Lung Nodule Segmentation and Classification using U-Net and Efficient-Net

Suriyavarman S[1], Dr. Arockia Xavier Annie .R[2]
Department of Computer Science Engineering-College of Engineering, Anna University, Chennai, India[1, 2]

***Abstract*—The ability to detect lung cancer has led to better health outcomes. Deep learning techniques are widely used in the medical field to detect lung tumors at an early stage. Deep learning models such as U-Net, Efficient-Net, Resnet, VGG-16, etc. have been incorporated in various studies to detect lung cancer accurately. To enhance the detection performance, this work proposes an algorithm that combines U-Net and Efficient-Net neural networks for lung nodule segmentation and classification. A feature-extraction-based semi-supervised method is used to take advantage of the huge amount of CT scan images with no pathological labels. Semi-supervised learning is achieved using a feature pyramid network (FPN) with ResNet-50 model for feature extraction and a neural network classifier for predicting unlabelled nodules. The main innovation of U-Net is the skip-connections, which give the decoder access to the features that the encoder learned at various scales and enable accurate localization of lung nodules. Efficient-Net uses depth, width, and resolution scaling, combined with a compound coefficient that uniformly scales all network dimensions, resulting in an efficient neural network for image classification. This work has been evaluated on the publicly available LIDC-IDRI dataset and outperforms most existing methods. The proposed algorithm aims to address issues such as a high false-positive rate, small nodules, and a wide range of non-uniform longitudinal data. Experiment results show this model has a higher accuracy of 91.67% when compared with previous works.**

***Keywords*—Cancer; CT; U-Net; efficient-net; feature; accuracy**

I. INTRODUCTION

Lung cancer is the second most lethal type of cancer in both men and women. Additionally, the coronavirus pandemic has increased the chances of mortality for lung cancer patients [1]. If not timely detected, lung tumor could lead to death. The small cell growths in the lung known as pulmonary nodules can be either cancerous (malignant) or noncancerous (benign) [2]. Given how closely early-stage cancer lung nodules resemble noncancerous nodules, a differential diagnosis based on the nodule's locations, morphological traits, and clinical biomarkers is required. According to the WHO, 2.21 million cases of lung cancer were reported in 2020, and 1.80 million lives were lost from it [3].

Lung cancer develops when cells in the lung tissue continue to grow uncontrollably. This results in tumor growth. They have the potential to spread to numerous other body parts and affect respiration [4]. A number of imaging techniques are used, including computed tomography, sputum cytology, chest X-rays, and magnetic resonance imaging, to detect lung cancers early (MRI). Tumors are briefly categorised into two types in order to be detected: i) Benign tumors (non-cancerous), and ii) Malignant tumors (cancerous) [5]. Image processing techniques can improve manual analysis. Many studies have demonstrated the occurrence of diagnostic errors in clinical practise [6]. These errors can be attributed to a variety of contributing factors, which can be broadly categorised as person-specific, nodule-specific, and environment-specific problems.

Fine-grained cross-sectional images of the human body are produced via a CT scan. A CT scan gathers many images as opposed to a standard x-ray, which only records one or two images. These images are eventually merged by a computer to create a slice of the body part being studied. A conventional chest x-ray is less likely to find lung cancer than a CT scan [7]. It can also show the size, shape, and location of any lung tumours and reveal enlarged lymph nodes that may house cancer that has spread [8].

Recent years have seen significant advancements in the fields of medicine, including lesion classification, tissue detection, and segmentation, owing to machine learning and particularly deep learning [9]. With the development of artificial intelligence, deep learning has become more popular in the analysis of medical images. A practical method for extracting richer and more powerful characteristics is provided by deep convolutional neural networks [10]. Automated algorithms can provide a faster and more accurate solution for the detection and classification of lung nodules, thus improving patient outcomes [11].

II. LITERATURE REVIEW

A. Preprocessing

Large amounts of unprocessed data from CT scans can contain noise, artefacts, and inconsistencies that may compromise the accuracy and reliability of subsequent analyses. In their work [12], Amalorpavam et al. present a very detailed systematic review of existing morphological operations in digital image processing. The primary goal of this work is to change images using mathematical morphology in order to normalize them for the intended analyses. This work shows many approaches, such as erosion, dilation, thresholding, and blurring, can be effectively incorporated in the pre-processing of CT scan images. These methods are used to enhance the contrast between various tissues, remove noise from the images, and remove any non-nodule structures.

B. Segmentation

Developed on the foundation of a fully convolutional neural network, U-Net is a semantic segmentation network. Olaf et al. [13] successfully modelled the convolutional network-based U-Net architecture, which was created specifically for the segmentation of images in the biomedical field. This work includes a traditional U-Net architecture with an encoder and a decoder and produced an average IOU ("intersection over union") of 92%.

A study by Chen et al. [14] proposed a U-Net-based method for segmenting lung nodules in CT images. The proposed method consisted of two stages: the first stage used a pre-trained U-Net model for initial segmentation, and the second stage used a 3D fully connected conditional random field (CRF) for refinement. For the segmentation of lung nodules, Ronneberger et al. [15] developed a U-Net network based on FCN. When segmenting medical images with hazy boundaries, U-Net combines low- and high-resolution information by skipping connections. Low-resolution information is used for target identification, and high-resolution information is used for localising segmentation.

