Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 4, April 2024

Detection and Classification of Lung Nodules Using Deep Learning Methods

Suchitra M, Trupti Manjunath Patgar, Varsha S A
Students, Department of Information Science and engineering
BMS Institute of Technology and Management
Bengaluru, India

Dr. M V Sudhamani
Professor, Department of Information Science and Engineering
BMS Institute of Technology and Management
Bengaluru, India

***Abstract*— The aim of our work is to present an approach towards early detection of abnormalities in the lungs, which is crucial for advancing human health. Leveraging Convolutional Neural Networks (CNN) and sophisticated image processing techniques such as Segmentation and Feature extraction, the primitive objective of this work is to develop an automated system for accurately classifying cases of lung cancer. With a subset of dataset obtained from LUNA 16 which consists of 800 CT Lung images, comprising 563 for training and 237 for testing, this work employs a pretrained CNN model that has demonstrated good accuracy. Through image segmentation and nodule identification using the watershed algorithm, the model is trained to discern between non-cancerous or benign and cancerous or malignant anomalies within lung structures. The resulting system boasts impressive performance metrics, with a precision of 94.2%, indicating a high proportion of true positive predictions among all positive cases, and a recall of 94.2%, signifying the model's ability to accurately captures positive instances from all the actual positive samples from the dataset. Moreover, with an overall accuracy of 97.1%, the system demonstrates exceptional proficiency in classifying cases across both positive and negative classes. Additionally, the integration of a user-friendly interface enhances accessibility, enabling medical professionals to submit CT images for automated cancer prediction, thereby facilitating timely interventions and ultimately improving patient outcomes.**

Keywords*— *Lung cancer, Convolutional Neural Networks (CNN), Deep Learning (DL), Computed Tomography (CT)

I. INTRODUCTION

Lung cancer is the second most common cause of death in the world, after cardiovascular diseases, and lung cancer ranks as the second most deadly disease. According to American Cancer Society, there were 1.762.450 new cancer cases in the United States in 2019, 13% of which were lung cancer, with 228,150 new cases [2]. In the United States, lung cancer ranks first in terms of cancer-related mortality, accounting for 24% [2] [3]. These figures highlight the need for more advanced lung cancer treatments. The lungs, the primary organs of the respiratory system, are segmented into lobes, with the right lung comprising three lobes, slightly larger than the left lung with two lobes. Separated by the mediastinum, this region houses the heart, trachea, oesophagus, and numerous lymph nodes. Encased in a safeguarding membrane called the pleura, the lungs are shielded from the abdominal cavity by the muscular diaphragm.

Chronic lung disease can affect human health. Figure 1 shows that abnormalities in certain anatomical regions can lead to lung disease like cancer.

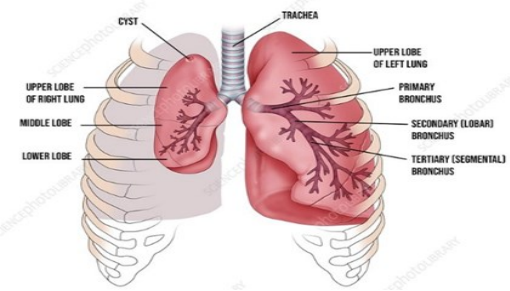


Figure 1: Anatomy of Lung organ (https://www.sciencephoto.com/)

The lungs are surrounded by a protective membrane known as the pleura, which protects them from the abdominal cavity.
Lung abnormalities can have significant effects on human health. Figure 2 shows how abnormalities in specific anatomical regions can significantly affect lung health, including lung cancer.

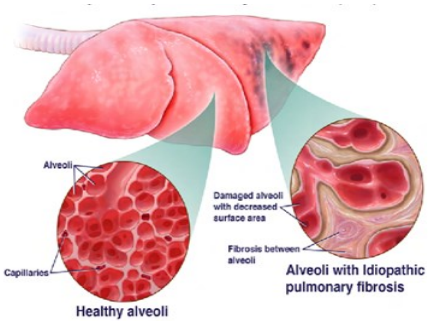


Figure 2: Normal vs Affected alveoli (https://www.semanticscholar.org/)

