Journal of
Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.98549430

LUNG CANCER DETECTION AND SEGMENTATION BASED ON DEEP LEARNING APPROACHES

R. Soundharya Devi [1], S. Nirmala Sugirtha Rajini[2]
1. Research Scholar 2. Professor
Department of Computer Applications
Dr.MGR. Educational and Research Institute, Maduravoyal, Chennai -95, India
rsoundharya92devi@gmail.com nirmalasugirtharajini.mca@drmgrdu.ac.in

Abstract: - *Lung cancer is a leading cause of death globally and early detection is crucial for successful treatment. Deep learning with image processing techniques has become a popular and effective tool for medical image analysis, particularly in the detection and segmentation of lung cancer. This survey paper provides a comprehensive overview of the recent advancements in the use of DL (Deep Learning) for lung cancer detection and segmentation. The aim of this study was to compare the accuracy and sensitivity of three different image segmentation methods for lung cancer detection and segmentation. The methods evaluated in the study were DB U-Net+LLIE, U-net and DenseNet & dilation block with U-Net. The study found that DenseNets&dilation block with U-Net achieved the highest accuracy and sensitivity rates of 95.05% and 90.52%, respectively.Accurate segmentation is crucial for reliable diagnosis and treatment planning in lung cancer using histopathological images, and advanced deep learning methods such as DenseNets & dilation block with U-Net should be used to achieve higher accuracy levels.*

Keywords: - *Deep Learning, Image processing, Lung cancer, Histopathilogical images, Prediction*

I INTRODUCTION

Lung cancer is a malignant growth that starts in the lungs and can spread to other parts of the body. It is one of the most common types of cancer and a leading cause of cancer-related deaths globally. The disease is characterized by uncontrolled cell growth in the lungs, which can eventually lead to the formation of a tumor. If left untreated, lung cancer can spread to other parts of the body, including the lymph nodes, bones, liver, and brain.

Small Cell Lung Cancer (SCLC) and Non-Small Cell Lung Cancer (NSCLC) are two main types of lung cancer, based on the appearance of the cancer cells under a microscope.

SCLC, also known as oat cell carcinoma, is a fast-growing type of lung cancer that tends to spread quickly to other parts of the body. It makes up approximately 10-15% of all lung cancer cases and is usually associated with a heavy history of tobacco use.

NSCLC, on the other hand, is the most common type of lung cancer, accounting for about 85-90% of all cases. NSCLC grows more slowly than SCLC and is more likely to be found at an earlier stage. NSCLC is typically divided into three subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma.

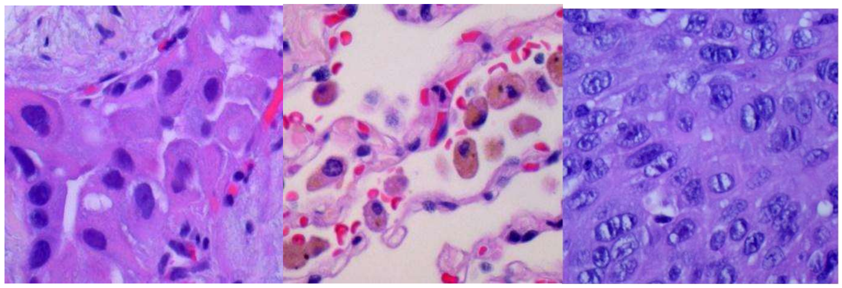

Figure 1: (a) (b) (c)

Figure 1 shows the Histopathological images of lung cancer. (a) lung_aca (lung adenocarcinoma cells), (b) lung_n (lung normal cells), and (c) lung_scc (lung squamous cell carcinoma.

Early detection is crucial for the best chance of survival with lung cancer. Common symptoms of lung cancer include a persistent cough, chest pain, shortness of breath, wheezing, and coughing up blood. If these symptoms are present, it is important to see a doctor for a complete evaluation. Diagnostic tests for lung cancer include chest X-rays, CT scans, PET scans, MRI scans, and biopsies. A biopsy is the most reliable test to diagnose lung cancer, as it involves removing a small piece of tissue from the lung to be examined under a microscope. Following table 1 shows a comparative analysis of various techniques utilized for the detection and segmentation of lung cancer. The table highlights the most effective methods identified through the analysis.