C. Feature Extraction

In a feature pyramid network, a convolutional neural network (CNN) is used to extract features from an input image and create a feature map. Xiaolong Wang et al. [16] proposed a semi-supervised learning method that uses a multi-level feature pyramid network (FPN) to leverage both labelled and unlabeled data. The FPN is used to extract features at different scales, and the method incorporates consistency regularisation to encourage the model to produce similar predictions on labelled and unlabeled data. With only 4,000 labelled samples, the suggested method achieves the best reported classification error rate of 4.42% on the test set. With 4,000 labelled samples and 50,000 unlabeled samples, it also achieves an error rate of 3.57%, which is also the best result to date.

Guangrui Mu et al., [17] proposed Feature pyramid networks with Relu Cascade for CT Pulmonary Nodule Detection. A detection network is first trained with few positive annotations (nodules) and randomly selected negative samples (background). In FPN, extraction is performed separately to produce feature maps of four different scales by using images onto each feature map. The "Relu cascade's" uniqueness lies in the method used to link these networks together in a cascade.

D. Classification

Convolutional neural network architecture called Efficient-Net has produced cutting-edge outcomes on a number of computer vision tasks, including image classification. Agrawal et al., [18] presented various deep learning based pre-trained CNN techniques for distinguishing benign and malignant brain tumor images. They used different optimizers to complete the tasks, namely Adam, RMSprop, and stochastic gradient descent (SGD). Their research demonstrated that a fine-tuned Alexnet could perform particularly well on challenges involving medical imaging. They used data pre-processing and augmentation techniques for boosting diversity in the data samples to decrease the overfitting of the previous models.

Hwejin Jung et al., [19] proposed a three-dimensional deep convolutional neural network (3D DCNN) with dense connections and shortcut connections was developed for the classification of lung nodules. By enabling the gradient to move swiftly and directly, shortcut connections and dense connections successfully address the gradient vanishing problem. This method achieved a highest CPM score of 0.910. Divya et al., [20] presented methods based on transfer learning that enhances current architecture in multi-class classification by relying on pretrained DCNN trained on ImageNet dataset. Pretrained weights of EfficientNetV2-B0 are transferred as initial weights and to distinguish between benign and malignant tissue samples in tumour cells and classify them, the model is fine-tuned.

The objective of this work is to create and construct a deep learning model based on the U-Net segmentation network and Efficient-Net architecture to segment and classify lung nodules in CT scan images. The work addresses various limitations, such as that improvements are needed in the detection rates of lung nodules, as the proposed models have only focused on training dataset image samples and also segmented anomalies as positive samples, an unequal number of benign and malignant lung nodules, resulting in an imbalanced dataset, an accurate diagnosis of the pathological type of lung cancer is crucial for effective treatment; and the performance of model relies heavily on the choice of optimizer.

Developed on the foundation of a fully convolutional neural network, U-Net is a semantic segmentation network. There are a total of 23 layers in the network, which is significantly fewer than the number of layers in other networks while maintaining accuracy [21]. Feature pyramid networks are used to extract feature vectors from segmented images. These feature vectors are used as input into a neural network classifier for an unlabelled nodule prediction task [22]. A family of convolutional neural networks called Efficient-Net was created to achieve cutting-edge performance while being computationally efficient. It accomplishes this by balancing network depth, width, and resolution in order to optimise performance within a predetermined computational budget [23].

The proposed algorithm will be put to the test on the LIDC dataset to see if it can accurately classify and identify specific features, and the outcomes will be compared to standard metrics like accuracy, recall, and F1-score [24]. The proposed algorithm aims to detect and segment lung nodules accurately. Using an ablation study, the project aims to examine the impact of various parameters (hyper-parameters) on the performance of the proposed lung nodule segmentation and classification algorithm [25].

To explore the possibility of employing semi-supervised learning and transfer learning to enhance the performance of the proposed algorithm on datasets with limited labelled data [26]. Adam optimizer is incorporated to stabilize the performance of the model and reduce the effect of errors [27].

III. Proposed Methodology

A. LIDC-IDRI: The Lung Image Database Consortium

The National Cancer Institute spearheaded the collection of the 1018 cases that make up the LIDC dataset, each of which was collaboratively labelled by four radiologists. The

LIDC-IDRI dataset (Armato et al., 2011; Armato III et al., 2015b; Clark et al., 2013) in the Cancer Imaging Archive (TCIA) contains 1018 clinical chest CT scans with lung nodules obtained from seven institutions. The locations of the nodules and the nine sematic attributes of subtlety, sphericity, internal structure, margin, lobulation, spiculation, and malignancy are detailed in an associated XML file for each CT scan that was evaluated for annotation by up to four radiologists (Qin et al., 2019).

It is a database that is widely utilised and devoted to the advancement of current methods for lung nodule segmentation. Three categories of nodules are distinguished: non-nodules (size >=3mm), nodules (size < 3mm), and nodules (size >=3mm). All of the images will be in the DICOM (Digital Imaging and Communications in Medicine) format and 512 x 512 in size. Based on the malignancy details provided in the metadata csv and xml files, around 12000 labelled CT scan images and 3000 unlabelled CT scan images are used.

