmonary embolism (PE). In the study of medical imaging, especially the identification of PE, machine learning (ML) algorithms have produced encouraging results. This study's objective is to assess how well machine learning (ML) algorithms perform in identifying PE in computed tomography (CT) scans. A range of ML approaches were used to the dataset, including deep learning algorithms such as convolutional neural networks. The effectiveness of PE detection systems can be greatly enhanced by the use of cutting-edge methodologies like deep learning, which lowers the possibility of incorrect diagnoses and enables the quick administration of therapy to individuals who require it. This work contributes to the growing body of evidence that supports the use of ML in medical imaging and diagnosis. Future research should examine how these algorithms might be included into clinical workflows, resolving any potential implementation challenges, and making sure their adoption is done so in a secure and efficient way. In this study, we provide a thorough evaluation of three different models: the streamlined architecture MobileNetV2 with an accuracy of 96%, compared to other models like the Xception model with an accuracy of 91%, and the Efficientnet B5 model with an accuracy of 97%, after observation and process following.

Keywords-Pulmonary Embolism, Machine Learning, CNN, MobileNet V2, Xception, EfficentNet, Segmentation.

I. INTRODUCTION

A wide variety of medical disorders can be diagnosed with the help of medical imaging. Yet, manual image analysis for medical purposes can be laborious and prone to human mistake. Machine learning (ML), which makes use of the creation of algorithms that can learn from and predict from massive volumes of data, has emerged as a viable remedy to these problems. Recent results from the use of ML approaches to medical imaging have been positive, offering chances to improve the precision, effectiveness, and affordability of medical imaging analysis. The purpose of this work is to review recent advances in medical image analysis using ML.

Pulmonary embolism is one of the most challenging diagnoses to make because it can be fatal if not caught in time. A blockage of a pulmonary artery in the lungs can result in pulmonary embolism (PE), a potentially dangerous cardiovascular illness. A blockage of one of the pulmonary arteries in the lungs can cause pulmonary embolism (PE), a potentially fatal cardiovascular condition. A blood clot that

299

_____

develops in a deep vein, such as the legs, and then moves to the lungs frequently causes PE. The term "deep vein thrombosis" refers to this kind of blood clot (DVT). Less blood flowing to the lung's afflicted area due to pulmonary artery obstruction can result in tissue damage and a reduction in the body's ability to absorb oxygen[1].

As the symptoms of PE frequently resemble those of other disorders, such as a heart attack or pneumonia, diagnosing PE can be difficult. A physician might prescribe tests like a computed tomography (CT) scan, a lung scan, or an MRI, as well as blood tests to check for the presence of clots, to confirm a diagnosis of PE. Blood tests may not always give a reliable diagnosis, and manual interpretation of medical imaging can take time and be susceptible to human error..

Machine learning techniques have recently gained popularity in medical research and diagnostics, offering a fresh and improved means of identifying specific medical disorders. ML is a subfield of AI that comprises formulating programmes specifically designed to learn new tasks that can generate predictions and learn from massive volumes of data. Medical imaging analysis using ML algorithms has recently demonstrated encouraging outcomes, including the diagnosis of PE. ML algorithms can analyse enormous amounts of data and find patterns and features that may be missed by human interpretation by applying cutting-edge techniques like deep learning.

The aim of this study is to see how well ML algorithms perform in detecting PE in CT scans. The collection and processing of a dataset of CT scans from patients with and without PE. The dataset was subjected to a range of ML techniques, including deep learning algorithms that make use of CNNs. The best algorithms for identifying PE were then determined by comparing the findings of this investigation. This study is notable because it illustrates the potential of utilising ML algorithms in the detection of PE and contributes to the growing volume of study showing the benefits of ML in medical imaging and diagnostics. Increased diagnosis speed and accuracy can lessen the chance of misdiagnosis and enhance patient outcomes[2].

In this paper, we give a general overview of PE, covering its aetiology, symptoms, and available diagnostic techniques. Then, we give a backdrop and rationale for utilising ML algorithms to identify PE, as well as a summary of the different ML approaches that were tested in this study. A discussion of the implications of these findings and recommendations for further research follow the presentation of the study's methodologies and findings. The research also emphasises the difficulties and limitations of using ML to medical imaging, such as the need for big, diverse datasets to train ML systems, as well as data privacy and security issues.

The results of this study show how machine learning (ML) possesses ability to revolutionise medical imaging analysis by providing better accuracy, speed, and cost-effectiveness than manual techniques. However, more investigation is required to fully understand the advantages of ML in medical imaging and to address any challenges that would prevent its deployment. Further study in the areas of medical imaging and ML is suggested as a possible direction in the paper's conclusion.

Machine learning (ML) models are required for the detection of pulmonary embolisms (PE) for several reasons:

Early and precise diagnosis are essential for the best results because PE is a disorder that can be fatal and necessitates quick medical attention. ML algorithms are able to analyse medical images quickly and automatically, including computed tomography (CT) scans, and can identify PE more effectively than manual techniques[3].

Increased precision: PE diagnosis can be tough because the symptoms are frequently similar to those of other illnesses and might be challenging to distinguish. Medical image analysis using machine learning (ML) algorithms can spot patterns and details that could be overlooked by human interpretation. Its increased precision may aid in lowering the possibility of a false positive and enhancing patient outcomes[3].

Cost-effectiveness: ML algorithms can automate the diagnostic procedure and cut down on the time and resources needed for manual analysis, making them cost-effective. This may contribute to lowering healthcare expenses and improving access to diagnostic services for those who require them. ML algorithms can automate the diagnostic procedure and cut down on the time and resources needed for manual analysis, making them cost-effective. This may contribute to lowering healthcare expenses and improving access to diagnostic services for those who require them[4].

Addressing manual methods' limitations: Manual analysis of medical imaging can be laborious and prone to human error. These restrictions can be solved by ML algorithms by delivering quicker, more precise, and more reliable results[4].

A. *Pulmonary Embolism*

A thrombus (clot of blood) that breaks free from a vein and lodges in the pulmonary circulation, blocking one or more pulmonary arteries, is known as pulmonary embolism (PE).Diagnosis of PE involves a combination of patient medical history, physical examination, and various diagnostic tests such as chest X-rays, CT scans, and blood tests measuring biomarkers associated with PE. The standard treatment for PE is anticoagulant medication, such as heparin or warfarin, to prevent further clots and dissolve existing ones. In severe cases, a procedure known as thrombolysis, in which a clot-dissolving medication is directly delivered to the clot, may be used. In

300

_____

certain instances, a filter may also be inserted in the inferior vena cava to prevent clots from reaching the lungs.

In the process of identifying and treating pulmonary emboli, machine learning can be quite helpful (PE). These are some applications for machine learning:

1. Predictive modelling: Machine learning algorithms can be trained on massive datasets of patient medical history, demographic data, imaging results, and other pertinent characteristics to predict the likelihood of a PE diagnosis. This can assist medical professionals in better deciding whether additional diagnostic procedures or medical therapy are necessary.

2. Imaging analysis: Machine learning algorithms can be used to analyze imaging data, such CT scans, to find PE symptoms and gauge how severe the blockage is. This can aid medical professionals in making more precise diagnosis and choosing the most appropriate course of action[5].