Table 1: Comparative study of different methods used in lung cancer detection and segmentation

Category	CT Scan	PET Scan	X-Ray	MRI	Biopsy
Definition	A medical imaging technique that uses X-rays to create detailed images of the internal structures of the body.	A nuclear medicine imaging technique that uses radioactive tracers to create images of the body's metabolic activity.	A diagnostic imaging technique that uses electromagnetic radiation to produce images of the internal structures of the body.	A medical imaging technique that uses a magnetic field and radio waves to produce detailed images of the internal structures of the body.	A procedure that involves taking a small tissue sample from the affected area for laboratory analysis.
Purpose	To identify the presence and location	To identify the metabolic	To identify the presence and location of	To provide detailed	To obtain a tissue sample for

	of tumors or abnormalities in the lung.	activity of tumors or abnormalities in the lung.	tumors or abnormalities in the lung.	images of the internal structures of the lung.	laboratory analysis and diagnosis.
Technique	X-rays are passed through the body, and the resulting images are captured and processed by a computer.	Radioactive tracers are injected into the body, and a specialized camera detects the radiation emitted by the tracer.	X-rays are directed at the body, and the resulting images are captured on film or digitally.	Magnetic fields and radio waves are used to create detailed images of the internal structures of the body.	A small tissue sample is obtained from the affected area using a needle or surgical instrument.
Radiation exposure	Involves exposure to ionizing radiation, which may increase the risk of cancer.	Involves exposure to small amounts of ionizing radiation from the radioactive tracer.	Involves exposure to ionizing radiation, but the amount is typically low.	Does not involve exposure to ionizing radiation.	No radiation exposure is involved.
Cost	Relatively expensive compared to other imaging techniques.	Relatively expensive compared to other imaging techniques.	Less expensive than CT or MRI scans.	More expensive than X-ray but less expensive than CT scan.	Cost varies depending on the type of biopsy performed.
Availability	Available in most hospitals and clinics.	Available in most hospitals and clinics.	Widely available in hospitals, clinics, and doctor's offices.	Available in specialized imaging centers	Available in hospitals and specialized clinics.

				and hospitals.	
Advantages	Provides detailed images of the lung structures, allowing for accurate diagnosis and staging of lung cancer.	Can detect metabolic activity in the lung, allowing for the detection of small or hidden tumors.	Quick and easy to perform, and widely available.	Provides detailed images of the lung structures without using ionizing radiation.	Provides a definitive diagnosis of lung cancer and can guide treatment decisions.
Disadvantages	Involves exposure to ionizing radiation, which may increase the risk of cancer.	May produce false-positive results, and the images may be difficult to interpret.	Images may not be detailed enough to detect small tumors or abnormalities.	More expensive than X-ray and may not be widely available.	Invasive procedure that carries a small risk of complications such as bleeding or infection.

DIFFERENCE BETWEEN CT, MRI & HISTOPATHOLOGICAL IMAGES

CT (Computed Tomography), MRI (Magnetic Resonance Imaging) and histopathological images are all types of medical imaging techniques used for diagnostic purposes, but they differ in their imaging principles, image characteristics, and clinical applications

CT images: CT images are obtained by using X-rays to create detailed, cross-sectional images of the body. CT scans are used to visualize internal organs, bones, blood vessels, and other structures in the body. CT images have high spatial resolution, meaning they can show small details in the body, and are commonly used for diagnosis of bone fractures, internal bleeding, and cancer

MRI images: MRI images are obtained by using a strong magnetic field and radio waves to create detailed images of soft tissues in the body, such as the brain, spinal cord, and internal organs. MRI images have high contrast resolution, meaning they can show differences between different types of soft tissues, and are commonly used for diagnosis of brain and spinal cord injuries, as well as abnormalities in the liver, kidneys, and other internal organs

Histopathological images: Histopathological images are obtained by examining thin slices of tissue samples under a microscope after staining with various dyes. These images provide detailed information about the microscopic structure of tissues, such as cells, nuclei, and blood vessels, and are commonly used for diagnosis of cancer and other diseases. Histopathological images have high spatial resolution and can reveal cellular and subcellular details of tissue

