Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

# A Hybrid Deep Learning Model Combining AgresNet, YOLO, and CNN for Lung Tumor Segmentation and Classification

**Princy Magdaline P.[1], Ganesh Babu T.R.[2], Praveena R.[3]**

**Khowshalya R.[4]**

[1]Department of Electronics and Communication Engineering, Saveetha School of Engineering, SIMATS, Saveetha University, Chennai, India.

[2,3]Department of Electronics and Communication Engineering, Muthayammal Engineering College, Namakkal, India

[4]Department of Biomedical Engineering, Muthayammal Engineering College, Namakkal, India

**E-mail:** [1]princymagdalinep.sse@saveetha.com

**Abstract**

Lung cancer is a malicious tumor that originates in the respiratory tissues when the cells lining the airways of the lungs proliferate and expand uncontrollably. While computed tomography (CT) scans are capable of illuminating problematic areas, they are not proficient at independently diagnosing lung tumors. The task of accurately determining the nodule distribution within the lung CT scans through automatic identification is challenging. As a consequence, the delineation of lung CT images facilitates the identification and classification of lung tumors. This study involved the acquisition of 1097 lung CT image datasets from the Iraq-Oncology Teaching Hospital/National Centre for Cancer Diseases (IQ-OTH/NCCD). In the initial stage of the work, lung CT images are segmented to find malignancies utilizing the U-Net and Attention Gate Residual U-Net (AGRes U-Net) algorithms. The Intersection over Union (IoU) and binary focal loss metrics are applied to evaluate the segmented lung images. AGRes U-Net achieves 97% greater accuracy than the standard U-Net architecture, as determined by performance evaluation. Subsequently, a YOLOv5 network is implemented to annotate the segmented lung CT images with lesions. Thus, the segmented lung CT outputs are labelled for tumors by the YOLOv5 network are fed into the VGG-19 architecture. With an accuracy of 94.8%, this VGG-19 framework distinguishes the lung tumors as normal, benign,


*DOI: https://doi.org/10.36548/jiip.2024.4.009*
*Received: 08.12.2024, received in revised form: 04.01.2025, accepted: 23.01.2025, published: 06.02.2025*

and malignant. To compare the performance of the segmented lung output to that of the classified output by means of a convolutional neural network (CNN), the labelled lung CT images are fed as input to a CNN model in the second phase of the work. In conjunction with the Adaptive Moment Estimation (Adam) optimizer, the CNN model is implemented. The Adam optimizer demonstrates a 98% accuracy rate for the classified tumor outputs. The study illustrates that the accuracy of AGResU-Net segmentation for lung cancer detection and classification, at 97%, is nearly equivalent to that of the CNN model, which achieves 98% accuracy.

**Keywords:** Lung cancer, CT images, AGRes U-Net, YOLOv5, VGG-19

## 1. Introduction

According to GLOBOCAN 2022 data on the occurrence, fatality rate, and commonness of 36 cancer types in 185 countries, the estimated global case count for lung cancer was numbered at 2,479,698 million. This figure encompasses both genders, with males accounting for 1,571,405 cases and females for 908,293 cases placing lung cancer in its first position among the most prevalent cancers worldwide.

In the past, males were significantly more susceptible to developing lung cancer than females. Males experienced 1.6 million incidences and 1.2 million deaths from trachea, bronchus, and lung cancer in 2022, with predictions for these numbers to rise to approximately 3.0 million and 2.4 million globally by 2050. In 2022, the incidence and mortality rates of trachea, bronchus, and lung cancer for females were 0.91 million and 0.58 million, respectively, with a predicted increase to 1.7 million and 1.1 million worldwide in 2050. Figure. 1. depicts the projected global incidence and mortality rates of cancer from 2022 to 2050.

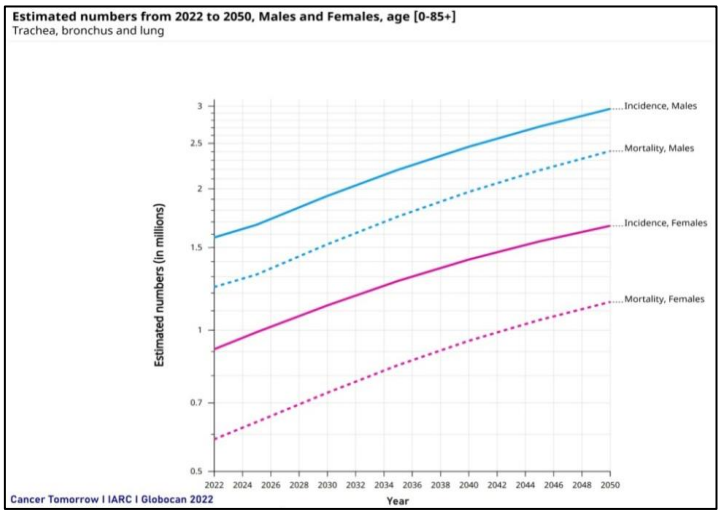



**Figure 1.** Estimated Global Prevalence and Mortality of Cancer (2022-2050) [https://www.uicc.org/what-we-do/thematic-areas/lung-cancer]

Advancements in non-invasive diagnostic methods for lung cancer have considerably increased treatment precision and success rates. Consequently, the research describes an automated system for lung cancer segmentation, detection, and classification using CT scans in this article.

The following constitute the primary contributions of the work:

- The segmentation process of lung CT images employs two distinct models: U-Net and AGResU-Net.
- The YOLOv5 architecture is employed for the purpose of tumor detection in the segmented lung CT images.
- The VGG-19 network is used for the classification of identified tumors.
- A deep CNN model is devised to sort the labelled lung CT images for tumors, evaluating its effectiveness against the results from previously segmented lung tumor.

Developed by Ronneberger et al. [1], the U-Net model is renowned for its effectiveness in segmenting biological microscope images. Its widespread application in medical image segmentation is attributed to its straightforward and versatile structure, coupled with its ability

to achieve accurate pixel-level results. The nomenclature of the U-Net, which signifies the interconnection of its layers, is derived from the letter U, signifying that the network is uniformly balanced. An expanding path (decoder) and a contracting path (encoder) compose the U-Net architecture. In their study, Chen Zhao et al. [2] developed a context-aware CNN along with a patch-based three-dimensional U-Net framework with the purpose of automatically segmenting and categorizing pulmonary nodules, with the intention of assisting radiologists in comprehending CT scans.

An investigation was conducted by Ruchita Tekade and Rajeswari [3] into the utility of CT scans in spotting the respiratory lesions and assessing cancer risk. An evaluation using artificial neural network was conducted on the collected datas utilized in their research to extract classification features. To achieve segmentation of tumors in lung CT scans, the U-NET framework was employed. A 3D multigraph VGG-like network was utilized to recognize lesions and validate their tumor aggression level. The integration of these two methodologies produced optimal results.