3. Risk assessment: Based on a range of variables, including age, medical history, lifestyle, and genetics, machine learning can also be used to determine an individual's risk of having PE. Healthcare professionals can utilise this information to monitor individuals who are at high risk of PE and to prioritize preventive interventions[6].

4. Monitoring and tracking: Machine learning algorithms can be used to monitor patients who have been given a PE diagnosis and to follow their development over time. Healthcare professionals can use this information to identify patients who could be at risk for issues and to modify treatment regimens as necessary[6].

B. Convolution Neural Networks (CNN)

For tasks like image identification, object detection, and image segmentation, convolutional neural networks (CNNs), a type of artificial neural network, are frequently employed in image and video analysis. The principle of convolution, which involves swiping a tiny matrix known as a kernel or filter over the input image to extract features, is the foundation for CNNs[7].

Due to their capacity to learn and recognize intricate patterns within images, convolutional neural networks (CNN) are frequently utilized for the detection of medical imaging, including X-rays, CT scans, and MRIs. CNN works the following way:

Input Layer: The input to the CNN is an image or a sequence of images. The image is taken in by the input layer and sent on to the following layer.

Convolutional Layers: At the center of the CNN are convolutional layers. They take features out of the input image using a collection of filters or kernels. Each filter performs a dot product with the image's pixel values as it slides over the input image. A feature map that highlights parts of the picture that contain the feature the filter is looking for is the output that is produced as a result.

Activation Function: For the network to gain nonlinearity, an activation function is used to each convolutional layer's output. Rectified Linear Unit (ReLU) and Sigmoid are often used activation functions.

Pooling Layers: By using pooling layers, the output of the convolutional layers are down sampled, causing the feature maps' spatial dimensions to be reduced. As a result, the network's parameters are reduced, improving its computational efficiency.

Fully Connected Layers: These layers perform classification or regression operations after receiving the output from the preceding levels. These layers are similar to the layers in a standard artificial neural network and employ the extracted features to classify the input image.

Output Layer: The output layer creates the CNN's ultimate output, which may be a regression or classification label.

To decrease the discrepancy between the expected and actual output, the CNN modifies the weights of the filters and fully connected layers during the training phase.

In order to detect medical images like X-rays, CT scans, and MRIs, convolutional neural networks (CNNs) are frequently utilised. This is because CNNs have the capacity to recognise and understand complicated patterns in images. The automated learning capacity of convolutional neural networks (CNNs) and recognize complex patterns within the pictures makes them a popular choice for the detection of abnormalities and diseases in CT scans[8]. The following are some of the crucial functions CNNs in the detection of CT scans:

Convolutional layers are used by CNNs to extract features from CT scan pictures. The image is subjected to a series of filters or kernels known as convolutional layers in order to recognise patterns and features including edges, curves, and textures. The retrieved features are then applied to the image to determine if it is normal or aberrant[8].

Classification: Having retrieved the features, the CNN classifies the image as normal or abnormal using a collection of fully linked layers. By changing the weights of the fully linked layers during training, the CNN learns to link the retrieved characteristics with the associated labels[9].

Localization: CNNs can also be utilized to pinpoint a disease or abnormality's exact location inside a CT scan image. Localization networks, which forecast the position and size of the anomaly inside the image, are used to do this.

Segmentation: Another use of CNNs in CT scan detection is segmentation, which entails separating the abnormality or disease from the surrounding image. Segmentation networks, which divide the image into regions of interest using convolutional layers, are used to achieve this.

By automatically learning and recognizing complicated patterns within the images CNNs play a crucial part in the

301

_____

evaluation of anomalies and diseases in CT scans. A number of medical diseases can now be diagnosed and treated more accurately and efficiently because to the use of CNNs in CT scan processing.

CNNs have shown encouraging outcomes in the early and precise diagnosis of pulmonary embolism when used for CTPA image detection. The application of CNNs in the early and accurate diagnosis of pulmonary embolism has demonstrated great levels of accuracy and performance in the detection of CTPA pictures. The standard of the training data, the intricacy of the medical condition being diagnosed, and the specific architecture of the CNN being utilized are all variables that might have an impact on the model's performance and accuracy.

II. RELATED WORK

A few of the most effective methods for integrating, analyzing, and drawing conclusions from copious amounts of data , diverse data sets is through machine learning (ML) [10]. Among the most important uses for artificial intelligence (AI) systems is the diagnosis of disease using image processing and machine learning [11]. Early work on employing clinical factors or perfusion imaging as inputs to neural networks rather than CTPA led to moderate generalization when ML was applied to the problem of PE detection [12]. Other early approaches relied on complicated feature engineering and pre-processing to automate identification without external validation, which made them expensive from multiple angles [13]. By directly creating the learning architectures, the use of Deep Learning minimizes or eliminates the requirement for human feature engineering on images and gives designers greater freedom when designing the end-to-end pipeline for this automated process. Diagnosing PE using CTPA imaging has become a focus of more recent strategies. This is advantageous since CTPA imaging is the imaging technique that PE examinations use the most, and the data collected allows for a firm diagnosis using various learning strategies. Furthermore, the newly-emerging field of applying Deep Learning techniques to medical applications enables setting-up an ever-improving process of collaboration between humans and machines by delegating the initial diagnosis to automated DL pipelines and further examination to human doctors. As a result, even if the model is not the most accurate, it would still speed up diagnosis time by enabling doctors to better identify where to conduct additional research. In contrast to [14], which separates each CTPA into smaller cubes instead of working with a full CTPA scan, [13] substantially preprocesses image features using segmentation and vessel-alignment. These features are then provided as an input to a 3D CNN. Our method works on whole CT scans rather than splitting them up, and it requires substantially less feature engineering than previous methods. It simply involves segmentation to extract the key features from the pictures and

applying basic transformations. Our findings therefore offer a stronger argument for integrability with current medical diagnostic processes if they are upgraded

III. PROPOSED MODEL

A. Dataset Preprocessing and Exploration

About Dataset:

A collection of computed tomography (CT) scans that have been labelled to show whether or not pulmonary emboli are present may be found in the RSNA Pulmonary Embolism (PE) dataset. The Radiological Society of North America (RSNA) Pulmonary Embolism Detection Challenge, which sought to increase the precision of automatic pulmonary embolism identification on CT images, inspired the creation of this dataset[15].

12,126 CT scans in all, including 10,330 for training and 1,796 for testing, make up the dataset. A series of two-dimensional CT images are combined to create a three-dimensional volume for each scan. The existence or absence of pulmonary embolism, as well as the location and severity of any emboli that are present, have all been identified on the pictures by board-certified radiologists[16].