II SIGNIFICANCE OF DEEP LEARNING

Deep learning is a type of artificial intelligence (AI) that is inspired by the structure and function of the human brain. It is a subset of machine learning.Deep learning algorithms can learn and make decisions or predictions from complex and large datasets without explicit instructions from humans. They are particularly useful for image and speech recognition, natural language processing, and medical imaging.In medical imaging, deep learning algorithms can be trained on large datasets of medical images to detect and segment medical conditions, such as lung cancer. Some of the popular deep learning algorithms used in medical image analysis include Convolutional Neural Networks (CNNs), Fully Convolutional Networks (FCNs), and Recurrent Neural Networks (RNNs). CNNs are designed to handle image data and can automatically learn the features relevant to a given task, such as the detection of lung nodules. FCNs are a variant of CNNs that are used for semantic segmentation, where the goal is to classify each pixel in an image into a specific class, such as lung tissue or lung tumor. RNNs are useful in medical image analysis when the input data has a temporal dimension, such as time-series images acquired from medical imaging modalities such as PET or MRI. There are various approaches to fine-tuning these deep learning algorithms for medical image analysis, such as transfer learning and end-to-end training.

In transfer learning, pre-trained deep learning models on large datasets are used as a starting point, and the final layers are retrained for a specific task, such as lung cancer detection. In end-to-end training, the entire deep learning model is trained from scratch on the target task, such as lung cancer segmentation. Overall, deep learning algorithms have shown promising results in the detection and segmentation of lung cancer from medical images and have the potential to provide accurate and fast results, leading to improved patient outcomes. Deep learning-based approaches in medical imaging have shown promising results and hold great potential for improving patient outcomes by enabling early and accurate detection of conditions such as lung cancer. However, like any other technology, deep learning has its limitations and challenges, such as the need for large and diverse datasets, the risk of overfitting specific datasets, and the need for robustness to handle variations in image quality.

III IMPORTANCE OF IMAGE PROCESSING

Image processing is the process of transforming a digital image from its initial form into a desired output. This involves applying a set of mathematical operations to manipulate the image data in order to improve its quality, extract useful information, or transform it into a desired format. The image processing techniques used in medical imaging include image enhancement, image segmentation, image registration, and image analysis. Image enhancement techniques are used to improve the visibility of an image, while image segmentation techniques are used to divide the image into meaningful regions or segments. Image registration techniques are used to align and compare multiple images of the same scene, while image analysis techniques are used to extract information from images and make quantitative measurements. These techniques can be applied to a variety of medical imaging modalities, such as Histopathological images, X-ray, CT, MRI, PET, and ultrasound images, to aid in the diagnosis and treatment of various medical conditions.

IV RELATED WORKS

Qu, Hui, et al., 2020 proposed a method that combines full image-level labels and partial point annotations of nuclei to train a CNN, generating probability maps for segmented nuclei masks, with promising results on two datasets and potential to reduce annotation effort in histopathology image segmentation.[1].Baranwal, Neha, PreethiDoravari, and RenuKachhoria., (2022) proposed deep learning-based approach using a CNN achieved 93.33% accuracy for the classification of histopathology images of lung cancer, outperforming state-of-the-art methods and potentially assisting pathologists in efficient and accurate diagnosis[2].

Chen, Zhe, et al. 2020 proposed weakly supervised histopathology image segmentation method, using sparse point annotations and an iterative refinement process, achieved state-of-the-art performance on publicly available datasets and has potential to reduce annotation time and cost while maintaining high accuracy. [3].Nishio, Mizuho, et al., 2021 suggested a approach for lung cancer segmentation utilizes transfer learning and a GAN-generated artificial dataset, achieving high accuracy on publicly available datasets and showing potential for improving accuracy with limited training data.[4].

Pawar, Vikul J., et al. 2020 recommend lung cancer detection system uses image processing and ML(Machine Learning) techniques, with a two-stage approach involving lung segmentation and feature extraction, and achieved high accuracy in detecting lung cancer through training and testing four machine learning models.[5]. Table 2 presents an overview of the recent DL (Deep Learning) techniques employed in the detection and segmentation of lung cancer in medical imaging studies. The table provides a detailed analysis of techniques used, advantages,disadvantages and future work for each paper.