B. System Architecture

This section provides the detailed architecture and description and algorithmic details of the various subsystems in the proposed system. The detailed architecture of the proposed system is represented in the below Fig. 1. An algorithm has been proposed to segment the lungs in images by using U-Net model and classify the images using Efficient-Net architecture. Unlabelled dataset is labelled by using FPN with Resnet-50 as backbone. FPN extracts features from labelled images and provides these features as input into a neural network classifier. The classifier predicts the class of unlabelled images and these images are added to the dataset for model training.

This model typically involves a series of preprocessing steps, such as Gaussian blur, thresholding, erosion and contouring, to enhance and segment the objects in the images. The Gaussian blur function smoothens the image and reduces noise, while thresholding converts the image to black and white, simplifying it for further processing.

Contouring then isolates the objects within the image by drawing lines around them. The model uses the FPN architecture to extract features from the images when preprocessing is completed. The Efficient-Net model is a state-of-the-art deep learning architecture that has achieved exceptional results in image classification tasks, making it a powerful tool for object recognition.

To achieve this goal, the system will use the U-Net model to segment the lung nodules from the surrounding tissue and the efficient-net architecture to classify the nodules as benign or malignant. The feature extraction is performed using Feature Pyramid networks. The system will also have the potential to be used in other medical imaging modalities, such as MRI and PET, for the detection and classification of other types of abnormalities and diseases.

C. Module Design

The architecture is divided into five modules. The modules are namely preprocessing, augmentation, segmentation, feature extraction and nodule classification, respectively.

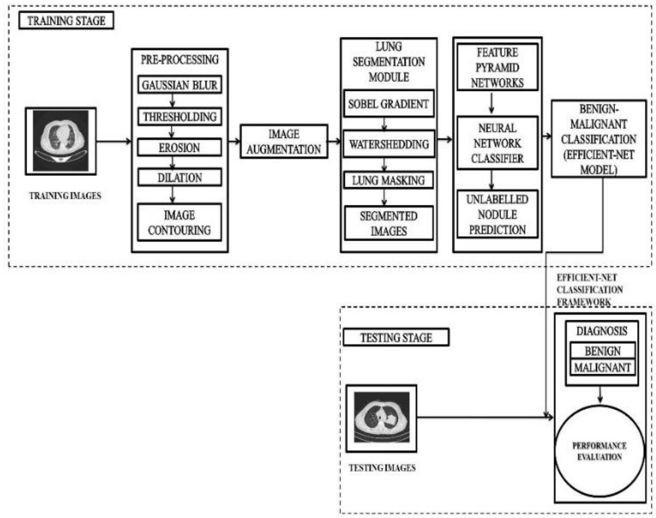

Fig. 1. Detailed system architecture.

1) Pre-Processing: The data pre-processing [28] step is essential to normalise the data in a way that enables the convolutional network to learn appropriate and meaningful features properly because CT scans might be acquired by different scanners in different medical clinics without the use of identical acquisition protocols. Directly obtained CT scans of the lungs have noise, no discernible variations in the greyscale border, and other characteristics that make them difficult to divide. Consequently, the Images Pre-processing is first task. To reduce the noise in the lung CT scans, it is necessary to remove certain regions that make it difficult to separate.

Meanwhile, to make the upcoming segmentation process easier, the boundary portion of the CT picture that changes smoothly needs to be sharpened. Fig. 2 represents the Preprocessing Module which consists of Gaussian Blur, Thresholding, Erosion, Dilation and contouring Techniques.

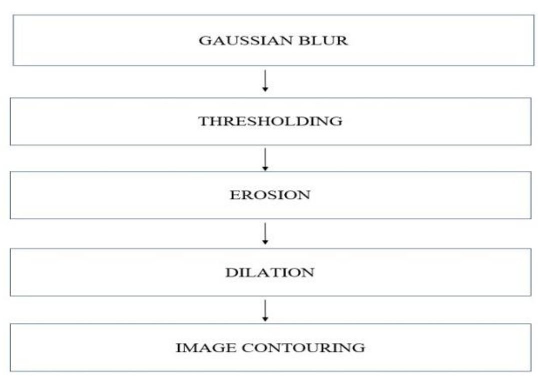

Fig. 2. Preprocessing flow diagram.

a)Gaussian Blur: Gaussian blur [29] is a linear low-pass filter, where the pixel value is calculated using the Gaussian function. The rate at which this weight diminishes is determined by a Gaussian function, hence the name Gaussian blur.

$$G(x, y) = (1 / (2\pi\sigma^2)) * \exp(-(x^2 + y^2) / (2\sigma^2)) \quad (1)$$

Where,

x → X coordinate value,

y → Y coordinate value,

π → Mathematical Constant PI (value = 3.13),

σ → Standard Deviation,

exp → exponential function

Algorithm 1: Gaussian Blur

Input: Image I with dimensions m x n, Gaussian kernel
Output: Image O with dimensions m x n

```
Normalize G, G = G / ∑∑G(x,y)
Pad the input image I with zeros to avoid boundary effects.
Amount of padding = (k-1)/2 on each side.
For each pixel (i,j) in the input image I:
    Initialize the output pixel O(i,j) to zero.
    For each kernel element (x,y):
        Compute the pixel coordinates (a,b):
            a = i - (k-1)/2 + x
            b = j - (k-1)/2 + y
        Add the product of the input pixel I(a,b) and the
        kernel element G(x,y) to the output pixel O(i,j):
            O(i,j) = O(i,j) + I(a,b) * G(x,y)
        End
    End
Clip the output image O to the range [0,255]
Return the output image O
```

b)Thresholding: The simplest thresholding [30] methods replace each pixel in an image with a black pixel if the image intensity is less than a fixed value called the threshold or a white pixel if the pixel intensity is greater than that threshold.