The RSNA PE CT Dataset, which is available to the public, is the biggest dataset of CTPA trials that has been expertly annotated that we are aware of. This dataset aims to further machine learning research and innovation, which will ultimately improve the calibre, efficacy, and accessibility of patient care around the globe. It is freely available to all researchers who want to use it for academic purposes.
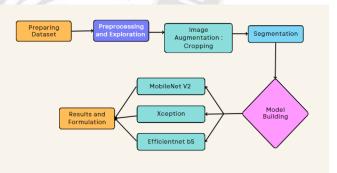

Fig 1. Proposed model and workflow

_____

Dataset Exploration:

The Dataset has following labels to work with:

Label Name	Explanation	Description
PE – Negative Exam	PE is not present	Analysis
Central PE	PE involving the saddle embolus that is present in either the left or right major pulmonary arteries	Analysis
Left Sided PE	PE found in the left pulmonary arterial tree at or above the lobar level	Analysis
Right Sided PE	PE seen in the right pulmonary arterial tree at or above the lobar level	Analysis
Chronic PE	Only those with persistent PE were observed.	Analysis
Acute and Chronic PE	The study includes both acute and chronic PE.	Analysis
Filling Defect – No PE	Examples of a true intraluminal filling defect that is not a normal PE include tumor invasion, thrombus in a stump, a catheter, and an embolized wire.	Analysis
QA-motion	Patient movement that is significant enough to compromise imaging quality and prevent PE from being detected or excluded	Analysis
QA-constant	Poor pulmonary arterial tree opacification that is severe enough to prevent PE from being detected or excluded from the picture	Analysis
RV to LV ratio < 1	Maximum short-axis diameter of RV to LV in a typical ratio	Analysis
RV to LV ratio >=1	Elevated RV/LV ratio, which denotes the possibility of right heart stress	Analysis
PE – present on the image	PE (acute, chronic, or both) is labelled at the image-level when it appears in a picture.	Image
Artifact flow	A filling deficiency that appears to be caused by PE but is actually caused by sluggish blood flow or contrast mixing	Analysis
Indeterminate	The poor image quality prevented either a negative or positive diagnosis from being made.	Analysis

Table 1. Introduction to labels of the dataset

A study has multiple images grouped into series. Thus, each image is characterized by identifiers:

• StudyInstanceUID: Indicates a unique identifier for the study to which the image belongs

• SeriesInstanceUID: Indicates the series in the study to which the image belongs

• SOPInstanceUID: A unique identifier for the image itself, defining the slice of a patient's data pertaining to the chest (in the context of CT scans, there are several images taken at different locations and angles from the patient body called slices).

For each training image there are 14 labels, that belong to 3 levels:

• Image-Level Labels Label for each image. Here, the only label at the image-level is the one used to indicate whether a PE is present on the image or not.

• Exam-Level Labels: These labels are used to detect the labels corresponding to characteristics of the PE and include the labels corresponding to the ratio of the right valve to left valve (greater than or less than 1), location of the PE (left, right,

center), nature of PE (acute, chronic or both) and a label to indicate whether the study was conclusive or not.

• Informational Labels: These labels are used to indicate whether the radiologists noted an issue with motion or contrast in the study, whether there were artifacts present or whether the thing diagnosed was something else other than PE. We used the insights obtained in the paper to determine the labels that need to be predicted, and finally narrowed the labels down to 9[16]:

acute_and_chronic_pe, central_pe,

chronic_pe, indeterminate, leftsided_pe,

pe_present_on_image, rightsided_pe,

rv_lv_ratio_gte_1, rv_lv_ratio_lt.

A worldwide standard for the sharing, storage, and transmission of digital medical images is called Digital Imaging and Communications in Medicine (DICOM), and this is how the images are saved[17]. A variety of 2D picture slices of various patients are included in the study. By giving the patient iodine and contrast to the chest, these images are produced. The information in the raw pixels comes from several examples of the exam's phases and through various tissues. The Hounsfield

_____

Unit (HU) scale, which ranges from -1000 HU for air to 0 HU for water, is used to estimate the density of tissues[18]. On disc versus in memory, CT scan images could be represented differently. This problem typically arises because DICOM images frequently store as unsigned integers, even though they may have negative values. We may convert the values to HU units by using the intercept and slope values from the DICOM data format (supplied by the device maker). So, we modified the CT numbers through a procedure known as segmentation to extract the relevant information from the images [19].
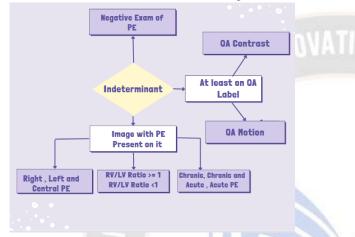

Fig 2. Digram of the workflow of labels of dataset

The CT scan records data regarding the radiodensity of a substance or tissue that has been subjected to x-rays. After gathering measurements in numerous different directions, a scan's transversal slice is recreated.

Since the spectrum makeup of x-rays varies on measurement conditions like acquisition parameters and tube voltage, we must convert to Hounsfield units. The images of various measures can be compared by normalising to the values of water and air (water has HU 0 and air -1000)[18].

Hounsfield Unit:

In medical imaging, a scale called the Hounsfield unit (HU) is used to quantify the tissue density. Sir Godfrey Hounsfield created the Hounsfield scale, which is used to quantify the tissue density in computed tomography (CT) scans. The scale goes from -1000 HU, which stands in for air, to +1000 HU, which stands in for bone[20].

Hounsfield units are used to discern between various tissues and structures, such as blood vessels, lung parenchyma, and pleural fluid, in computed tomography pulmonary angiography (CTPA) pictures. Hounsfield units can be used to distinguish between healthy and unhealthy tissues, which makes it simpler to spot and identify diseases like pulmonary embolism[21].

For instance, blood arteries are often represented by higher Hounsfield units in a CTPA image because they are denser than the nearby tissues. Lower Hounsfield units are used to represent

pleural fluid and pulmonary enema because of their lower densities. Healthcare professionals can precisely identify the presence of aberrant tissues, such as clots or obstructions in the pulmonary arteries, which may suggest the existence of PE, by analysing the Hounsfield units in CTPA images.

Hounsfield units are an essential part of CTPA images and are crucial for the detection of diseases like pulmonary embolism. Hounsfield units help healthcare professionals properly distinguish between normal and pathological tissues and make more informed judgements about patient treatment by quantitatively describing the density of tissues in CT scans.
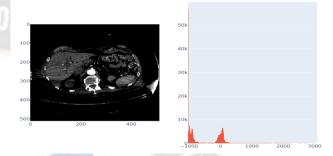

Fig3.1 HU values for a single CTPA Image

In computed tomography pulmonary angiography (CTPA) pictures, the use of Hounsfield units in the identification of pulmonary embolism (PE) is advantageous for a number of reasons:

1. More accuracy: Hounsfield units quantify tissue density and make it simpler to discern between normal and diseased tissues. The likelihood of misdiagnosis is decreased thanks to the increased accuracy, which also enables healthcare professionals to make better informed decisions about patient care[21].

2. Removing subjectivity from the diagnosis process is made possible by the use of Hounsfield units, which offer a consistent way to assess tissue density. This guarantees that different healthcare professionals receive the same results and helps to eliminate variability in the interpretation of CTPA images[21].

3. Automation: The PE detection procedure can be automated with the use of machine learning algorithms and Hounsfield units. Due to the absence of human mistake, manual interpretation requires less time and resources[21].

4. Consistency: By incorporating Hounsfield units into CTPA pictures, results are guaranteed to be uniform across various imaging systems and institutions. This makes it simple to compare the results and enhances the precision of diagnosis[21].

Using Hounsfield units enables accurate, effective, and uniform PE diagnosis, resulting in the best management for patients.