Mukherjee et al. [4] demonstrated the utilization of a thresholding technique to segment and classify pulmonary nodules. In order to segment the input lung images, conventional histograms and iterative thresholding methods are utilized. By utilizing rule-based filtration, lung nodules can be identified. For the purpose of classification, the KNN and SVM classifiers are employed. A modified variant of quality threshold clustering algorithm was proposed by Stelmo Magalhaes Barros Netto et al. [5] as a means of quantifying, visualizing, and analyzing alterations in respiratory abnormalities. To determine the density of pulmonary abnormalities, two data sources comprising pulmonary tract lesions were utilized: one contained exclusively aggressive tumors and the other contained non-cancerous tumors. Harmful lesions are characterized by an increase in volume and a decrease in density. Additionally, harmless lesions undergo a volumetric reduction, resulting in an increase in density.

The study conducted by Nidhi S. Nadkarni and Sangam Borkar [6] centres on the classification of lung images into two distinct categories: healthy and pathological. The researchers eliminated the random noise from the samples using a median filter. Gunaydin et al. [7] performed a comparative analysis of lung cancer detection algorithms. The statistical technique of principal component analysis is utilized to identify new patterns. In the extraction procedure, eigenvectors and entropy attributes are utilized. Classification tasks frequently demand the implementation of classifiers including Naive Bayes, KNN, and others.

Yohei Momoki et al. [8] suggested an innovative training methodology for the construction of an image classifier—one that autonomously extracts image labels from medical imaging summaries and eliminates the requirement for manual image annotation.

In the course of their research on the identification of lung nodules and the categorization of knowledge transfer candidates using multi-resolution CNN, Zuo et al. [9] undertook an inquiry. Wang et al. [10] identified and classified pulmonary tumors from CT images by utilizing CNN. The images are retrieved from the input by CNN. In order to detect nodules, false positive and false negative rates are utilized. In addition to identifying distinct nodule varieties, their sizes and locations are also specified. The process of segmentation employs a clustering technique. By employing a Faster Region CNN-based classifier after candidate detection, the FP reduction design is simplified.

In their study, Wenkai Huang and Lingkai Hu [11] proposed a noisy U-Net (NU-Net) architecture that incorporates a distinct noise component within the hidden convolution layers. The objective of this modification is to enhance the sensitivity of the neural network in identifying small lung cancer nodules. Fangzhou Liao et al. [12] devised two modules with the purpose of identifying and assessing lung cancer: three-dimensional region proposal and a leaky noisy-OR gate is utilized to correlate the five most significant nodules with the subject's lung cancer risk. Every module implemented a modified U-net.

A radiomic sequencer, a method for predicting lung carcinoma that employs a SISC architecture composed of interpretable sequencing cells, was proposed by Devinder Kumar et al. [13]. By developing a greater collaboration between radiologists and algorithms, this architecture is intended to produce more precise diagnoses. In their study, Chao Zhang et al. [14] demonstrated the results of a sensitivity analysis utilizing a dataset collected from multiple centres. Two categories have been applied to the results: pathological outcome and diameter.

Sunyi Zheng et al. [15] in their work showcased the efficacy of Maximum Intensity Projection (MIP) images in the detection of lung nodules during CT scans. MIP images representing slab thicknesses are utilized as input. By augmenting 2DCT slice images with more representative spatial data; this approach enables the differentiation of nodules from vessels according to their morphologies. Consequently, an increase in sensitivity is accompanied by a reduction in the occurrence of false positives. A joint lung nodule

identification and segregation network was devised by Liu Chenyang et al. [16] using label ambiguity CT images. Classification and nodule detection subnetworks of the joint network employ three-dimensional encoder-decoder architecture. In order to enhance the performance of classification, the nodule classification subnetwork of the combined network utilizes multiscale nodule-specific features in addition to those extracted from the encoder output of the detection subnetwork.

Based on the vesselness filter and the Multi-Scene DL Framework (MSDLF), Qinghai Zhang et al. [17] devised a method for identifying pulmonary nodules. Two discrete categories are relevant to this notion. This study develops an autonomic mechanism for mending the lung wall so as to prevent the omission of juxta-pleural nodules. A technique known as 3D-Res2UNet, which combines 3D-UNet and Res2Net, was proposed by Zhitao Xiao et al. [18] to segment lung nodules in CT images. Featuring a robust capability to extract features at various dimensions, the 3D-Res2Net convolutional neural connection network is symmetrical and hierarchical. This feature enhances the coverage area of each network layer while enabling the network to more precisely express attributes at multiple levels. By employing this framework, the underlying issue is resolved. Enhancing the accuracy of detection and segmentation, the network is impervious to gradient disappearance and gradient eruption. Additionally, this improves the segmentation process's comprehensiveness and accelerates the model's training. This strategy outperforms the alternatives with regard to the dice coefficient and recall rate.

Divya and Ganesh Babu [19] have conducted a comprehensive examination of optimization algorithms. By connecting the input and output of the multi-layer convolutional layers and utilizing the resultant feature representation of one layer as the input for the subsequent layer, a natural inter-layer flow is established. The model is trained using the optimization methods Adam, SGDM, and RMSProp, and the resulting predictions are compared. An improved 3D dense connected U-Net (I-3D Dense U-Net) was developed by Guobin Zhang et al. [20] to segment various lung tumors from CT images. The nested dense skip connection of the I-3D Dense U-Net attempts to supply encoder and decoder subnetworks with identical feature maps. Encoder-decoder blocks employ dense connections to facilitate feature propagation and reuse.

In the study they performed, Yu Fu et al. [21] employed chest CT scans to accurately identify and classify five distinct pathological classifications of lung nodules. A reverse adversarial classification network was used to represent the five adverse types, and a distinct

regression task was formulated. The regression task enables the reverse adversarial classification network to extract specific features that can be incorporated into a CNN. Naseer et al. [22] proposed a classification system for lung cancer that uses a modified U-Net-based model with three stages to detect and segment nodules. The lobe was divided into segments using a CT slice and a prediction mask applied in a modified U-Net architecture during the initial phase. Afterward, the candidate node was extracted using the prediction mask and a modified U-Net architecture card.

A computer-aided design model that precisely identifies and describes the lung nodule was proposed by I. Shafi et al. [23]. The first training of the model entails identifying lung cancer by analyzing and comparing distinct profile values in the CT scans of patients and controls during the diagnosis stage. Afterward, the model is evaluated and validated by using CT scans from patients and control subjects who were not part of the training phase.

## 2. Materials and Methods

This portion presents details concerning the data collected, as well as the AGResU-Net, YOLOv5, and VGG-19 architectures.

### 2.1 Dataset Description

Across the fall season of 2019, 1097 CT images were amassed from the IQ-OTH/NCCD to support this research project, spanning three months. The collected data's are grouped into three divisions, normal, benign, and malignant instances. The utilization of benign classes to augment the data is necessitated by the imbalanced dataset's disproportionate class assignment. 480 additional benign cases originated following augmentation. The total number of lung images following enhancement reaches 1457 samples. A dataset of 1020 samples was involved in training the segmentation models, U-Net and AGRes U-Net. While the remaining 437 samples were used for validation purposes. Figure.2. illustrates the number of instances in each stage of the dataset before and after augmentation.