Table 2: An overview of Deep Learning Methods for Lung Cancer Detection and Segmentation

Title	Techniques used	Advantages	Dis Advantages	Future work
Coarse-to-Fine Lung Nodule Segmentation in CT Images-(2021)[9]	Dual-branch network, Image Enhancement, Coarse-to-fine approach	Dual-branch network can capture both global and local features to improve segmentation accuracy. Coarse-to-fine approach allows for efficient and accurate segmentation of lung nodules. Image enhancement techniques	Evaluation only performed on one dataset. Requires a GPU for efficient processing. Complexity of the method may limit its generalizability.	Further exploration and optimization of the dual-branch network architecture. Investigation into additional image enhancement techniques. Expansion to multi-modal imaging.

		improve the quality of CT images and assist in nodule detection.		
Automatic Lung Nodule Segmentation and Intra-Nodular Heterogeneity Image Generation-(2022)[10]	GAN Generalized intersection over union Faster-R-CNN LNHG model	The proposed approach offers an end-to-end design that simplifies the segmentation process, thereby making it more accessible for use in clinical practice.	The accuracy of automated segmentation heavily relies on the quality and quantity of the data used for training the algorithm. Limited data can lead to overfitting or underfitting of the model, which can result in inaccurate segmentation.	Expanding the dataset and exploring the network topology of WGAN-GP for improved segmentation accuracy and intra-nodular heterogeneity image evaluation
Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images—The ACDC@LungHP Challenge 2019-(2020)[11]	Deep learning (U-Net architecture), data augmentation, post-processing (thresholding and connected components)	1. Achieved high performance for lung cancer segmentation in whole-slide histopathology images. 2. Fast and efficient segmentation.	1. Limited to the dataset used in the challenge. 2. Limited to the U-Net architecture.	Investigate the use of other deep learning architectures and evaluate the generalizability of the approach on other datasets.

V PROPOSED MODEL

Lung cancer prediction using DenseNet and Dilation Block U-Net is an approach that involves using two deep learning architectures for detecting lung cancer from medical images. The DenseNet architecture is used for feature extraction, while the Dilation Block U-Net architecture is used for segmentation. DenseNet is a deep convolutional neural network architecture that has shown impressive performance in image classification tasks. It works by densely connecting each layer to every other layer in a feed-forward fashion. This creates a network that has fewer parameters than traditional networks, while still being able to extract complex features from the input images.

The Dilation Block U-Net architecture is an extension of the original U-Net architecture, which is a popular deep learning architecture for image segmentation. It uses dilated convolutional layers to expand the receptive field of the network, allowing it to capture more context from the input image. This helps improve the accuracy of the segmentation

Following figure 2 shows the general structure of the proposed system.