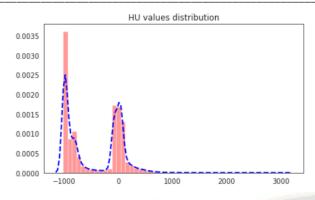The result of HU exploration was as follows:

_____

Fig.3.2 HU values of all the images in the dataset

With -320 we are separating between lungs (-700) /air (-1000) and tissue with values close to water (0).
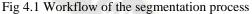
B. Segmentation

A system is trained to recognize and distinguish particular features or regions of interest within an image as part of the machine learning process known as medical image segmentation. This is accomplished by segmenting an image into many areas or segments, every area having their own unique set of attributes and properties. Images from CTPA (Computed Tomography Pulmonary Angiography) may be segmented into its component parts so that the pulmonary arteries, which are crucial in cases of pulmonary embolism, can be isolated, enabling a more precise identification of PE[22].

The outcome of image segmentation method is a collection of segmented images that clearly show the pulmonary arteries and enable a more precise PE detection. These segmented pictures can be utilized to confirm a PE diagnosis in conjunction with other diagnostic procedures, such as blood testing and clinical examinations.

A crucial step in the process of applying machine learning to detect PE is segmenting CTPA images. Algorithms for machine learning can identify and measure the presence of PE more precisely and increase the precision of diagnosis by isolating and highlighting the pulmonary arteries[23].

Steps followed during the Implementation were as follows:

Fig 4.1 Workflow of the segmentation process

Step 1: For scans that are washed out, find the average pixel value near to the lungs.
This is given as:

image = image-mean
image = image/std

middle = image [(column size/5) : (column size/5*4), (row size/5) : (row size/5*4)]
mean = mean(middle)
max = max(image)
min = min(image)

Step 2: Moving the underflow and overflow on the pixel spectrum will improve threshold detection.

In image processing methods like segmentation, which divides several tissues based on their pixel values, the threshold value is an important component. The threshold value establishes which pixel values are regarded as background noise and which as being a part of the target tissue.

The range of pixel values in the image must be taken into account when choosing a threshold value. The CT machine's technical limits or the existence of picture distortions, however, may cause some pixel values to fall outside of this range.

The threshold value needs to be changed to include these extreme pixel values in order to move the underflow and overflow on the pixel spectrum. The complete range of pixel values is used in this way, which may produce better segmentation outcomes[24].

This is given by:
image [image = = max] =mean
image [image = = min] =mean
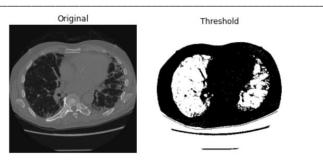The observed result is as follows.

_____

Fig 4.2 Output after changing the threshold values of the orignial image

Step 3: Separating the background (lung/air) from the foreground (soft tissue/bone) using Kmeans.

The goal of utilising k-means to separate the foreground in CT scans is to divide the image into distinct regions that correspond to various structures or tissues in the body.

The technique requires a set of features that describe the picture pixels in order to be used for foreground segmentation in CT scans. These characteristics are commonly represented by the pixels' Hounsfield Units (HU) in CT images. several people in the community, the government and the defence of the bone[24].

It is also necessary to provide the number of clusters that k-means will produce. The quantity of clusters on a CT scan often reflects the variety of tissue types shown in the image.

K-means will iteratively allocate each pixel to the nearest cluster centre after determining the characteristics and the number of clusters, and it will change the centres to decrease the total squared distance between each pixel and its corresponding centres. When the maximum number of iterations has been reached or the assignment of pixels to clusters stops changing, this process is repeated[25].

Using k-means, we may divide the image pixels into foreground and background clusters in CT scans to distinguish the foreground. By calculating the distance between each pixel and the two cluster centres, it is possible to determine which pixels are closest to the foreground cluster. A binary mask that distinguishes the foreground from the background can be made using the chosen pixels[25].

Step 4: First erode the finer elements, then dilate to include some of the pixels surrounding the lung.

In morphological image processing, erosion and dilation are two fundamental procedures that are used to change the geometry of objects in an image. Other image processing methods, such segmentation and edge detection, are frequently combined with these procedures.

Erosion:

Erosion is a morphological technique in which the boundaries of objects in an image are shrunk. The method involves sliding a structuring element (typically a tiny binary image) over the original image and replacing each pixel in the image with the minimum value of all pixels in the structuring element that overlap with the pixel. As a result, the boundaries of the items in the image are lost, making the objects smaller[26]. Little airways and blood arteries that are not a part of the lung tissue can be removed using erosion in a CT image of the lungs. The result is a segmented image that solely displays lung tissue. The following equation defines Erosion:

$$M \ominus N = \{x | (N^\wedge)A \in M\}$$

The structural component is only considered in the equation when it is either equal to or a fraction of the source images M. Fig. shows this procedure. Once more, the white square denotes O and the black square denotes 1.[27].
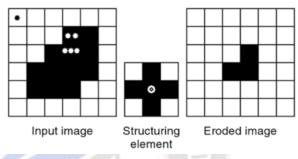

Fig 4.3.1 Visualization of the Erosion Process

Dilation:

The process of dilation, which is the reverse of erosion, entails enlarging the boundaries of the objects in an image. In order to perform the process, a structuring element is slid over the original image, and The maximum value of all the pixels is then used to swap each pixel in the image in the structuring element that overlap with the pixel. The outcome is an expansion of the image's object boundaries, which makes the objects larger[28].

In a CT scan of the liver, dilation can be used to join minor blood veins and fill in spaces between the liver lobes. The end result is a segmented picture that displays the entire liver tissue.

The following equation defines Dilation:

$$M \oplus N = \{x | [(N^\wedge)A \cap M] \in M\}$$

where N is the copy of N that has been rotated about the centre. According to Formula, N will include at least one component that as the structural component N dilates the image M, it coincides with one of the components in M.

If this is the case, "ON" will appear at the location where the structural component is cantered on the image. This procedure is shown in Fig. 4.3.2 is represented by the black square, and 0 by the white square.[27].

_____

Fig 4.3.2 Visualization of the Dilation Process

The Results of the following were as follows:

Fig 4.3.3 Output after Erosion and dilation process

Step 5: Applying colour labels.

The visibility and interpretation of the image can be improved by applying a colour label during the processing of CT scan images.

Based on the attenuation or absorption of X-rays by the tissue or material being imaged, different portions of the image are given different pixel values during the processing of CT scan images. Depending on the density of the tissue or material, the pixel values can range from low to high, with low pixel values denoting low density and high pixel values denoting high density.

The image can be color-coded to draw attention to particular elements or structures by giving distinct ranges of pixel values a colour designation. Bone tissue, for instance, might be given a white colour label, while soft tissue might be given a grey or blue colour label[29].

Colour labelling can also be used to divide or separate various areas of the image for analysis or visualisation. For instance, colour labelling in medical imaging can be used to distinguish between distinct tissue types, such as the brain's grey and white tissue, or to separate the brain from the skull.

Overall, applying a colour label on CT scan image processing can enhance accuracy and efficiency of image analysis, making it simpler to recognise and interpret important features or structures[30].

The results of the following are as follows:

Fig 4.4 Output after applying colour labels

Step 6: Applying Mask.

While processing CT scan image data, applying a mask entail choosing and isolating a particular area or region of interest within the image. Because the mask is a binary picture, it only comprises the values 0 and 1, with 1 denoting the region of interest and 0 denoting the rest of the image[30].