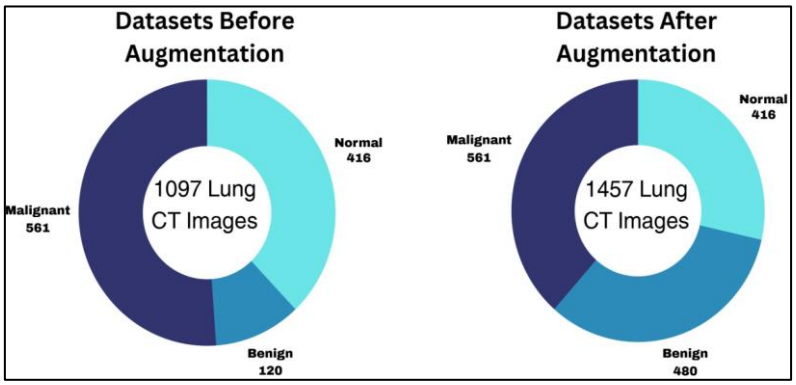



**Figure 2.** Number of Instances in Each Stage of the Dataset Before and After Augmentation

## 2.2 Proposed Method

In the first phase of the work, the lung CT images are input into two segmentation models, of which one is U-Net and the other is AGRes U-Net. Figure. 3. depicts the suggested process diagram for segmentation and identification of pulmonary nodules. To perform lung region segmentation, 70% of the dataset's samples are utilized in the training process for both the U-Net and AGRes U-Net frameworks, while 30% of the samples are utilized for evaluation. For performance metrics, Intersection over Union (IoU) and binary focal loss are involved in validating both segmentation models. The AGRes U-Net exhibits a greater degree of accuracy in comparison to U-Net. Subsequently, the AGRes U-Net output is utilized in conjunction with the YOLOv5 to label the lesions in the segmented image.

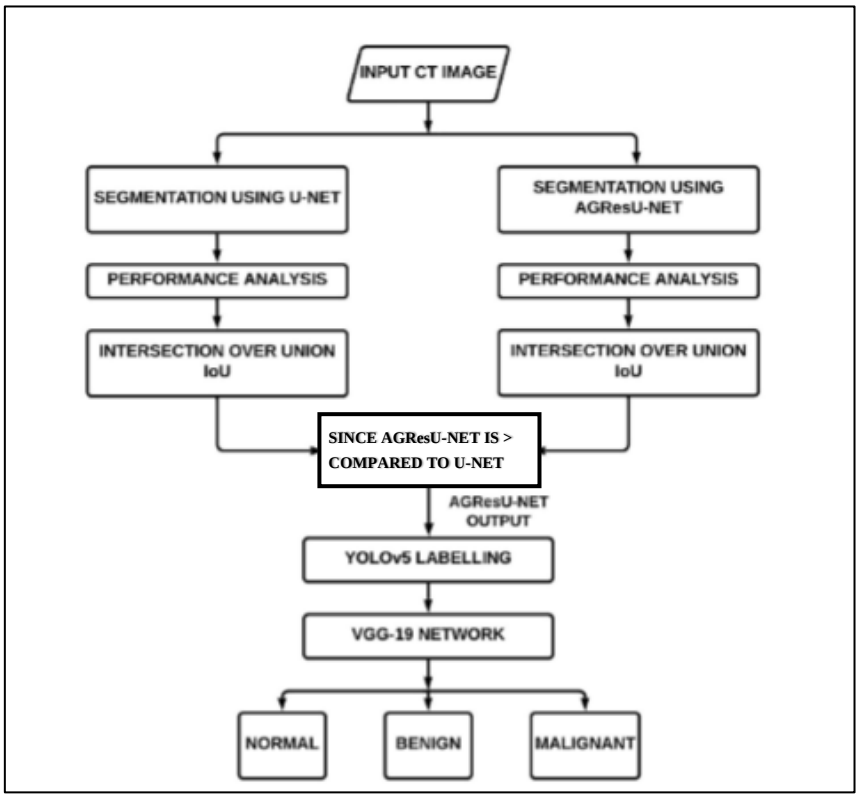



**Figure 3.** Process Diagram for Segmentation and Identification of Lung Tumor

Using bounding boxes, the YOLOv5 identifies and highlights the tumors. After labelling, the annotated image is processed by the VGG-19 network to determine whether the tumor is normal, harmless, or harmful.

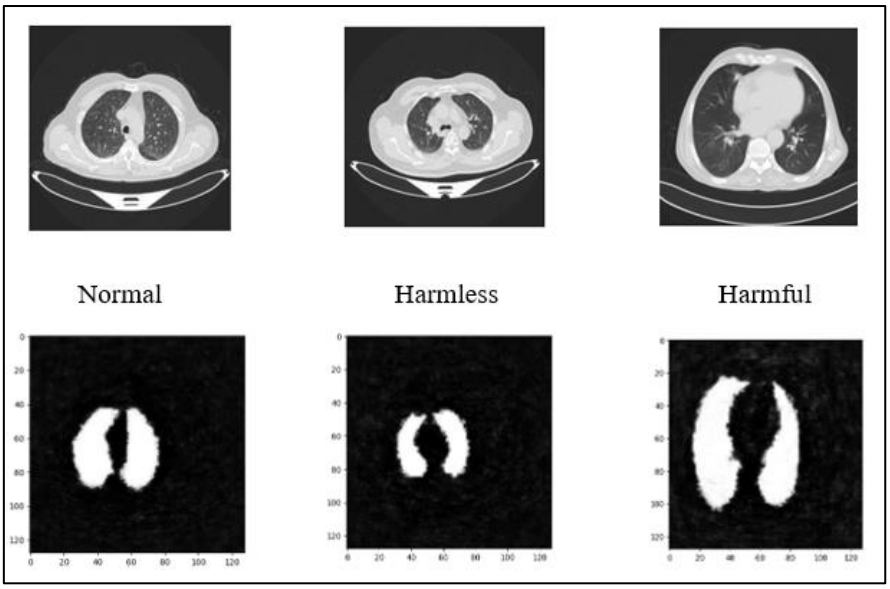
### 2.3 Lung Region Schematic Segmentation using U-Net

One computational model that has been used for biomedical image partitioning is U-Net. Ensuring that the input and output are identical in size, it classifies each individual pixel of an image. The U-Net architecture is symmetrical; the left side has an encoder that downsamples the input image, while the right side contains a decoder that upsamples. The encoder part is made up of numerous convolutional blocks, each of which is followed by a downsampling operation. The decoder part comprises many upsampling operations, each followed by a convolutional block that includes upsampling layers, convolutional layers, and concatenations. The U-Net architecture utilizes the transposed convolution operation to enhance the fidelity of the downscaled image produced by convolution to the original dimensions. This improves the accuracy with which all pixels are classified during image

segmentation. Throughout the process of successive convolutions and irregular transitions, it is typical to lose spatial data and locational specifics from the high-level output maps.

### 2.3.1. Segmentation output of U-Net

This section represents how a selection of learning datasets of lung CT scans from three cases: normal, harmless, and harmful are selected and segmented using the U-Net architecture. The envisioned images from the training dataset depict the segmented lung regions. Figure. 4. illustrates the partitioned outputs of U-Net for a selection of learning datasets.