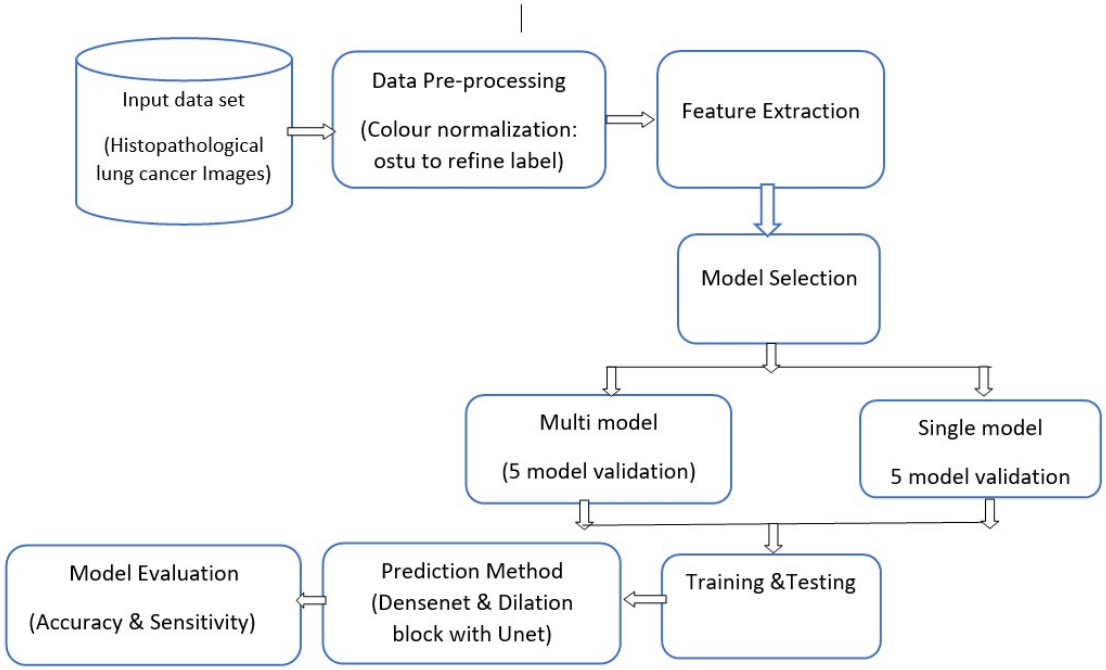

Figure 2: Flow Diagram of Lung Cancer Prediction System

Step 1: The input dataset consists of histopathological images of lung cancer, which can be used for training Deep learning models for automated cancer detection and segmentation.

Step 2: The data preprocessing step involves color normalization of histopathological lung cancer images using Otsu thresholding method to refine the labeling process.

Step 3: CNN and FCN are deep learning methods used for feature extraction in image processing tasks. CNNs are widely used for image feature extraction. They use convolutional layers to extract spatial features from images and pooling layers to reduce the spatial dimensions of the feature maps.

Step 4: Five multi-model approach uses diffetent models to predictions for the final output . Five single-model approach uses different models to predictions for the final output.A single model approach involves building a single deep neural network model that is trained on a given task. The model takes input data and learns a mapping to output the desired predictions. Single models can be trained on various types of tasks, such as classification, segmentation, or regression, and can be implemented with different types of architectures. On the other hand, a multi-model approach involves building multiple deep neural network models that are trained separately on the same task or different tasks. The predictions from each model are then combined in some way, such as averaging or ensembling, to obtain the final prediction. Multi-model approaches can be useful when a single model may not be sufficient to capture all the complexities of the problem, or when different models can be trained on different subsets of the data or different types of features.

Step 5: A total of 10 models were used in the training and testing of the images, including 5 multi-model and 5 single-model approaches. The best performing model was found to be a combination of DenseNet and dilation blocks with a UNet architecture, achieving the highest accuracy and sensitivity.

Convolution, normalisation, and ReLU functions are all combined to create the densenet function [Eq. (1)]. Transferring inputs with zero mean value and unit variance allows for batch normalisation. Then, using the ReLU function, the negative values are changed to zero[23].

$$Densenet(F) = D_l([F, f_1, f_2, \ldots, f_{l-1}]) --------(1)$$

To separate biological images semantically using multi-level dilated residual convolutions. In a multi-level residual of residual connection, each level denoted as L/N[24].

VI RESULTS AND DISCUSSION

The purpose of this study is to evaluate the accuracy and sensitivity of three different image segmentation methods. The methods evaluated in this study include DB U-Net+LLIE, U-net, and DenseNets&dilation block with U-Net. The following sections present the results obtained from our study, followed by a discussion of their implications and potential applications. The findings of this study are expected to provide valuable insights into the effectiveness of different image segmentation methods and can guide the development of more accurate and efficient methods for Lung cancer detection and segmentation. Comparative results with respect to the above-said methods are shown in Table 3. A graphical comparative analysis between the existing methods is shown in figure 2.

Table 3: Comparative analysis of accuracy & sensitivity of the various methods

Various methods	Accuracy (%)	Sensitivity (%)
DB U-Net+LLIE	81.97	84.57
U-net	82.80	87.81
DenseNets&dilation block with U-Net	95.05	90.52

The formula of accuracy & sensitivity is described as follows:

$$Accuracy = \frac{(TP + TN}{(TP + TN + FP + FN)} ----- (2)$$

$$Sensitivity = \frac{TP}{(TP + FN)} ------(3)$$

where:

TP = True Positives (number of instances correctly predicted as positive)

TN = True Negatives (number of instances correctly predicted as negative)

FP = False Positives (number of instances wrongly predicted as positive)

FN = False Negatives (number of instances wrongly predicted as negative)