Algorithm 2: Thresholding

Input: Image I with dimensions m x n, Threshold value
Output: Image O with dimensions m x n

```
For each pixel (i,j) in the input image I:
        If the pixel value I(i,j) is lesser than or equal to the
        threshold value (255):
            set the output pixel O(i,j) to 255 (white)
        End
        Else:
            The output pixel O(i,j) remains the same
        End
    End
Return the thresholded output image O
Return the output image O
```

c)Erosion: Erosion is a morphological image processing technique that is used to remove small details or isolated pixels from an image.

Algorithm 3: Erosion

Input: Image I with dimensions m x n, structuring element B with dimensions k x k
Output: Image O with dimensions m x n

```
Pad the input image I with zeros to avoid boundary effects.
Amount of padding = (k-1)/2 on each side
For each pixel (i,j) in the input image I:
        Initialize a flag variable f to true
        For each structuring element element (x,y):
            Compute the pixel coordinates (a,b):

                a = i - (k-1)/2 + x                  (2)

                b = j - (k-1)/2 + y                  (3)
            If (I(a,b) == 0 & B(x,y) ==1):
                set f to False.
            End

            If (f == True):
                set the output pixel O(i,j) to 1
            End
            Else:
                set the output pixel O(i,j ) to 0.
            End
        End
    End
Remove the padding from the output image O
Return the output image O
```

d)Dilation: Dilation is a morphological image processing technique that is used to enlarge or extend the details in an image.

Algorithm 4: Dilation

Input: Image I with dimensions m x n, structuring element B with dimensions k x k

Output: Image O with dimensions m x n

```
Pad the input image I with zeros to avoid boundary effects.
Amount of padding = (k-1)/2 on each side
For each pixel (i,j) in the input image I:
    Initialize a flag variable f to false
    For each structuring element element (x,y):
            Compute the pixel coordinates (a,b):

                a = i - (k-1)/2 + x                  (3)

                b = j - (k-1)/2 + y                  (4)
            If (I(a,b) == 0 & B(x,y) ==1):
                set f to True.
            End
            If (f == True):
                set the output pixel O(i,j) to 1
            End
            Else:
                set the output pixel O(i,j ) to 0.
            End
        End
```

```
    End
Remove the padding from the output image O
Return the output image O.
```

e) Contouring: Contours are defined as the line joining all the points along the boundary of an image that are having the same intensity.

Algorithm 5: Contouring

Input: Image I with dimensions m x n
Output: Image O with dimensions m x n

```
For each pixel (i,j) in the input image I:
    Compute the biggest contour
        biggest_contour = max(contours, contourArea)
    Compute the extreme points
        leftmost=(image,8,(0, 0, 255),-1)
        rightmost=(image,8,(0, 255, 0),-1)
        topmost=(image,8, (255, 0, 0),-1)
        bottommost=(image,8,(255,255,0),-1)
    End
Draw the contour and extreme points in the image I
Crop the image
Return the cropped image O
```

2) Augmentation: Image data augmentation is a technique that can be used to artificially expand the size of a training dataset by creating modified versions of images in the dataset. The label for all the images will be the same and that is of the original image which is used to generate them.

Flipping the images vertically and horizontally can help the model become more robust to variations in the orientation of the patient within the CT scan. This is because CT scans can be taken in different positions, and flipping the images can simulate these different positions and help the model generalize better to new data.

3) U-Net segmentation: In the field of biomedicine, image segmentation applications frequently employ the U-Net architecture. It has two sides: one that contracts and the other that expands symmetrically. Convolutional, Rectified Linear Unit (ReLU), dropout, and pooling layer portions make up the contracting side. In down sampling, the number of feature channels is increased by a factor of two. Sections of the up sampling (transpose convolution) layer, concatenation with the relevant feature channel from the contracting side, convolutional layers, and dropout layer are included in the expansive side.

The network is able to get the required features from the relevant layer using the connections between the contracting and expanding sides. Every area on the expanded side sees a halving of the amount of feature channels. At each section on the expansive side, the number of feature channels is halved. The feature vector is mapped to the expected number of classes in the final convolution layer.

In U-Net, the encoder network is the contracting side of the architecture. It consists of convolutional and max-pooling layers that gradually reduce the spatial resolution of the input image while increasing the number of feature channels. This contraction path is designed to capture the contextual information and high-level features of the image. On the other hand, the decoder network is the expansive side of the architecture. It consists of up-sampling and convolutional layers that gradually increase the spatial resolution of the feature maps while reducing the number of feature channels.