**Predicted Output Images of Training Dataset**

**Figure 4.** U-Net Partitioned Outputs for a Selection of Training Datasets

The proportion of the image is 128 x 128 x 3, and an Adaptive moment optimizer is used in this work with a learning rate of 0.0001. The number of epochs is 50. The U-Net architecture is evaluated for its performance in segmenting lung CT images. This evaluation involves calculating their binary focal loss and Intersection over Union (IoU). Table 1 illustrates the IoU and binary focal loss values for a sample of ten images out of 1457 images. For the U-Net segmented lung image, the accuracy and binary focal loss values are 87% and 23%, respectively.

**Table 1.** IoU and Binary Focal Loss Values of U –Net

| U-Net | |
|---|---|
| **IoU** | **Binary Focal Loss** |
| 0.68 | -5.8 |
| 0.76 | -6.6 |
| 0.74 | -6.4 |
| 0.78 | -6.8 |
| 0.75 | -6.5 |
| 0.74 | -6.4 |
| 0.68 | -6.6 |
| 0.58 | -6.6 |
| 0.58 | -6.6 |
| 0.59 | -6.5 |

The U-Net can struggle with preserving fine details in images, especially when dealing with complex structures. A classic U-Net approach struggles to distinguish between pulmonary arteries and nodules. The fixed receptive field in a U-Net may not be sufficient to capture varying scales of features in an image. Small receptive fields might miss the global context, while large ones might overlook fine details. Different parts of an image might require different receptive field sizes for optimal segmentation. A fixed receptive field cannot adapt to these varying requirements. To overcome these disadvantages in U-Net and to improve the accuracy, the AGResU-Net architecture is used.

### 2.4 Lung Region Schematic Segmentation using AGResU-Net

Therefore, to improve the efficiency of segmentation, the AGResU-Net model was developed by Aasia Rehman et al. [24] by integrating residual modules and attention gates into the U-Net framework. The AGResU-Net combines both attention gates and residual modules in U-Net; the attention gates assist the model in focusing on the most important features, improving segmentation accuracy. Additionally, the residual modules in AGResU-Net efficiently solve the vanishing gradient problem, making the training process more stable and resulting in improved performance. Furthermore, the model is highly flexible and can be adapted to various types of segmentation tasks, making it a versatile choice. During the downsampling phase, the AGResU-Net model extracts more precise density feature

information as the residual connections help to prevent vanishing gradients and promote better gradient flow by feeding the input to the output. During the upsampling phase, attention gate units are utilized to transmit relevant location information and spatial details from low-level feature maps. The network configuration of AGResU-Net is depicted in Figure 5. AGResU-Net's attention gate units are responsible for emphasizing crucial feature details while removing non-relevant and distracting feature details. They direct increased attention towards very small lesions and extract more accurate details regarding their locations, while residual modules enhance the extraction and expression of features.

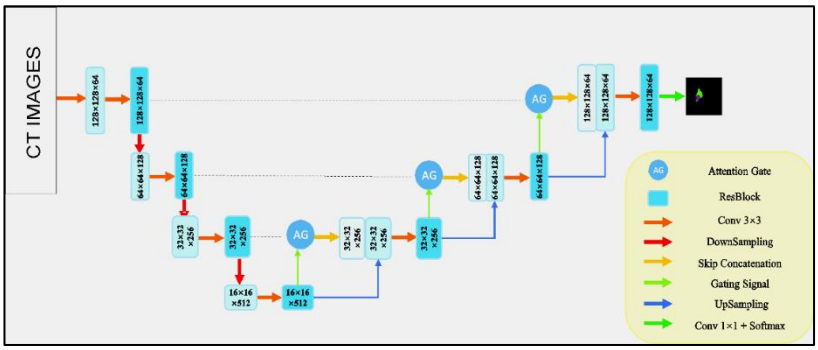



**Figure 5.** Network Configuration of AGRes U-Net

The mathematical model of AGRes U-Net is described here.

1. Input

   The input to AGRES U-Net is a CT image of size $H * W * C$

   Where

   **H** - height, **W** - width and **C** - number of channels of the image

2. Encoder

The encoder consists of several convolutional blocks, each followed by a downsampling operation that is carried out by convolutional layers, Batch normalization, ReLU activation, and residual connection.

The operation of the convolutional block is defined by,

$$y_{i,j,k} = \sigma\left(BN\left(Conv\left(x_{i,j,c,}W_k\right)\right)\right) + x_{i.j.k} \quad (1)$$

Where $x_{i,j,c}$ $is\ the\ input\ feature\ map$

$$W_k \text{ are the convolutional filters}$$
$$Conv \text{ denotes the convolution operation}$$
$$BN \text{ denotes the batch normalization}$$
$$\sigma \text{ denotes the ReLU activation function}$$
$$x_{i.j.k} \text{ back to the result of the operation}$$

3. Bottleneck

The bottleneck is very similar to the encoder block but without downsampling.

4. Decoder

The decoder consists of several upsampling operations and it includes Upsampling layers, convolutional layers, concatenation, and attention mechanism.

The upsampling block can be defined as

$$y_{i,j,k} = Conv(\sigma(BN\left(Upsample\left(x_{i,j,c}\right)\right)) \quad (2)$$

$$y_{i,j,k} = Attention(y_{i,j,k} * skipconnection) \quad (3)$$

5. Attention Mechanism

The attention mechanism is represented as,

$$e_{i,j,k} = tanh\left(W_x x_{i,j,k} + W_g g_{i,j,k}\right) \quad (4)$$

$$\alpha_{i,j,k} = \frac{exp(e_{i,j,k})}{\sum_{k\prime} \quad exp(e_{i,j,k\prime})} \quad (5)$$

$$y_{i,j,k} = \alpha_{i,j,k}.\, x_{i,j,k} \quad (6)$$

Where

$$W_x \text{ and } W_g \text{are learnable weight matrices}$$

$$g_{i,j,k} \text{ is the gating signal}$$

$$\alpha_{i,j,k} \text{ are the attention coefficients}$$

6. Output

The last layer is a convolutional one, succeeded by a softmax activation function, to produce pixel-wise class probabilities.

$$y_{i,j,k} = softmax\left(Conv\left(x_{i,j,c}, W_{out}\right)\right) \quad (7)$$

Where

$$W_{out}\ are\ the\ convolutional\ filters$$

7. Loss Function

The loss function of AGRES U-Net is given by,

$$L = CrossEntropy(y, \hat{y}) + Diceloss(y, \hat{y}) \quad (8)$$

Where $y$ is the ground truth segmentation map.

$$\hat{y}\ is\ the\ predicted\ segmentation\ map$$

The cross-entropy loss is given as

$$Cross\ Entropy(y, \hat{y}) = -\sum_{i,j} \quad \sum_{c} \quad y_{i,j,c} log\left(\hat{y}_{i,j,c}\right) \quad (9)$$

The Dice loss is represented as

$$Dice\ loss(y, \hat{y}) = 1 - \frac{2\sum_{i,j} \quad y_{i,j}\hat{y}_{i,j}}{\sum_{i,j} \quad y_{i,j}^2 + \sum_{i,j} \quad \hat{y}_{i,j}^2} \quad (10)$$

### 2.4.1. Segmentation output of AGRes U-Net

This part explains about the illustration of the utilization of the AGRes U-Net framework to segment a selection of learning datasets consisting of lung CT images as suggested by Princy Magdaline and Ganesh Babu [25] for three cases, namely normal, harmless, and harmful tumors. In this work, the same hyper-tuning parameter and the same set of images are used for U-Net. Segmented lung regions are depicted in the predicted images derived from the learning dataset. The predicted images acquired for the training dataset correspond to segmented regions of the lungs. Figure. 6. represents the AGRes U-Net partitioned outputs for a selection of learning datasets.