The main airways to the lungs, bronchioles (small airways) and alveoli (tiny air sacs that contain oxygen), are areas of the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 4, April 2024

lung where lung cancer can spread. In most cases, lung cancer begins in cells in the bronchi and other structures in the lungs. Abnormal and uncontrolled growth of epithelial cells can lead to lung dysfunction, which often leads to these cancers. As cancer, it can enter neighboring tissues and spread to other organs through higher-level lymph nodes or blood vessels. Rapid identification and understanding of many cell types involved are essential for effective discovering and medical care of lung malignancies. DL uses deep neural networks to gain insights from data. The stage of the cancer indicates how far the disease has spread. The terms "stage I" and "stage II" apply to breast cancer, while the term "advanced" refers to cancer that has spread to other organs. Today, blood biopsies and imaging practices such as computed tomography (CT) and x-ray scans are used to make a diagnosis.

Early diagnosis of lung cancer is challenging because there are very less symptoms present. Lung cancer can spread to other diseases. The word "metastasis" refers to the way cancer spreads to different parts of the body. The difference between damaged and healthy alveoli is shown in Figure 3. To solve this important health problem, our research uses CNN, DenseNet, SVM and YOLOv5 deep learning models.
Measurements such as recall, accuracy, precision, and F1 score were used to estimate the efficacy of this model in identifying pulmonary nodules. Using data representing real-world situations, the study sought to determine which machine learning framework was most effective at saving lives. These results provide key facts to doctors and help improve early and reliable detection of pulmonary nodule.

II. LITERATURE SURVEY

Lately, DL has unfolded as a promising avenue for improving identification of lung cancer. Leveraging complex neural networks and sophisticated algorithms, DL models exhibit the capability to amplify the accuracy and efficiency of early detection methods. This literature survey explores the evolving landscape of lung cancer detection, examining key advancements, challenges, and the transformative impact of artificial intelligence in revolutionizing diagnostic approaches for this critical health concern.
In paper [1], using CT scan images, the process of detecting lung cancer consisted of two stages: feature recognition and machine learning application. Relevant features are taken out of the CT scan images in the initial stage. Techniques including feature extraction, segmentation, and image preprocessing might be used for this. During the second stage, the retrieved features are classified as either non-cancerous or malignant by utilizing machine learning techniques. There are several machine learning strategies that have been proposed for identifying lung cancer, including SVM, Naive Bayes, artificial neural networks, and decision trees.The study suggests a computer-aided system that makes use of deep learning techniques for lung cancer detection in order to lessen this difficulty in [2]. Utilising the AlexNet architecture, a CNN technique is used to analyse the input dataset that was collected from hospitals in Iraq. The introduced model demonstrates high accuracy, reaching up to 93.548%, showcasing its effectiveness in distinguishing between normal, benign, and malignant cases. Further performance metrics highlight the agility of the suggested model, with sensitivity at 95.714% and specificity at 95%. This study represents a prominent advancement in the operation of AI for early discovery, which will increase the survival estimation of lung cancer patients.

Used CT images from the Lung Image Database Consortium (LIDC) to classify lung cancer using Deep Convolutional Neural Networks (DCNN) in [3]. Improving the precision of distinguishing between malignant and noncancerous lung nodules was the aim, as this is a critical stage in effective lung cancer therapy. Robust identification of lung nodules was difficult ascribed to their complexity. The study aimed to bloom the accuracy of lung cancer identification and recognition from CT images by using DCNN to outperform current methods in this regard. This underscores the potential of advanced machine learning techniques.

Given the difficulty of diagnosing lung cancer at a later stage, [4] is concentrated on the problem of early identification of lung cancer using LDCT images, which is essential for improving survival rates. LDCT scans offer lower X-ray doses but poorer image quality compared to normal CT scans. To address this, the study proposed a methodology leveraging both machine learning and deep learning, specifically employing CNN for attribute selection and categorization. The capability for early cancer identification with LDCT images was enhanced by the CNN technique, which demonstrated promising results in nodule identification and categorization. The study also covered available datasets, illuminating potential paths and obstacles in the field, and pointing researchers in the direction of more efficient approaches for early lung cancer identification using LDCT images.