This expansion path is designed to reconstruct the high-resolution details of the image and refine the segmentation mask. The encoder and decoder networks are connected by a bottleneck layer that preserves the high-resolution features of the image. The skip connections between the encoder and decoder networks help to combine the high-level features from the encoder with the low-level features from the decoder, resulting in a more accurate segmentation.

4) Feature extraction: A well-versed method in computer vision for various tasks like object detection, segmentation, and classification is feature extraction using Feature Pyramid Networks (FPN) with ResNet as the backbone. FPN is a feature extraction network that enhances the performance of a CNN by utilizing multi-scale feature maps, allowing the network to identify objects at different scales. ResNet, on the other hand, is a deep residual network architecture that has been shown to outperform traditional CNN architectures by allowing for deeper network depths without encountering the vanishing gradient problem.

When used as the backbone for FPN, ResNet provides a strong base for feature extraction and is capable of detecting high-level features in the input image. The FPN architecture takes feature maps from ResNet's intermediate layers and combines them to form a pyramid of multi-scale feature maps.

The top-down pathway in FPN involves up sampling the feature maps of lower resolutions and then combining them with the feature maps of higher resolutions.

Unlabelled nodule prediction:

The FPN is designed to address the problem of detecting objects of various scales in an image. The backbone network is typically a deep convolutional neural network such as ResNet, which is used to extract feature maps from the input image.

- The input image is passed through the first convolutional layer with 64 filters and a kernel size of 5x5.
- The output of the first layer is passed through a batch normalization layer, a ReLU activation layer, and a max pooling layer with a kernel size of 3x3 and stride of 2.
- The output of the first stage is passed through the second stage, which consists of 3 residual blocks with 256 filters and a kernel size of 3x3.

- The output of the second stage is passed through the third stage, which consists of 4 residual blocks with 512 filters and a kernel size of 3x3.
- The output of the third stage is passed through the fourth stage, which consists of 6 residual blocks with 1024 filters and a kernel size of 3x3.
- The output of the fourth stage is passed through the fifth stage, which consists of 3 residual blocks with 2048 filters and a kernel size of 3x3.
- The output of each stage is then passed through a feature pyramid network (FPN) module to create a feature pyramid. The feature pyramid is then used for Classification or other downstream tasks.

Classifier:

The classifier model is trained using two different class numpy array files, which are typically generated from a training dataset. One array file contains the extracted feature vectors of benign images, while the other array file contains the extracted feature vectors of malignant images.

5) Nodule classification: Efficient Net [31] is a convolutional neural network architecture and scaling method that uniformly scales all dimensions of depth/width/resolution using a *compound coefficient*. Unlike conventional practice that arbitrary scales these factors, the Efficient Net scaling method uniformly scales network width, depth, and resolution with a set of fixed scaling coefficients.

Efficient net architecture provides compound scaling method (scaling all depth, width, and resolution dimensions) can help the model achieve the greatest accuracy gains. The baseline network has a significant impact on how well models scale. An extensive range of image classification tasks can be handled by Efficient Net.

It is a good model for transfer learning because of this feature. Efficient-Net performs Compound Scaling - that is, scale all three dimensions (depth, width, image resolution) while maintaining a balance between all dimensions of the network. The architecture of Efficient-Net consists of different Convolution and MBConvultion blocks interconnected together to produce feature maps of images.

MBConv:

This developed architecture uses the mobile inverted bottleneck convolution (MBConv). Then two more ideas are borrowed from MobileNet-V2 (which is a second improved version of MobileNet) including inverted residual connections, Linear bottlenecks. Mobile inverted bottleneck convolution (MBConv) is the main building block of Efficient-Net model family. MBConv is based on concepts borrowed from the MobileNet models.

Operations:

- The first operation is a linear transformation, which is performed on the input features using the weights of the first layer.
- The input features X are multiplied by a weight matrix W1, which has dimensions (input_dim, units). The result of this multiplication is a matrix of activations with dimensions (batch_size, units).
- The next operation is the application of a non-linear activation function to the activations.
- The process is repeated for subsequent layers, with each layer applying a new linear transformation to the activations of the previous layer, followed by a non-linear activation function.
- The final layer has a single unit and uses the sigmoid activation function, which maps the activations to a value between 0 and 1, representing the predicted probability of belonging to the class.
- During training, the model uses the back propagation algorithm to compute the gradients of the loss with respect to the weights of each layer, which are used to update the weights during the optimization process.
- These gradients are computed using the chain rule of calculus, which involves computing the derivative of the loss with respect to the output of each layer, and then propagating these derivatives backwards through the layers of the network.
- The model can learn to predict the correct label for each input sample, based on the features in the input data.

IV. Results and Discussions

The project was executed using python language and Anaconda Jupyter tool. Python is among the most widely used programming languages used for deep learning due to its flexibility and large number of available libraries, such as TensorFlow, PyTorch, Keras, and Scikit-learn. Python programming language's distribution Anaconda includes a collection of commonly used data science packages and tools. Jupyter is included as part of the Anaconda distribution. GPUs are designed to handle massive amounts of parallel processing, making them ideal for training complex neural networks.

The handling of large datasets, which is necessary for training our deep learning models, is made possible by the 16GB of memory. Running deep learning software and tools like TensorFlow and PyTorch requires a stable and user-friendly environment, which the Windows operating system offers.