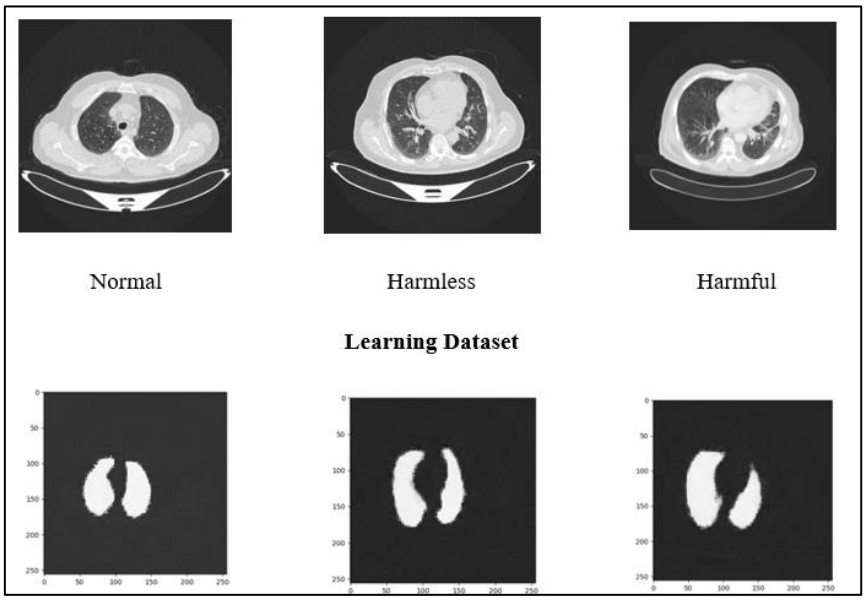



**Figure 6.** AGRes U-Net Partitioned Outputs for a Selection of Learning Datasets

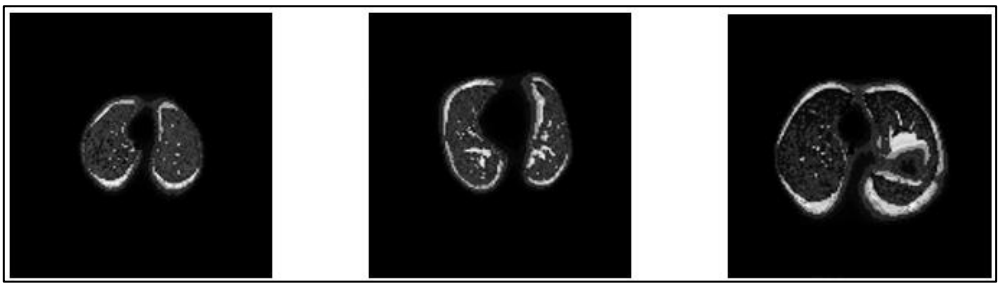
### 2.4.2. Analysis of Segmented Output from AGRes U-Net

This evaluation involves calculating their binary focal loss and Intersection over Union (IoU). Table 2 indicates IoU and pixel loss of ten lung CT images out of 1457 images. The accuracy and pixel loss of the proposed algorithm are 97% and 5.2%, respectively. The analysis indicates that the AGResU-Net outperforms the U-Net in terms of higher IoU values.

**Table 2.** IoU and Binary Focal Loss of AGRes U-Net

| AGResU-Net | |
|---|---|
| **IoU** | **Binary Focal Loss** |
| 0.81 | -8.1 |
| 0.78 | -8.5 |
| 0.84 | -8.3 |
| 0.88 | -8.3 |
| 0.95 | -8.5 |
| 0.95 | -.8.4 |
| 0.85 | -8.3 |
| 0.87 | -8.2 |
| 0.89 | -8.2 |
| 0.92 | -8.9 |

Therefore, the partitioned lung region images from AGResU-Net are considered for further processing. Utilizing AGResU-Net, the image displays an obscured section of the lungs with lung tumor regions extracted through a Pixel-specific, bit-level AND operation, utilizing the primary sample and the masked sample. The appearance of white spots in the lung region suggests the existence of a tumor. Figure. 7. illustrates the segmented regions of the lungs.



Normal Case Benign Case Malignant Case

**Figure 7.** Segmented Lung Regions

### 2.5 Tumor Outputs Labelled in YOLOv5

The YOLOv5 is employed for indicating the tumors within the lung CT image that has been segmented. The tumor is identified by this network and subsequently captured in bounding boxes. The Figure. 8. represents the single-stage architecture of YOLO.

The main features of YOLOv5 architecture are

- The Single-stage detection process in YOLOv5 processes the entire image in a single pass, thus reducing the latency and enabling faster decision-making.
- The image processing rate of YOLOv5 is 150 FPS (frames per second) for smaller models and 45 FPS for larger models. Thus it is faster than many other object detection models.
- YOLOv5 uses CSPNet (Cross Stage Partial Network) in its backbone and Mosaic Augmentation thus reducing the computational cost while balancing speed and maintaining high accuracy.

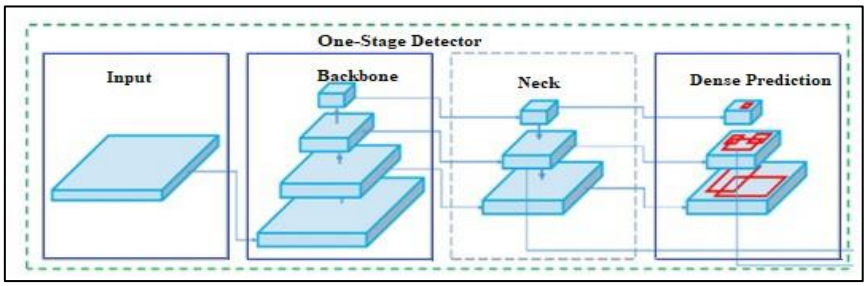



**Figure 8.** Representation of Single Stage Architecture of YOLO

The mathematical model of YOLO model is described here

The source to YOLOV5 is an image of dimensions H*W*C

Where, H - height, W - width and C - number of channels of the image

**Backbone:** The backbone is one of the CNN that extract details from the input image. Backbone in CNN includes convolutional layers, batch normalization, neck and head.