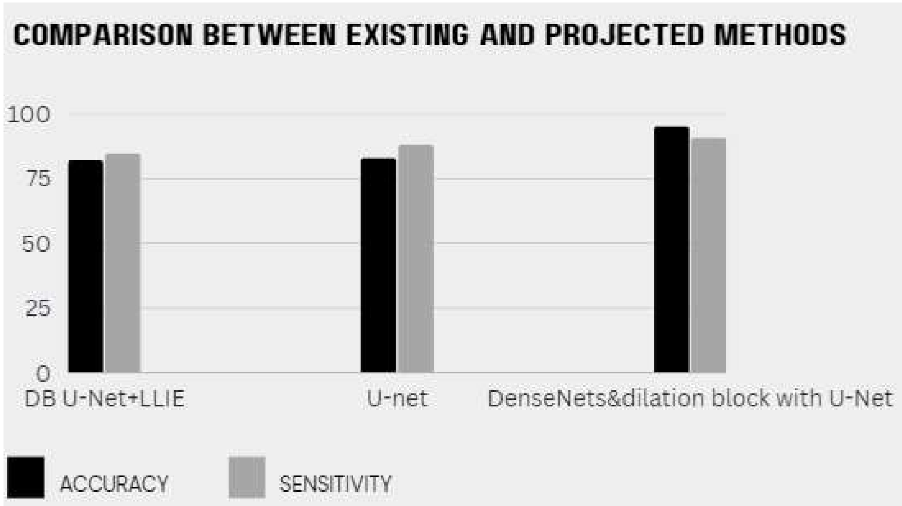

Figure 3: Graphical view of accuracy & sensitivity

This approach can provide accurate and efficient lung cancer prediction using deep learning techniques. However, it requires a large amount of annotated medical image data and computational resources to train and optimize the model

VII CONCLUSION

This study aimed to compare the accuracy and sensitivity of different methods for lung cancer detection and segmentation using various methods. The results show that DenseNets& Dilation block with U-Net achieved the highest accuracy and sensitivity rates of 95.05% and 90.52%, respectively. This finding is particularly noteworthy as accurate segmentation is crucial for reliable diagnosis 7and treatment planning in lung cancer using with histopathological images. The study highlights the importance of using advanced deep learning methods such as DenseNets & dilation block with U-Net. However, additional research is required to gather real-time lung cancer histopathological images and incorporate them into the existing detection and segmentation methods in deep learning techniques to enhance their accuracy level, ultimately improving patient outcomes.

REFERENCES

1. Qu, Hui(2020), "Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images." *IEEE transactions on medical imaging* 39.11, pp. 3655-3666.

2. Baranwal, Neha, PreethiDoravari&RenuKachhoria(2022), "Classification of histopathology images of lung cancer using convolutional neural network (CNN)." *Disruptive Developments in Biomedical Applications*. CRC Press, pp. 75-89.

3. Chen, Zhe, Zhao Chen, Jingxin Liu, QiangZheng, Yuang Zhu, YanfeiZuo, Zhaoyu Wang, Xiaosong Guan, Yue Wang& Yuan Li(2020). "Weakly supervised histopathology image segmentation with sparse point annotations." *IEEE Journal of Biomedical and Health Informatics* 25, no. 5, pp. 1673-1685.

4. Nishio, Mizuho, Koji Fujimoto, Hidetoshi Matsuo, ChisakoMuramatsu, Ryo Sakamoto, & Hiroshi Fujita(2021), "Lung cancer segmentation with transfer learning: usefulness of a

pretrained model constructed from an artificial dataset generated using a generative adversarial network." *Frontiers in Artificial Intelligence* 4, 694815.

5. Pawar, Vikul J., Kailash D. Kharat, Suraj R. Pardeshi&Prashant D. Pathak(2020). "Lung cancer detection system using image processing and machine learning techniques." *Cancer* 3 (2020): 4.

6. Han, Yong, Yuan Ma, Zhiyuan Wu, Feng Zhang, DeqiangZheng, Xiangtong Liu, Lixin Tao(2021), "Histologic subtype classification of non-small cell lung cancer using PET/CT images." *European journal of nuclear medicine and molecular imaging* 48, pp 350-360.

7. Srivastava, Shrey, Amit VishvasDivekar, ChanduAnilkumar, IshikaNaik, Ved Kulkarni, &V. Pattabiraman(2021). "Comparative analysis of deep learning image detection algorithms." *Journal of Big Data* 8, no. 1, pp. 1-27.