The U-Net was given a collection of pre-processed CT images to train the model on. The weights were initialised with normal initialization and the ReLU activation function. The learning rate was set at 0.001 and the batch size to 32.

The binary cross-entropy loss function was utilised to calculate the loss function. In comparison, 80% of the data are utilised for training and 20% for validation. The model was trained for 250 epochs. The average dice score and best dice score resulted in values of 0.4273 and 0.5009. The average loss of the model was found to be 0.27. The nodule segmented using U-Net is represented in Fig. 3.

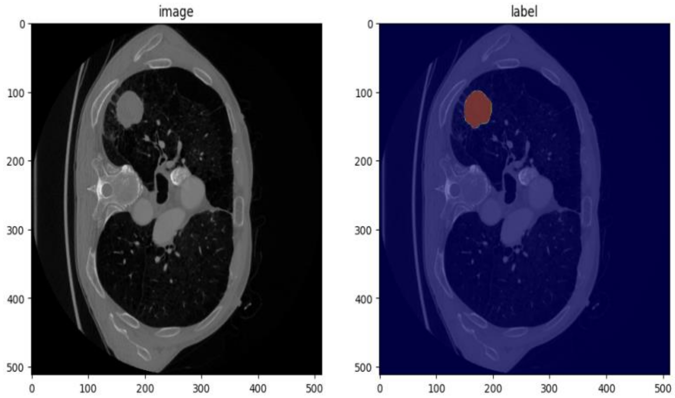

Fig. 3. Nodule segmented image.

An alternative to the stochastic gradient descent approach for deep learning models is the Adam optimizer. This approach handles sparse gradients on noisy situations and provides optimisation by integrating the key characteristics of AdaGrad and RMSProp. ReLU function resists simultaneously stimulating all neurons since it is a non-linear activation function.

In order to get a good result, the suggested U-Net model uses ReLU activation. The neuron will become inactive if the outcome of the linear transformation is less than 0. Results were satisfactory when the ReLU activation function was used with the Adam optimizer.

The training vs validation accuracy graph is a plot that shows the accuracy of a model on both the training and validation datasets. In Fig. 4, the blue line represents the training accuracy, while the red line represents the validation accuracy. The graph shows that the model is improving its accuracy on both the training and validation datasets from the first few epochs.

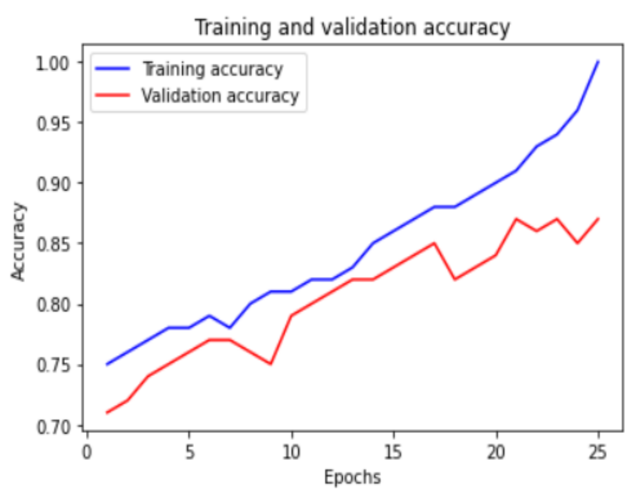

Fig. 4. Training vs validation accuracy graph.

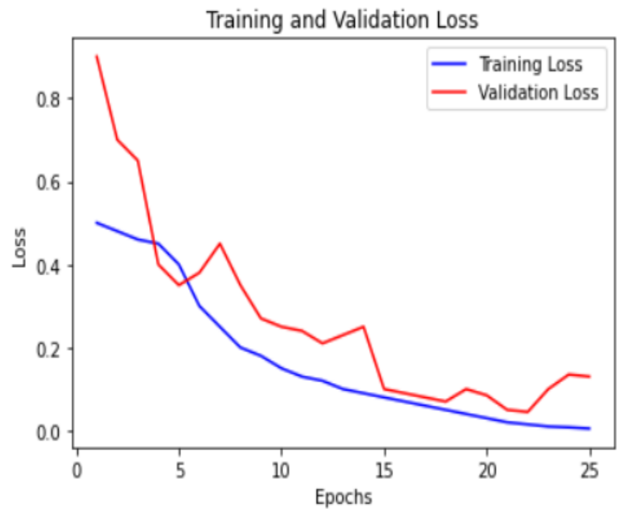

Fig. 5. Training vs validation loss graph.

The training vs validation loss graph is a plot that shows the loss of a model on both the training and validation datasets during the training and validation processes. During the training process, the model is trained on a training dataset to minimize the loss function. In Fig. 5, the blue line represents the training loss, while the red line represents the validation loss.

Accuracy (Table I) measures the percentage of correctly classified instances. It is the most basic metric and gives a good overall measure of the model's performance.