**Batch Normalization:** Normalizes the activations to improve training stability and speed

Leaky ReLU Activation: Applies non-linearity to the feature maps

The convolution operation carried out in YOLOv5 is explained by mathematical equation

$$y_{i,j,k} = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{c=1}^{C} x_{i+m-1,j+n-1,c} \cdot w_{m,n,c,k} + b_k \quad (11)$$

Where

$$y_{i,j,k} \; is \; the \; ouput \; of \; the \; feature \; map$$

$$x_{i,j,c} is \; the \; input \; feature \; map$$

$$w_{m,n,c,k} \; are \; the \; weights \; of \; the \; filter$$

$$b_k \; is \; the \; bias \; term$$

**Neck:** The neck is responsible for aggregating features from different stages of the backbone. YOLO v5 uses a PANet for this purpose. It composes of two types one is feature pyramid networks and other is upsampling and concatenation. This term is mathematically expressed by

$$y_{i,j} = x_{\lfloor \frac{i}{n} \rfloor, \lfloor \frac{j}{n} \rfloor} \quad (12)$$

Where s is the upsampling factor

**Head:** The head of YOLO v5 is responsible for making final predictions.

$$Bounding\ Box\ Coordinates: (t_x, t_y, t_w, t_h) \quad (13)$$

$$Objectness\ Score: P_0$$

$$Class\ Probabilities: P_c$$

The final bounding box coordinates are computed as,

$$b_x = \sigma(t_x) + c_x \quad (14)$$

$$b_y = \sigma(t_y) + c_y \quad (15)$$

$$b_w = pwe^{t_w} \quad (16)$$

$$b_h = phe^{t_h} \quad (17)$$

Where,

$$(c_x, c_y) is\ the\ top - left\ corner\ of\ the\ grid\ cell$$

$$(p_w, p_h) are\ the\ anchor\ box\ dimensions$$

$$\sigma\ is\ the\ sigmoid\ functio$$

**Loss Function:** The YOLOv5 loss function combines three types of losses that are boundary box regression loss, objectness loss, and class probability loss.

The mathematical model of total loss is given by

$$L = L_{box} + L_{obj} + L_{cls} \quad (18)$$

Where

$$L_{box}\ is\ the\ bounding\ box\ regression\ loss$$

$$L_{obj}\ is\ the\ objectness\ loss$$

$$L_{cls}\ is\ the\ class\ probablity\ loss$$

As depicted in Figure. 9, the YOLO network does not display any bounding boxes for a normal input sample because there are no indications of a tumor. Bounding boxes are utilized to depict the tumor in both benign and malignant images. Figure. 10. and Figure. 11. illustrate the labeled cancerous outputs of YOLOv5 for images that are benign and malignant, respectively.

**Figure 9.** Normal Lung Image Output in YOLOv5

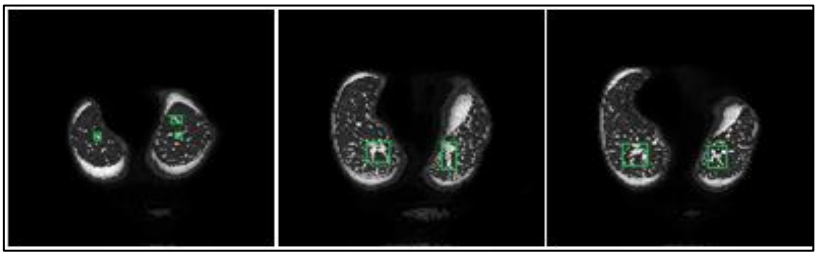


**Figure 10.** YOLOv5 Labelled Tumors in Benign Case Images

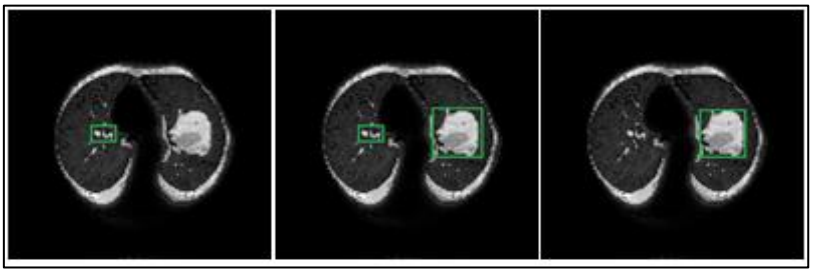


**Figure 11.** YOLOv5 Labelled Tumors in Malignant Case Images

## 2.6 Analysis of VGG-19's Classified Outputs

The architecture VGG-19 classifies the predicted and marked output from the YOLOv5 model into the following categories: Normal, Benign, and Malignant. A performance evaluation is conducted for each individual case by employing a range of metrics. Figure 12 presents the performance metrics computed in VGG-19.

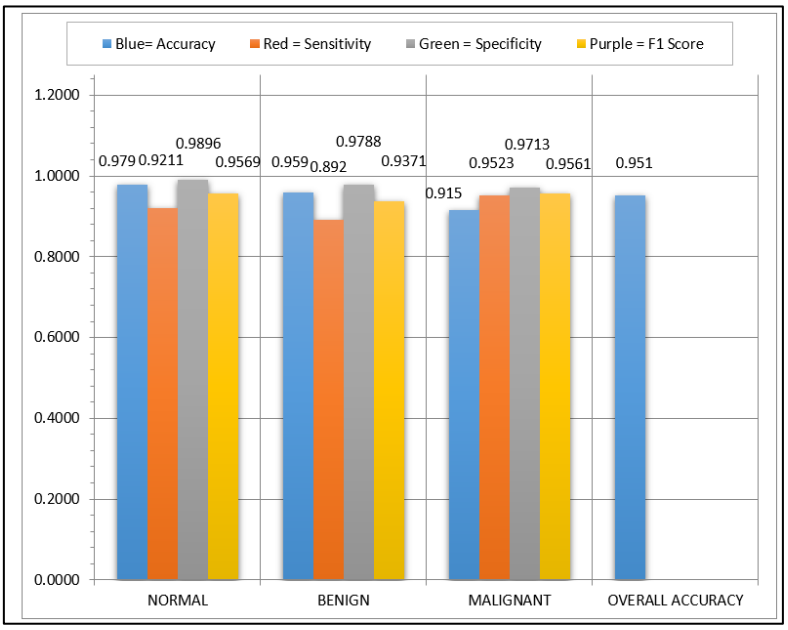



**Figure 12.** Performance Measurements Evaluated within VGG-19

The segmented output acquires an accuracy of 95.1%, as shown in Figure 12 for the VGG-19 architecture. Furthermore, when compared to the normal and benign sensitivity values of 0.9211 and 0.892, respectively, the sensitivity metric for the malignant case is a commendable 0.9523. Sensitivity values for benign and normal cases are challenging for the VGG-19 to predict.

## 2.7 Classification of Lung Cancer Using CNN

In the second phase of the work, CNN architecture and Adam optimization techniques are involved. The marked CT lung images are fed into the CNN model for the purpose of classifying and detecting lung carcinomas. We carry out this to determine whether the segmentation and classification approach or the deep CNN method is effective and appropriate for lung cancer detection and classification. The block diagram depicting the classification of lung tumors is referenced in Figure. 13. The CT images of the lungs are first resized and input into the deep learning CNN that has been proposed. A compilation of 1457 labelled lung CT images was utilized, comprising 561 malignant cases, 416 benign cases, and 480 normal cases.