Using a mask could be advantageous when processing CT scan images for several reasons. A mask, for instance, can be utilised for:

Eliminate unwanted structures or noise: By using a mask, it is possible to remove undesired structures or noise from the image that can obstruct analysis or interpretation. When working with photos that have artefacts or other kinds of noise, this is quite helpful.

Using a mask can assist in concentrating the analysis on structures or areas of interest in the image. The area of medical imaging might benefit from this, where it is frequently necessary to isolate anatomical components in order to make a diagnosis or determine a course of therapy.

Increase efficiency: Using a mask can increase the efficiency of image processing by confining the evaluation to a narrow region of interest, enabling faster and more accurate analysis[31].
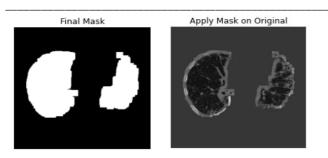
The results are as follows:

_____

Fig 4.5 Output after applying final mask

Final results of the segmentation process was as follows:

Fig 4.6 Final output after the segmentation process

C.   Model Building

The image is sent into the model network for classification, which then extracts latent features and averages them using a Global Average Pooling Layer. After that, a dropout of 25% is performed, followed by 5 back-to-back fully connected layers that reduce the features to 64, which are then fed to 9 fully connected layers—1 for each label—activated by sigmoid to forecast the label probability.

The Architecture is divided into two sections:
(1) A convolutional base to extract the image's latent characteristics.
The three model networks were utilized as the convolutional foundation. Three flows are used to transfer the data: The input flow has five convolutional layers with batch-normalization and residual connection, the middle flow has three separable convolutions that are repeated eight times, and the exit flow has two layers of normalized convolutions. A 16x16x2048 vector of latent features is the result of the model's feature extractor. A Global Average Pooling (GAP) Layer is used to transmit this. This layer conducts dimensionality reduction by averaging the

16x16 values, resulting in a feature vector of 2048 length. A dropout layer is then applied to this feature vector, randomly setting 25% of the 2048 values to 0. The benefit of this dropout layer, as discussed in, is that the classifier is compelled to learn more reliable weights that are not impacted by losing random Neurons, helping to stabilize the network and lowering over-fitting.

After the dropout, the feature vector is passed through five dense, completely linked layers to extract more precise features. By learning the necessary parameters, the first dense layer reduces the feature vector to half its size, or from 1024 to 512. With each subsequent layer, the feature vectors' size is cut in half, from 512 to 256 to 128. This 64-length vector is fed into a multi-output configuration with nine outputs, each corresponding to one label. In contrast to all of the preceding layers, which were all activated by ReLu functions, each of these output layers uses a sigmoid function to combine the 64 attributes and output the probability of one label for each. The models employ the Adam optimizer, which computes an exponential moving average of the gradient and the squared gradient before decaying both at rates that can be manually adjusted[32]. Since the output was divided into nine probabilities between 0 and 1, we utilized a binary cross-entropy loss function to make the gradients decay according to errors. The batch size for each training was eight.
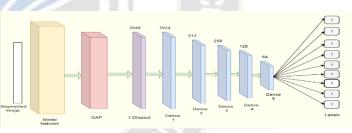

Fig 5.   Architecture employed for classification

Mobilenet V2:

In contrast to the traditional convolution, the revised MobileNet V2 design incorporates a "depth-wise-separable convolution" into each of its architecture's hidden layers. This allows for a significant reduction in the total number of variables as well as the generation of a lightweight neural-network. In place of the conventional convolution, the depth-wise-convolution employs a solitary filter and is preceded by a pointwise convolution that is referred to as a "depth-wise-seperable convolution". This new convolution method is intended to improve image quality.

MobileNetV2 is built on the concept of inverted residuals, which implies that rather than the conventional method of increasing the number of channels using 3x3 convolutional layers, the input and output of a residual block are enlarged and contracted using 1x1 convolutional layers. This minimises the

308

amount of parameters and computations, while preserving accuracy[33].

MobileNetV2 also features:
Linear bottlenecks In MobileNetV2, the non-linear activation function (ReLU) is applied after the depth-wise convolution to improve accuracy and decrease the amount of computations.
Expansion layer: In each inverted residual block, a 1x1 convolutional layer is utilised to increase the amount of nodes, followed by a depth wise convolutional layer to reduce the number of calculations.

MobileNet V2 architecture:
Among the most popular deep learning architectures for mobile devices is MobileNet V1, which is not only compact but also computationally effective, attaining great performance. The primary concept of MobileNet is that the process is broken into depthwise-separable 3X3 convolution filters followed by 1X1 convolution rather than utilizing standard 3X3 convolution filters. The new design does the same filtering with lesser steps and parameters and combination process as a normal convolution. In MobileNet V1, The amount of channels needed to be maintained or quadrupled for the pointwise convolution. The pointwise convolution in MobileNet V2 has the opposite effect: it reduces the number of channels. Because it converts data with many reducing the dimension while adding dimensions (channels) to a tensor, this layer has come to be known as the projection layer[35].
The extension layer is the first brand-new function introduced by MobileNet V2. A 1-1 convolution makes up the expansion layer. Before beginning the depth-wise convolution, its purpose is to expand the image data's channel count. Because of this, the expansion layer always has a greater number of output channels compared to input channels, which is opposed to the projection layer.
concentrate on key elements by adjusting the channel weights.
Numerous input/output sizes: MobileNetV2 is adaptable for a variety of applications thanks to its support for multiple input and output sizes[34].
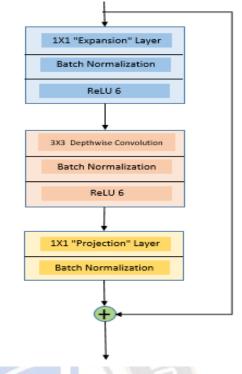

Fig 6. Architecture of MobileNet V2 Model

The residual connection shown in Figure is the second new feature of MobileNet V2. This functions similarly to the ResNeT and aids in the gradient flow through the network. The feature channels are expanded by a factor of t.
While the MobileNet V2 is primarily concerned with reducing latency, it also gives small networks the ability to function effectively and accept input of any size. This can result in improved performance due to the inclusion of batch normalization and ReLU6 as the activation function in every layer[33].

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 512, 512, 3)	0	
mobilenetv2_1.00_224 (Functiona	(None, None, None, 1	2257984	input_2[0][0]
global_average_pooling2d (Globa	(None, 1280)	0	mobilenetv2_1.00_224[0][0]
dropout (Dropout)	(None, 1280)	0	global_average_pooling2d[0][0]
dense (Dense)	(None, 1024)	1311744	dropout[0][0]
dense_1 (Dense)	(None, 512)	524800	dense[0][0]
dense_2 (Dense)	(None, 256)	131328	dense_1[0][0]
dense_3 (Dense)	(None, 128)	32896	dense_2[0][0]
dense_4 (Dense)	(None, 64)	8256	dense_3[0][0]
acute_and_chronic_pe (Dense)	(None, 1)	65	dense_4[0][0]
central_pe (Dense)	(None, 1)	65	dense_4[0][0]
chronic_pe (Dense)	(None, 1)	65	dense_4[0][0]
indeterminate (Dense)	(None, 1)	65	dense_4[0][0]
leftsided_pe (Dense)	(None, 1)	65	dense_4[0][0]
pe_present_on_image (Dense)	(None, 1)	65	dense_4[0][0]
rightsided_pe (Dense)	(None, 1)	65	dense_4[0][0]
rv_lv_ratio_gte_1 (Dense)	(None, 1)	65	dense_4[0][0]
rv_lv_ratio_lt_1 (Dense)	(None, 1)	65	dense_4[0][0]

Fig 7 Model of MobileNet V2

Xception Model:

Xception is a deep-convolutional-neural-network architecture with "Depthwise Separable Convolutions". Convolutional neural networks utilize Inception modules Following a depth-wise convolution comes a pointwise convolution, which serves as a transitional stage between the depth-wise-separable convolution process and the conventional convolution that comes after it. This makes a depth-wise separable convolution comparable to an Inception module with the most towers possible. Using this insight, they suggest a novel deep convolutional neural network structure that draws inspiration from Inception model, replacing Inception modules with depth-wise separable convolutions.[36].