TABLE I. MODEL ACCURACY EVALUATION

S. No.	Epochs	No. of Images	Batch Size	Activation Function	Test Accuracy
1	25	13500	32	RELU	81.33%
2	25	16000	32	RELU	91.67%
3	25	16000	64	SWISH	83.70%
4	100	16000	32	RELU	91.34%

The False Positive Rate (FPR) is the ratio of all the benign nodules that are falsely identified as malignant nodules. The False Negative Rate (FNR) it is the ratio of all the malignant nodules that are incorrectly identified as benign nodules.

TABLE II. RESULTS OF DIFFERENT PARAMETERS

S. No	True Negative	True Positive	False Negative	False Positive	FPR	FNR	DR	Recall
1	1447	1658	191	101	0.06	0.10	0.89	0.86
2	1627	1150	222	398	0.19	0.16	0.83	0.74
3	1920	1227	127	73	0.03	0.09	0.86	0.85
4	1087	1594	251	465	0.29	0.13	0.81	0.76

Detection Rate (DR) is the ratio of benign nodules that are correctly identified as malignant nodules and vice-versa. Table II represents the different parameters calculated using the classification results.

V. Conclusion and Future Works

The proposed model for lung nodule segmentation and classification using U-Net and Efficient-Net neural network has shown promising results in detecting and classifying lung nodules accurately and efficiently. The proposed algorithm outperforms state-of-the-art methods in terms of accuracy, recall, loss. The ablation study conducted in this project revealed the contribution of each component of the proposed algorithm to the overall performance.

Furthermore, the proposed algorithm has the potential to be applied to larger datasets and to be optimized for real-time processing. The successful implementation of this algorithm can potentially improve the accuracy and efficiency of lung cancer screening programs and contribute to the early detection and treatment of lung cancer. However, the proposed algorithm still has limitations, such as the need for large annotated datasets and the potential for over fitting.

The proposed algorithm can be improved by incorporating explainable artificial intelligence (XAI) techniques to enhance the interpretability and transparency of the algorithm. The inclusion of XAI techniques can enable the identification of the features and patterns that the algorithm uses to make decisions, thus providing insights into the reasoning behind the algorithm's output. This can be particularly valuable for medical applications, where the decision-making process needs to be transparent and understandable to clinicians and patients.

The proposed algorithm can be optimized for multimodal imaging, such as computed tomography (CT) and positron emission 40 tomography (PET) scans, to improve the accuracy of lung nodule segmentation and classification. Additionally, the proposed model can be extended to address other types of lung cancers, such as small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). Therefore, future work can focus on developing a more comprehensive algorithm that can detect and classify various types of lung abnormalities accurately and efficiently.

References

[1] Wan C, Ma L, Liu X, Fei B, “Computer-aided Classification of Lung Nodules on CT Images with Expert Knowledge,” Proc SPIE Int Soc Opt Eng. 2021 Feb;11598:115982K.

[2] Walter JE, Heuvelmans MA, Ten Haaf K, Vliegenthart R, van der Aalst CM, Yousaf-Khan U, van Ooijen PMA, Nackaerts K, Groen HJM, De Bock GH, de Koning HJ, Oudkerk M, “Persisting new nodules in incidence rounds of the NELSON CT lung cancer screening study,” Thorax. 2019 Mar;74(3):247-253.

[3] Siegel RL, Miller KD, Jemal A, “Cancer statistics”, 2019. CA Cancer J Clin. 2019 Jan;69(1):7-34.

[4] Oudkerk M, Devaraj A, Vliegenthart R, Henzler T, Prosch H, Heussel CP, Bastarrika G, Sverzellati N, Mascalchi M, Delorme S, Baldwin DR, Callister ME, Becker N, Heuvelmans MA, Rzyman W, Infante MV, Pastorino U, Pedersen JH, Paci E, Duffy SW, de Koning H, Field JK, “European position statement on lung cancer screening. Lancet Oncol,” 2017 Dec;18(12):e754-e766.

[5] J. Mukherjee, A. Chakrabarti, S. H. Shaikh and M. Kar, "Automatic Detection and Classification of Solitary Pulmonary Nodules from Lung CT Images," Fourth International Conference of Emerging Applications of Information Technology, Kolkata, India, 2014.

[6] Y. Zhang, B. Dai, M. Dong, H. Chen, and M. Zhou, "A Lung Cancer Detection and Recognition Method Combining Convolutional Neural Network and Morphological Features," IEEE 5th International Conference on Computer and Communication Engineering Technology (CCET), Beijing, China, 2022.

[7] Liang M, Tang W, Xu DM, Jirapatnakul AC, Reeves AP, Henschke CI, Yankelevitz D, “Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology,” 2016 Oct; 281(1):279-88.

[8] Liang M, Tang W, Xu DM, Jirapatnakul AC, Reeves AP, Henschke CI, Yankelevitz D, “Low-Dose CT Screening for Lung Cancer: Computer-aided Detection of Missed Lung Cancers. Radiology,” 2016 Oct; 281(1):279-88.

[9] Juanyun Mai, Minghao Wang, Jiayin Zheng, Yanbo Shao, Zhaoqi Diao, Xinliang Fu, Yulong Chen, Jianyu Xiao, Jian You, Airu Yin, Yang Yang, Xiangcheng Qiu, Jinsheng Tao, Bo Wang, Hua Ji, "MHSnet: Multi- head and Spatial Attention Network with False-Positive Reduction for Lung Nodule Detection," 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Las Vegas, NV, USA, 2022, pp. 1108-1114.