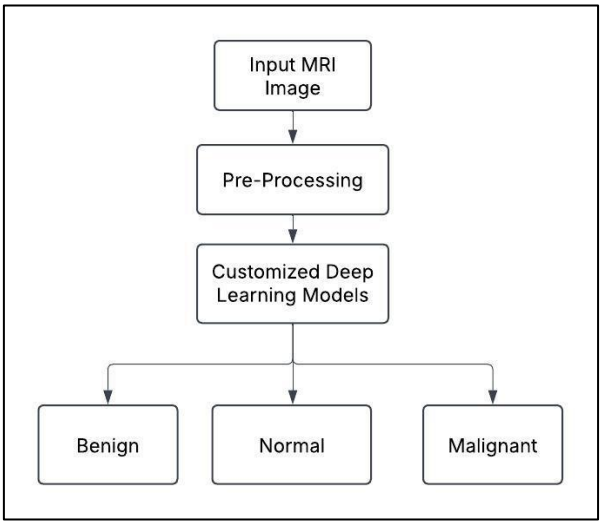



**Figure 13.** System Layout of Categorization of Lung Nodule using CNN

In this work, the hyperparameter tuning involves setting the learning rate to 0.0001. The dataset is divided into 70% for training and 30% for testing. The model is trained for 50 epochs. The input size for the model is defined as 244x244x3, and the batch size is set to 16. The Adam optimizer is used for optimization, while the categorical cross-entropy function is applied as the loss function. Additionally, the activation function employed in the output layer is SoftMax.

The CNN model, Lecun et al. [26], performs convolution operations using convolutional filters designed to acquire and retrieve critical properties that aid in the effective comprehension of medical images. A convolved feature is created by applying a dot product function to the filter entries and the input data. This resultant feature map is passed to the subsequent layer. Each convolution layer includes such filters to generate feature maps. By transforming an image to numerical values, this layer makes it easier for the CNN to analyze the image and identify patterns that are significant. Post-extraction, the feature maps are handed off to the Rectified linear unit (ReLU) activation function layer, to enable nonlinear data transformation, which improves the deep learning model's ability to encompass, adjust, and segregate across diverse sets of input and outputs.

### 2.8 Adaptive Moment Estimation (Adam) Optimization

Optimization algorithms in a deep learning model are tasked with minimizing losses, increasing training speed, and giving the best possible results. When correlating inputs to

outputs, an optimization technique selects the suitable weight that minimizes error.

Adam merges the functionalities of AdaGrad and RMSProp; in deep learning training within a networked neural model, it independently modifies the learning rate for each network weight. Based on past gradients and second moments, the Adam optimizer dynamically adjusts individual learning rates. The training speed employed in this study is 0.0001, with 50 cycles. Figures 14 (a) and (b) shows Adam's iterations (epochs) vs. accuracy rate graph and performance matrix[25].

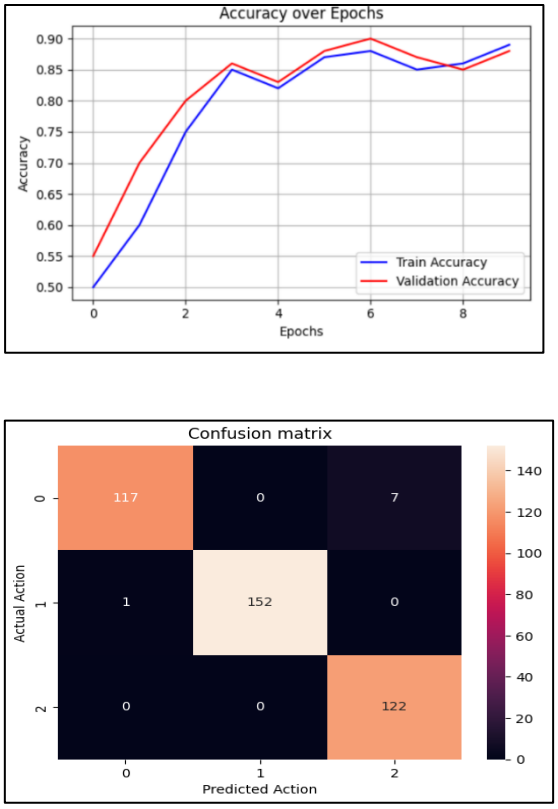



**Figure 14.** (a) Iterations Vs Accuracy Rate Plot (Adam) (b) Performance Matrix (Adam)

The metrics of sensitivity, specificity, precision, and accuracy are determined using the elements of the performance matrix listed in Table 3. It is observed that CNN achieves model overall accuracy rate of 98%. Thus a CNN classifier with an Adam optimizer is implemented, and the CT lung images are classified.

**Table 3.** Performance measurements evaluated within proposed CNN

| Cases | Sensitivity | Specificity | Precision | Accuracy |
|---|---|---|---|---|
| Normal cases | 96.4 | 98.98 | 98.24 | 96.84 |
| Benign cases | 99.86 | 98.89 | 98.12 | 98.56 |
| Malignant cases | 93.37 | 99.87 | 99.90 | 96.12 |

## 3. Results and Discussion

In this study, semantic segmentation, specifically AGRes U-Net segmentation, and a CNN model was proposed and employed for lung cancer detection and classification.

U-Net segmentation and AGRes U-Net segmentation are the two methods used in the initial stage of the work to semantically segment CT images of the lung. Subsequently, both units undergo performance analysis, and their IoU and Binary focal loss are computed. The IoU values in AGRes U-Net are greater than those in U-Net when compared. Additionally, the binary focus loss decreases with increasing IoU, allowing the AGRes U-Net to surpass the conventional U-Net architecture with 97% accuracy. Using the YOLOv5 network, which marks the tumors with bounding boxes, the AGRes U-Net partitioned result is marked for lung carcinoma. The detected tumors are then, with 94.8% accuracy, classified as normal, benign, or malignant using the VGG-19 architecture. To identify and classify lung cancers, the accuracy of the segmented image must be compared to the accuracy rate of the labelled image produced by the CNN model. The second stage of the work is completed for this.

In the latter part of the work, the annotated lung scan image is processed by a CNN model, which trains and classifies lung cancer using the Adam optimizer, achieving 98% accuracy. Lung cancer identification and classification using AGRes U-Net segmentation achieves 97% accuracy, which is almost identical to the CNN model's 98% accuracy. Thus, AGRes U-Net lung CT image partitioning with the YOLOv5 and VGG-19 networks is effective for lung carcinoma identification and classification.

## 4. Conclusion

Thus, the focus of this research is on lung image segmentation initially, followed by classification of lung tumors. Tumors in the lungs cannot be seen or mapped without segmentation. The importance of this study is that semantic partitioning is used to identify lung regions in CT lung images using the AGRes U-Net model. The YOLOv5 architecture is utilized to determine the particular location of tumors in the lung region, while classifiers are employed to classify lung tumors. The segmentation, detection, and classification of lung tumors will aid in the unveiling of a patient's lung cancer, and depending on their health status, extra treatments can be offered to improve their survival rate.

## References

[1] Ronneberger O, Fischer P, Brox T, 'U-net: Convolutional Networks for Biomedical Image Segmentation', International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, pp.234–241.