By maintaining or even increasing the model's accuracy, this method drastically minimises the parameters and calculations that the network must perform. Xception can achieve equivalent or higher performance with lesser components than previous top of the line and latest models by swapping the ordinary convolution for the depth wise separable convolution operation. With a sequence of "convolutional-layers", "a GAP layer", and a "fully-connected layer" following, Xception has a hierarchical structure. The design consists of numerous blocks, each of which has multiple depth-wise separable convolutional layers. Depending on the needs of the application, the number of blocks and the amount of levels within each block can be modified.

The capacity of Xception to learn complicated features from incoming data is one of its main advantages, and this property makes it particularly beneficial for tasks like segmentation, object detection, and image categorization. Fine-grained picture recognition, medical image analysis, and scene understanding are just a few of the image-related applications that have used Xception.[36].

Across a number of benchmark datasets, including ImageNet, CIFAR-10, and CIFAR-100, Xception has also been demonstrated to outperform other cutting-edge CNN architectures. For other applications including speech recognition and natural language processing, deep learning models have also been developed using Xception.

In general, Xception is a robust and effective CNN architecture that is frequently utilised in a variety of applications involving images. It is an important tool in the field of deep learning due to its capacity to learn intricate features from input data and its exceptional performance on benchmark datasets.[37].

The success of Xception's two fundamental concepts makes for a very effective architecture:

Depth wise Separable Convolution

Shortcuts between Convolution blocks as in Reset[38].

XCeption provides a structure that is built with depth-wise separable convolution blocks + Maxpooling, all connected using shortcuts, similar to ResNet implementations.

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 512, 512, 3)	0	
xception (Functional)	(None, None, None, 2	20861480	input_2[0][0]
global_average_pooling2d (Globa	(None, 2048)	0	xception[0][0]
dropout (Dropout)	(None, 2048)	0	global_average_pooling2d[0][0]
dense (Dense)	(None, 1024)	2098176	dropout[0][0]
dense_1 (Dense)	(None, 512)	524800	dense[0][0]
dense_2 (Dense)	(None, 256)	131328	dense_1[0][0]
dense_3 (Dense)	(None, 128)	32896	dense_2[0][0]
dense_4 (Dense)	(None, 64)	8256	dense_3[0][0]
acute_and_chronic_pe (Dense)	(None, 1)	65	dense_4[0][0]
central_pe (Dense)	(None, 1)	65	dense_4[0][0]
chronic_pe (Dense)	(None, 1)	65	dense_4[0][0]
indeterminate (Dense)	(None, 1)	65	dense_4[0][0]
leftsided_pe (Dense)	(None, 1)	65	dense_4[0][0]
pe_present_on_image (Dense)	(None, 1)	65	dense_4[0][0]
rightsided_pe (Dense)	(None, 1)	65	dense_4[0][0]
rv_lv_ratio_gte_1 (Dense)	(None, 1)	65	dense_4[0][0]
rv_lv_ratio_lt_1 (Dense)	(None, 1)	65	dense_4[0][0]

Fig 8. Model of XCeption Model

EfficientNet B5 Model:

EfficientNet B5 is a convolutional neural network architecture that was introduced in the paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks" by Tan et al. in 2019[39].

EfficientNet B5's design is built on a scaling method that balances the model's depth, width, and resolution. As a result, the model is computationally effective while achieving cutting-edge accuracy. The following list of building blocks makes up the architecture:

Stem: After batch normalisation and activation, a convolutional layer is applied to the input image with a small kernel size. This helps to extract low-level information from the supplied image.[40].

Basic network: The base network is made up of a succession of repeated blocks, each with numerous layers. Using skip connections and squeeze-and-excitation modules, each block is composed of a mixture of depth-wise separable convolution and regular convolution. These building components aid in the extraction of more intricate details from the source image.[40].

Head: The output of the primary network is then subjected to a pooling-layer, a fully-connected layer, and a SoftMax-activation function. This helps assign the correct class as the destination for the characteristics extracted.[40].

EfficientNet B5 has a total of 30.5 million parameters, and it has been shown to achieve sophisticated and latest performance on several benchmark datasets, including ImageNet, COCO, and CIFAR-100[41].

_____

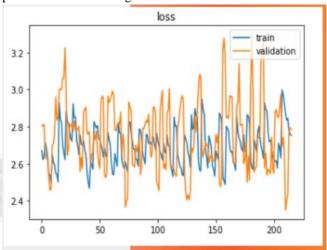| Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	[(None, 512, 512, 3)]	0	
efficientnetb5 (Functional)	(None, None, None, 2	28513527	input_2[0][0]
global_average_pooling2d (Globa	(None, 2048)	0	efficientnetb5[0][0]
dropout (Dropout)	(None, 2048)	0	global_average_pooling2d[0][0]
dense (Dense)	(None, 1024)	2098176	dropout[0][0]
dense_1 (Dense)	(None, 512)	524800	dense[0][0]
dense_2 (Dense)	(None, 256)	131328	dense_1[0][0]
dense_3 (Dense)	(None, 128)	32896	dense_2[0][0]
dense_4 (Dense)	(None, 64)	8256	dense_3[0][0]
acute_and_chronic_pe (Dense)	(None, 1)	65	dense_4[0][0]
central_pe (Dense)	(None, 1)	65	dense_4[0][0]
chronic_pe (Dense)	(None, 1)	65	dense_4[0][0]
indeterminate (Dense)	(None, 1)	65	dense_4[0][0]
leftsided_pe (Dense)	(None, 1)	65	dense_4[0][0]
pe_present_on_image (Dense)	(None, 1)	65	dense_4[0][0]
rightsided_pe (Dense)	(None, 1)	65	dense_4[0][0]
rv_lv_ratio_gte_1 (Dense)	(None, 1)	65	dense_4[0][0]
rv_lv_ratio_lt_1 (Dense)	(None, 1)	65	dense_4[0][0]

Fig 9. Model of Efficientnet B5

IV. EXPERIMENTS AND RESULTS

Because to the size of the competition's data—roughly 1 TB—and the inability of our systems to handle or train it, our experimentation pipeline was severely constrained to the metrics and time constraints of Kaggle kernels. The total amount of time spent in training for our model was about four or five hours of training time on Kaggle GPUs (40 hours of GPU usage per week), allowing us to train no more than 400 epochs in a single run. Here, the loss over-training and validation as well as the model's classification accuracy served as our measuring measures. a point in the direction of the next step, which is why the next step is to be based. For all labels, the accuracy in all experiments reached a value at the conclusion of training on the validation set. Since the issue was divided into nine binary classification problems, we chose to train our model utilizing the "Binary-Cross-Entropy loss" and assess the degree of agreement between the predicted labels and the labels in the validation set.
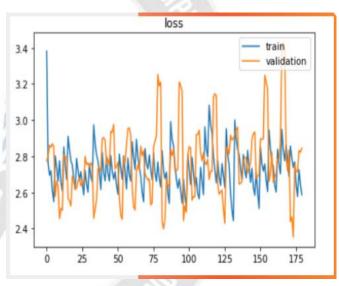
The three models were trained to equal epochs, and a comparison of the three was conducted to assess the models, determine which model is most effective for the task of PE detection, and construct the results appropriately.