[10] Salama, W.M., Shokry, Aly, M.H, “A generalized framework for lung Cancer classification based on deep generative models,” Multimed Tools Appl 81, 32705– 32722 (2022).

[11] Liu K, Li Q, Ma J, Zhou Z, Sun M, Deng Y, Tu W, Wang Y, Fan L, Xia C, Xiao Y, Zhang R, Liu S, “Evaluating a Fully Automated Pulmonary Nodule Detection Approach and Its Impact on Radiologist Performance,” Radiol Artif Intell. 2019 May 29; 1(3):e180084.

[12] Wei Y, Shen G, Li JJ, “A fully automatic method for lung parenchyma segmentation and repairing. J Digit Imaging,” 2013 Jun; 26(3):483-95.

[13] Amalorpavam & T, Harish & Kumari, Jyoti & Mallaiah, Suresha, “Analysis of Digital Images Using Morphlogical Operations,” International Journal of Computer Science and Information Technology, 2013.

[14] Tran, Song-Toan, Ching-Hwa Cheng, Thanh-Tuan Nguyen, Minh-Hai Le, and Don-Gey Liu,"TMD- U-Net: Triple- U-Net with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation" Healthcare , 2021, no. 1: 54.

[15] Ronneberger, Olaf and Fischer, Philipp and Brox, Thomas, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” LNCS. 9351. 234-241.

[16] Shi F, Chen B, Cao Q, Wei Y, Zhou Q, Zhang R, Zhou Y, Yang W, Wang X, Fan R, Yang F, Chen Y, Li W, Gao Y, Shen D, “Semi-Supervised Deep Transfer Learning for Benign-Malignant Diagnosis of Pulmonary Nodules in Chest CT Images,” IEEE Trans Med Imaging, 2022 April.

[17] Guangrui Mu, and Yanbo Chen, “Relu Cascade of Feature Pyramid Networks for CT Pulmonary Nodule Detection,” Machine Learning in Medical Imaging, 10th International Conference, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China.

[18] Agrawal, Anand, Ansari, and Mehrotra, “A transfer learning approach for AI-based classification of brain tumors,” Machine Learning with Applications, 2020.

[19] Jung, Hwejin & Kim, Bumsoo & Lee, Inyeop & Lee, Junhyun and Kang, Jaewoo, “Pulmonary Nodule Classification in Computed Tomography Image Using a 3D Deep Convolutional Neural Network,” KIISE Transactions on Computing Practices, 2018.

[20] Divya Anwesh Sahu, Nagaraju.Y., Sheela Rachel.K., and Venkatesh, "Histopathological Image Classification of Breast Cancer using EfficientNet," 3rd International Conference for Emerging Technology (INCET), Belgaum, India, 2022.

[21] C. Zhao, J. Han, Y. Jia, and F. Gou, "Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network," International Conference on Networking and Network Applications (NaNA), Xi'an, China, 2018.

[22] Zhe Li , Shuo Wang, He Yu, Yongbei Zhu , Qingxia Wu, Liusu Wang , Zhangjie Wu, Yuncui Gan, Weimin Li, Bensheng Qiu , and Jie Tian, "A Novel Deep Learning Framework Based Mask-Guided Attention Mechanism for Distant Metastasis Prediction of Lung Cancer," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 7, no. 2, pp. 330-341, April 2023.

[23] Tan, Mingxing and Le Quoc, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," International Conference on Machine Learning, 2019.

[24] Suji RJ, Bhadouria SS, Dhar J, Godfrey WW, "Optical-Flow Metthods for Lung Nodule Segmentation on LIDC-IDRI Images," J Digit Imaging, 2020 Oct.

[25] Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Trans Med Imaging, 2016 May.

[26] Bianconi F, Fravolini ML, Pizzoli S, Palumbo I, Minestrini M, Rondini M, Nuvoli S, Spanu A, Palumbo B, "Comparative evaluation of conventional and deep learning methods for semi-automated segmentation of pulmonary nodules on CT," Quant Imaging Med Surg., 2021 July.

[27] Kingma, Diederik and Ba, Jimmy, "Adam: A Method for Stochastic Optimization," International Conference on Learning Representations, 2014.

[28] Jambek A.B, and Said K.A.M, "Analysis of Digital Images Using Morphological Operations," Journal of Physics Conference Series, 2021.

[29] Ahmed Abou ElFarag, Nahla M. Ibrahim, and Rania Kadry, "Gaussian Blur through Parallel Computing," International Conference on Image Processing and Vision, Engineering, 2021.

[30] Goutami Dey, Nilanjan Dey, Saurab Dutta, Sayan. C Payel Roy, and Ruben Ray, "Adaptive thresholding: A comparative study," International Conference on Control Instrumentation, Communication and Computational Technologies (ICCICCT).

[31] Haikel Alhichri, Nassim Ammour, Naif A. Alajlan, Asma S. Alswayed, and Yakoub Bazi, "Classification of Remote Sensing Images using EfficientNet-B3 CNN Model with Attention," Advanced Lab for Intelligent Systems Research (ALISR), Computer Engineering Department, College of Computer and Information Sciences, King Saud University, Saudi Arabia.