[2] Chen Zhao, Jungang Han, Yang Jia and Fan Gou, 'Lung Nodule Detection via 3D U-Net and Contextual Convolutional Neural Network', International Conference on Networking and Network Applications (NaNA), Publisher: IEEE, 2018,

[3] Ruchita Tekade & Rajeswari K, 'Lung Cancer Detection and Classification using Deep Learning', Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), IEEE XPLORE, 2018

[4] Mukherjee M & Biswal PK, 'Segmentation of Lungs Nodules by Iterative Thresholding Method and Classification with Reduced Features', International Conference on Inventive Communication and Computational Technologies, 2018, 450-455.

[5] Stelmo Magalhães Barros Netto, João Otávio Bandeira Diniz, AristófanesCorrêa Silva, Anselmo Cardoso de Paiva, Rofolfo Acatauassú Nunes & Marcelo Gattass, 'Modified Quality Threshold Clustering for Temporal Analysis and Classification of Lung Lesions', IEEE Transactions on Image Processing, 2019, vol. 28, no. 4, 1813-1823.

[6] Nidhi S. Nadkarni and Prof. Sangam Borkar, 'Detection of Lung Cancer in CT Images using Image Processing', Proceeding of the Third International Conference on Trends and Informatics (ICOEI), IEEE XPLORE, 2019, DOI:10.1109/ICOEI.2019.8862577.

[7] Günaydin Ö, Günay M &Şengel Ö, Comparison of Lung Cancer Detection Algorithms, Electrical-Electronics & Biomedical Engineering and Computer Science,2019, pp.1-4.

[8] Yohei Momoki, Akimichi Ichinose, YutaroShigeto, Ukyo Honda, Keigo Nakamura & Yuji Matsumoto 'Characterization of Pulmonary Nodules in Computed Tomography Images Based on Pseudo-Labeling Using Radiology Reports', IEEE Transactions on Circuits and Systems For Video Technology, 2022, vol. 32, no. 5, 2582-2591.

[9] Zuo W, Zhou F, Li Z & Wang L, 'Multi-Resolution CNN and Knowledge Transfer for Candidate Classification in Lung Nodule Detection', IEEE Access, 2019, vol.7, 32510-32521.

[10] Wang J, Wang J, Wen Y, Lu H, Niu T, Pan J & Qian D, 'Pulmonary Nodule Detection in Volumetric Chest CT Scans Using CNNs-Based Nodule-Size-Adaptive Detection and Classification', IEEE Access, 2019, vol.7, 46033-46044.

[11] Wenkai Huang &Lingkai Hu, 'Using a Noisy U-Net for Detecting Lung Nodule Candidates', IEEE Open Access Journal, 2019, vol. 7, pp. 67905-67915.

[12] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu &Sen Song, 'Evaluate the Malignancy of Pulmonary Nodules Using the 3-D Deep Leaky Noisy-OR Network', IEEE Transactions on Neural Networks and Learning Systems, 2019, vol.30, no. 11, 3484-3495.

[13] Devinder Kumar, Vignesh Sankar, David Clausi, Graham W. Taylor & Alexander Wong, 'SISC: End-to-End Interpretable Discovery Radiomics-Driven Lung Cancer Prediction via Stacked Interpretable Sequencing Cells', IEEE Open Access Journal, 2019, vol. 7, 145444-145454.

[14] Chao Zhang, Xing Sun, Kang Dang, Ke Li, Xiao-wei Guo, Jia Chang, Zong-qiao Yu, Fei-yue Huang, Yun-sheng Wu, Zhu Liang, Zai-yi Liu, Xue-gong Zhang, Xing-lin Gao, Shao-hong Huang, Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolutional Neural Network, The Oncologist, 2019, Volume 24, Issue 9, 1159–1165.

[15] Sunyi Zheng, Jiapan Guo, Xiaonan Cui, Raymond N. J. Veldhuis, Matthijs Oudkerk& Peter M. A. van Ooijen, 'Automatic Pulmonary Nodule Detection in CT Scans Using

Convolutional Neural Networks Based on Maximum Intensity Projection', IEEE Transactions on Medical Imaging, 2020, vol. 39, no. 3, 797-805.

[16] Liu Chenyang & Shing-Chow Chan, 'A Joint Detection and Recognition Approach to Lung Cancer Diagnosis from CT Images with Label Uncertainty', IEEE Open Access Journal, 2020, vol. 8, 228905-228921.

[17] Qinghai Zhang &Xiaojing Kong, 'Design of Automatic Lung Nodule Detection System Based on Multi-Scene Deep Learning Framework', IEEE Open Access Journal, 2020, vol. 8, 90380-90389.

[18] Zhitao Xiao, Bowen Liu, Lei Geng, Fang Zhang &Yanbei Liu, 'Segmentation of Lung Nodules Using Improved 3D-UNet Neural Network', MDPI, Symmetry, Special Issue Computational Intelligence and Soft Computing: Recent Applications, 2020, vol. 12, no. 11, 1787

[19] Divya & Ganesh Babu, TR, 'Fitness Adaptive Deer Hunting- Based Region Growing and Recurrent Neural Network for Melanoma Skin Cancer Detection', International Journal of Imaging System Technology, Wiley Publication, 2020, vol.30, no. 3, pp.731-752.

[20] Guobin Zhang, Zhiyong Yang & Shan Jiang, 'Automatic lung tumor segmentation from CT images using improved 3D densely connected UNet', Article, Medical, Biological Engineering and Computing, 2022, DOI: 10.1007/s11517-022-02667-0.

[21] Yu Fu, Peng Xue, Taohui Xiao, Zhili Zhang, Youren Zhang &Enqing Dong, 'Semi-Supervised Adversarial Learning for Improving the Diagnosis of Pulmonary Nodules', IEEE Journal of Biomedical and Health Informatics, 2023, vol. 27, no. 1, 109-120.

[22] Naseer, I, Akram, S, Masood, T, Rashid, M & Jaffar, A, 'Lung Cancer Classification using Modified U-Net based Lobe Segmentation and Nodule Detection', IEEE Access, 2023, vol. 11, 60279-60291.

[23] Shafi I, Din S, Khan A, Díez IDLT, Casanova RdJP, Pifarre KT &Ashraf I, 'An Effective Method for Lung Cancer Diagnosis from CT Scan Using Deep Learning-Based Support Vector Network', Cancers, 2022, vol. 14, no. 21, 5457.

[24] Aasia Rehman, Muheet A. Butt & Majid Zaman, 'Attention Residual UNet With Focal Tversky Loss for Skin Lesion Segmentation', International Journal of Decision Support System Technology, 2023, vol. 15, no. 1, 1-17.

[25] Princy Magdaline P and Ganesh Babu T R, 'Detection of Lung Cancer Using Novel Attention Gate Residual U-Net Model And KNN Classifier From Computer Tomography Images', Journal of Intelligent and Fuzzy Systems, 2023, vol.11, no.4, 1334-1340.

[26] Lecun Y, Bottou L, Bengio Y& Haffner P, 'Gradient Based Learning Applied to Document Recognition', In Intelligent Signal Processing, IEEE Press, 2001, 306-351.