Our starting point was the simple architecture discussed in part II, which performed admirably on the validation set, with an average loss of 2.166 and an average accuracy, equally weighted for all metrics, of roughly 94.7% across all 3 models. In order to improve accuracy and outcomes, we implemented a
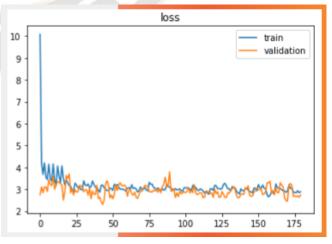
variety of modifications by adding a variety of segmentation processes to the training model.

a) *Fig 10.1 Loss Plot for MobileNet V2*

b) *Fig 10.2 Loss Plot for EfficentNet B5*

c) *Fig 10.3 Loss Plot for Xception Model*

311

Compared to the existing works, our suggested approach can reach higher accuracy. Also, it should be emphasized that our approach results in a very small MSE loss compared to what is shown in Fig. 10. The gap between our anticipated value and actual value is quite minimal thanks to the decreased MSE value. In each epoch, the MSE of. 10 changes. To solve the overfitting issue, it is crucial to identify the best epoch. To do this, the model is stored if the outcome obtained using an epoch is superior to the previous one. Evaluation of the set of test accuracy is used to determine whether or not the outcome is superior. We also take into account the natural change in MSE with epoch to combat the overfitting issue. Training is considered complete when the MSE loss of the testing curve has been stable after four iterations. Then, we take into account both the accuracy score and the MSE loss to guarantee that the overfitting problem is not influencing our approach. The foregoing discussion leads us to the conclusion that our technique is superior to existing ones in its ability to detect the severity of pulmonary embolism, as evidenced by its greater accuracy, lower MSE, and assurance that no overfitting occurs. From the table, In terms of precision, we find that EfficientNet performs best, followed by MobileNet V2. With less computational time and less loss, MobileNet V2 offers good accuracy.

		MOBILENET V2	EFFICIENTNET B5	XCEPTION
PE_PRESENT_ON_IMAGE	Accuracy	96	97	91
	Loss	0.1478	0.1302	0.3524
RV_LV_RATIO_GTE_1	Accuracy	86	89	85
	Loss	0.4122	0.3332	0.4473
RV_LV_RATIO_LT_1	Accuracy	79	80	78
	Loss	0.5339	0.5044	0.5482
CENTRAL_PE	Accuracy	90	95	92
	Loss	0.3487	0.2097	0.3045
CHRONIC_PE	Accuracy	91	94	92
	Loss	0.3303	0.2389	0.2949
INDETERMINATE	Accuracy	98	100	97
	Loss	0.0933	0.0275	0.1298
LEFTSIDED_PE	Accuracy	72	76	73
	Loss	0.6257	0.5795	0.6185
RIGHTSIDED_PE	Accuracy	74	73	69
	Loss	0.5752	0.5724	0.6541
ACUTE_AND_CHRONIC_PE	Accuracy	99	98	95
	Loss	0.0613	0.1261	0.2254

d) Tabel 2. Reading of the accuracy and Loss for all the labels of after using all three models .

V. CONCLUSION

In this piece, we presented a full-pipeline method for picking up PE from chest CTs. We worked with the model and the data that the RSNA provided. Using the Xception, MobileNet V2, and EfficientNet CNN will use AI to identify latent details in the scans, we then applied a classifier to predict the 9 labels of the classification job, resulting in an overall loss on the training and validation sets. These labels were predicted by breaking the multi-category classification problem down into 9 binary classification units, which were then used to assess the model using accuracy as a metric. On the validation test set, With this approach, we had a 92% accuracy rate. Although this is not yet up to par for direct application in the medical industry, we do present an architecture that can be improved upon to raise the caliber of the classifications. The two ways that additional enhancements can be made have been determined: First, the feature extractor can be combined with RNN units that treat the study's collection of images as a series to increase accuracy. Moreover, the data can be further pre-processed to improve the label balance.

REFERENCES

[1] A. K. Tarbox and M. Swaroop, "Pulmonary embolism," Int J Crit Illn Inj Sci, vol. 3, no. 1, p. 69, 2013, doi: 10.4103/2229-5151.109427.

[2] X. Yang et al., "SPECIAL SECTION ON DATA-ENABLED INTELLIGENCE FOR DIGITAL HEALTH A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images", doi: 10.1109/ACCESS.2019.2925210.

[3] M. L. Giger, "Machine Learning in Medical Imaging," Journal of the American College of Radiology, vol. 15, no. 3, pp. 512–520, Mar. 2018, doi: 10.1016/J.JACR.2017.12.028.

[4] D. Shen, G. Wu, and H.-I. Suk, "Deep Learning in Medical Image Analysis," 2017, doi: 10.1146/annurev-bioeng-071516.

[5] H. Khachnaoui, M. Agrebi, S. Halouani, and N. Khlifa, "Deep Learning for Automatic Pulmonary Embolism Identification Using CTA Images," in 2022 6th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), May 2022, pp. 1–6. doi: 10.1109/ATSIP55956.2022.9805929.

[6] R. Pillai, P. Oza, and P. Sharma, "Review of Machine Learning Techniques in Health Care," 2020, pp. 103–111. doi: 10.1007/978-3-030-29407-6_9.

[7] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in 2017 International Conference on Engineering and Technology (ICET), Aug. 2017, pp. 1–6. doi: 10.1109/ICEngTechnol.2017.8308186.

[8] R. Chauhan, K. K. Ghanshala, and R. C. Joshi, "Convolutional Neural Network (CNN) for Image Detection and Recognition," in 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Dec. 2018, pp. 278–282. doi: 10.1109/ICSCCC.2018.8703316.

[9] A. Esteva et al., "Deep learning-enabled medical computer vision", doi: 10.1038/s41746-020-00376-2.

[10] M. Lalli and S. Amutha, "Applications of Deep Learning and Machine Learning in Healthcare Domain-A Literature Review," International Journal of Electrical Engineering and Technology, vol. 11, no. 8, pp. 113–126, 2020, doi: 10.34218/IJEET.11.8.2020.011.

[11] M. Al-Ayyoub, G. Husari, O. Darwish, and A. Alabed-Alaziz, "Machine learning approach for brain tumor detection," ACM International Conference Proceeding Series, 2012, doi: 10.1145/2222444.2222467.