9. Wu, Zhitong, Qianjun Zhou, and Feng Wang(2021), "Coarse-to-fine lung nodule segmentation in CT images with image enhancement and dual-branch network." *IEEE Access* 9 , pp. 7255-7262

8. Malathi, M., P. Sinthia, U. Madhanlal, K. Mahendrakan& M. Nalini(2022). "Segmentation of CT Lung Images Using FCM with Active Contour and CNN Classifier." *Asian Pacific journal of cancer prevention: APJCP* 23, no. 3, pp. 905-910.

10.Song, Jiangdian, Shih-Cheng Huang, Brendan Kelly, Guanqun Liao, Jingyun Shi, Ning Wu, Weimin Li(2021) "Automatic Lung Nodule Segmentation and Intra-Nodular Heterogeneity Image Generation." *IEEE Journal of Biomedical and Health Informatics* 26, no. 6, pp. 2570-2581.

11. Li, Zhang, Jiehua Zhang, Tao Tan, XichaoTeng, Xiaoliang Sun, Hong Zhao&Lihong Liu (2019), "Deep learning methods for lung cancer segmentation in whole-slide histopathology images—the acdc@ lunghp challenge 2019." *IEEE Journal of Biomedical and Health Informatics* 25, no. 2, pp 429-440.

13. Li, Min, Xiaojian Ma, Chen Chen, Yushuai Yuan, Shuailei Zhang, Ziwei Yan &Cheng Chen(2021). "Research on the auxiliary classification and diagnosis of lung cancer subtypes based on histopathological images." *Ieee Access* 9, pp. 53687-53707.

14. Puttagunta, Muralikrishna& S. Ravi. "Medical image analysis based on deep learning approach." *Multimedia tools and applications* 80 (2021): 24365-24398.

15. Shimazaki, Akitoshi, Daiju Ueda, Antoine Choppin, Akira Yamamoto, Takashi Honjo, Yuki Shimahara&Yukio Miki(2022). "Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method." *Scientific Reports* 12, no. 1, 727.

16.Uzelaltinbulat, Selin&BuseUgur(2017). "Lung tumor segmentation algorithm." *Procedia computer science* 120, pp. 140-147.

17.Wang, Xi, Hao Chen, CaixiaGan, Huangjing Lin, Qi Dou, EfstratiosTsougenis, Qitao Huang, MuyanCai&Pheng-Ann Heng(2019). "Weakly supervised deep learning for whole slide lung cancer image analysis." *IEEE transactions on cybernetics* 50, no. 9, pp. 3950-3962.

18.Gami, Bhavin, Khushi Chauhan&Brijeshkumar Y. Panchal(2023), "Breast Cancer Detection Using Deep Learning." *Mobile Radio Communications and 5G Networks: Proceedings of Third MRCN 2022*. Singapore: Springer Nature Singapore, pp. 85-95.

19.Vijayaraj, J(2021). "Various segmentation techniques for lung cancer detection using CT images: a review." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.2, pp. 918-928.
20.Senthil Kumar, K., K. Venkatalakshmi&K. Karthikeyan(2019). "Lung cancer detection using image segmentation by means of various evolutionary algorithms." *Computational and mathematical methods in medicine* 2019.
21.Rahane, Wasudeo, HimaliDalvi, YaminiMagar, Anjali Kalane&SatyajeetJondhale(2019). "Lung cancer detection using image processing and machine learning healthcare." In *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, pp. 1-5. IEEE, 2018.9.
22.Wang, Shidan, Donghan M. Yang, RuichenRong, Xiaowei Zhan, Junya Fujimoto, Hongyu Liu, John Minna, Ignacio Ivan Wistuba, Yang Xie&Guanghua Xiao(2019), "Artificial intelligence in lung cancer pathology image analysis." *Cancers* 11, no. 11, 1673.
[23] Nancy Girdhar, Aparna Sinha & Shivang Gupta(2022), “DenseNet-II: an improved deep convolutional neural network for melanoma cancer detection“,Soft Computing , Springer
[24] Naga RajuGudhe, Hamid Behravan, MazenSudah, Hidemi Okuma, RitvaVanninen, Veli-MattiKosma & ArtoMannermaa (2021), “Multi-level dilated residual network for biomedical image segmentation, “Scientific Reports, Vol. 11.