[12] G. D. Tourassi, C. E. Floyd, H. D. Sostman, and R. E. Coleman, "Acute pulmonary embolism: artificial neural network approach for diagnosis," Radiology, vol. 189, no. 2, pp. 555–558, 1993, doi: 10.1148/RADIOLOGY.189.2.8210389.

[13] N. Tajbakhsh, J. Y. Shin, M. B. Gotway, and J. Liang, "Computer-aided detection and visualization of pulmonary embolism using a novel, compact, and discriminative image representation," Med Image Anal, vol. 58, p. 101541, Dec. 2019, doi: 10.1016/j.media.2019.101541.

[14] X. Yang et al., "A Two-Stage Convolutional Neural Network for Pulmonary Embolism Detection From CTPA Images," IEEE Access, vol. 7, pp. 84849–84857, 2019, doi: 10.1109/ACCESS.2019.2925210.

[15] "Radiological Society of North America | RSNA." https://www.rsna.org/ (accessed Mar. 19, 2023).

[16] "RSNA STR Pulmonary Embolism Detection | Kaggle." https://www.kaggle.com/competitions/rsna-str-pulmonary-embolism-detection (accessed Mar. 19, 2023).

[17] W. D. Bidgood, S. C. Horii, F. W. Prior, and D. E. Van Syckle, "Understanding and Using DICOM, the Data Interchange Standard for Biomedical Imaging," Journal of the American Medical Informatics Association, vol. 4, no. 3, pp. 199–212, May 1997, doi: 10.1136/jamia.1997.0040199.

[18] Hounsfield and Godfrey Newbold, "Computed Medical Imaging," Journal of Computer Assisted Tomography 4, pp. 665–674, 1980.

[19] " ♎ Pulmonary Embolism Dicom preprocessing & EDA ♎ | Kaggle." https://www.kaggle.com/code/nitindatta/pulmonary-embolism-dicom-preprocessing-eda/notebook (accessed Mar. 19, 2023).

[20] S. NAKASU, T. ONISHI, S. KITAHARA, H. OOWAKI, and K. MATSUMURA, "CT Hounsfield Unit Is a Good Predictor of Growth in Meningiomas," Neurol Med Chir (Tokyo), vol. 59, no. 2, pp. 54–62, 2019, doi: 10.2176/nmc.oa.2018-0209.

[21] S. Kamalian, M. H. Lev, and R. Gupta, "Computed tomography imaging and angiography – principles," Handb Clin Neurol, vol. 135, pp. 3–20, Dec. 2016, doi: 10.1016/B978-0-444-53485-9.00001-5.

[22] C. Zhang, M. Sun, Y. Wei, H. Zhang, S. Xie, and T. Liu, "Automatic segmentation of arterial tree from 3D computed tomographic pulmonary angiography (CTPA) scans," 2019, doi: 10.1080/24699322.2019.1649077.

[23] C. Zhou et al., "Automatic pulmonary vessel segmentation in 3D computed tomographic pulmonary angiographic (CTPA) images," Mar. 2006, p. 61444Q. doi: 10.1117/12.655343.

[24] K. U. Ahamed et al., "A deep learning approach using effective preprocessing techniques to detect COVID-19 from chest CT-scan and X-ray images," Comput Biol Med, vol. 139, p. 105014, Dec. 2021, doi: 10.1016/J.COMPBIOMED.2021.105014.

[25] M. R. Islam and M. Nahiduzzaman, "Complex features extraction with deep learning model for the detection of COVID19 from CT scan images using ensemble based machine learning approach," Expert Syst Appl, vol. 195, p. 116554, Jun. 2022, doi: 10.1016/J.ESWA.2022.116554.

[26] Y. Kimori, "Synchrotron Radiation Morphological image processing for quantitative shape analysis of biomedical structures: effective contrast enhancement," J. Synchrotron Rad, vol. 20, pp. 848–853, 2013, doi: 10.1107/S0909049513020761.

[27] "Erosion and Dilation in Digital Image Processing – Buzztech." https://buzztech.in/erosion-and-dilation-in-digital-image-processing/ (accessed Mar. 19, 2023).

[28] H. Hassanpour, N. Samadiani, and S. M. Mahdi Salehi, "Using morphological transforms to enhance the contrast of medical images," The Egyptian Journal of Radiology and Nuclear Medicine, vol. 46, no. 2, pp. 481–489, Jun. 2015, doi: 10.1016/J.EJRNM.2015.01.004.

[29] S. Dai, K. Lu, J. Dong, Y. Zhang, and Y. Chen, "A novel approach of lung segmentation on chest CT images using graph cuts," Neurocomputing, vol. 168, pp. 799–807, Nov. 2015, doi: 10.1016/J.NEUCOM.2015.05.044.

[30] A. Saood and I. Hatem, "COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet," BMC Med Imaging, vol. 21, no. 1, pp. 1–10, Dec. 2021, doi: 10.1186/S12880-020-00529-5/FIGURES/5.

[31] J. Hamwood, B. Schmutz, M. J. Collins, M. C. Allenby, and D. Alonso-Caneiro, "A deep learning method for automatic segmentation of the bony orbit in MRI and CT images," Scientific Reports 2021 11:1, vol. 11, no. 1, pp. 1–12, Jul. 2021, doi: 10.1038/s41598-021-93227-3.

[32] D. P. Kingma and J. Lei Ba, "ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION".

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks".

[34] C. Mahanty, · Raghvendra Kumar, · S Gopal, K. Patro, and R. Kumar, "Internet of Medical Things-Based COVID-19 Detection in CT Images Fused with Fuzzy Ensemble and Transfer Learning Models · CT images · Pneumonia · Transfer learning · SqueezeNet · DenseNet-201 · MobileNetV2 · Trainable ensemble · Sugeno fuzzy integral," New Gener Comput, vol. 40, pp. 1125–1141, 2022, doi: 10.1007/s00354-022-00176-0.

313

_____

[35] J. Manokaran, F. Zabihollahy, A. Hamilton-Wright, and E. Ukwatta, "Detection of COVID-19 from chest x-ray images using transfer learning," Journal of Medical Imaging, vol. 8, no. S1, Aug. 2021, doi: 10.1117/1.JMI.8.S1.017503.

[36] E. Alshehri, M. Kalkatawi, F. Abukhodair, K. Khashoggi, and R. Alotaibi, "COVID-19 Diagnosis from Medical Images Using Transfer Learning," Saudi Journal of Health Systems Research, vol. 2, no. 2, pp. 54–61, Feb. 2022, doi: 10.1159/000521658.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," pp. 248–255, Mar. 2010, doi: 10.1109/CVPR.2009.5206848.

[38] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jul. 2017, pp. 1800–1807. doi: 10.1109/CVPR.2017.195.

[39] M. Tan and Q. V Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks".

[40] "Complete Architectural Details of all EfficientNet Models | by Vardan Agarwal | Towards Data Science." https://towardsdatascience.com/complete-architectural-details-of-all-efficientnet-models-5fd5b736142 (accessed Mar. 19, 2023).

[41] S. Poudel and S. W. Lee, "Deep multi-scale attentional features for medical image segmentation," Appl Soft Comput, vol. 109, p. 107445, Sep. 2021, doi: 10.1016/J.ASOC.2021.107445.