The proposed method in [5] comprises three stages: preprocessing, segmentation, and classification. To preserve image sharpness for precise segmentation, preprocessing uses a geometric mean filter to eliminate noise from CT scans. In segmentation, lung nodules are identified by pixel similarity using K-means clustering because nodules are generally brighter and more homogeneous than the surrounding lung tissue. Artificial Neural Networks (ANN) are used in classification to categorize nodules as benign or malignant. ANN are best for nodule categorization because of their capability to learn complex patterns from data, which is inclined by the system of the human brain. The approach achieved a 95% classification accuracy for lung nodules when tested on an input dataset of CT scan from patients with lung cancer. Comparing this to conventional diagnostic techniques, which typically yield accuracy levels of 80% or higher, shows a significant improvement.
The six-stage design in [6] involves symptom-based lung cancer identification using a Random Forest (RF) Classifier, CT scan classification with a CNN, and nodule detection making use of a UNet model. Additionally, a RF Regressor predicts medical insurance costs. The web application

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 4, April 2024

incorporates interactive Plotly graphs for data analysis and provides a user-friendly interface with four main buttons for different functionalities. The system achieved high accuracies, including 96.9% for symptom-based detection, 92.42% for CTsnapshot classification, and 98% for nodule detection. The proposed strategy, encapsulated in a Flask web application, promises accurate results in real-time, making it an organized aid for lung cancer identification and cost estimation. The study also details data processing steps and emphasizes the application's user-friendly design and accessibility.

Leveraging the capability of CNNs to pull out critical attributes from images, the project involves the analysis of CT scan slices to establish a machine learning model. The 3D CNN are employed in [7] to discover the existence of cancer by evaluating and preprocessing the data. The goal is to identify and address cancerous cells at their earliest stages. The research utilized the “Cancer Imaging Archive” dataset, implementing a pre-processing step to standardize the size and format of CT images before model training. The resulting model demonstrated a high accuracy of 93%, and the analysis metrics, including Precision, Recall, Kappa, and F1 score, further underscore the model's effectiveness in categorizing lung nodules as either cancerous or non-cancerous. With precision at 0.68669, recall at 0.64384, kappa score at 0.39733, and F1 score at 0.62699, the model exhibits promising performance in lung cancer identification by using 3D CNN.

Employing the LIDC/IDRI dataset from the Lung Nodule Analysis (LUNA16) challenge, the CNN version is built following a successive approach with convolutional loops, max pooling layer, flattening layer, fully connected layer, and the output layer in [8]. The algorithm, named CNN-based Automatic Lung Cancer Detection (CNN-ALCD), is designed for supervised learning, and demonstrates the capability to diagnose lung cancer from newly arrived test samples. The suggested CNN model, trained and tested on an input dataset of 343 lung CT images, exhibits a correctness of 94.11%, surpassing existing designs like ANN (90.24%) and Multilayer Perceptron (MLP) (92.12%) in predicting lung cancer. The study emphasizes performance evaluation using metrics such as accuracy and confusion matrix.

There are limitations in the utility of thoracic radiography, particularly chest X-rays, due to a shortage of skilled radiologists. To prevail over this obstacle, the paper [9] suggests the utilization of a modified model, MobileNet V2, for the categorization and prediction of frontal thoracic X-rays. Highlighting the probable life-saving impact of (CT) in detecting tumours early, the study highlights the load on radiologists in analysing a large volume of CT lung images and the associated observer fatigue. The proposed technique, evaluated using the NIH Chest-Xray-14 database, outperforms contemporary pathology classification algorithms, with an AUC of 0.811 and an accuracy exceeding 90%. Notably, resampling the dataset remarkably improves the model's performance, aiming to outline a model that is easily trainable, requires less computational energy, and hence could be deployed on smaller IoT devices.

To prevail over the flaws of prior lung cancer identification, the paper [10] suggested Deep Ensemble 2D CNN model which achieves better accuracy, precision, and recall because it is specifically made for lung tumour identification using CT scan data. The productiveness of the design is attributed to its role to extract relevant information via CNN blocks and merge predictions from numerous deep neural networks, hence augmenting the classification precision overall.

In summary, the literature survey emphasizes the significant progress made in deploying deep learning (DL) techniques for the identification of lung cancer through medical imaging, with a focus on CT scans and X-rays. The incorporation of various machine learning strategies and sophisticated DL architectures, alongside the growth of computer-aided systems, reflects substantial advancements in early detection capabilities. The utilization of innovative methods, including CNNs, SVMs, and ensemble models, showcases promising accuracy rates, indicating the potential to upgrade patient issues through timely identification and intervention. Despite persistent challenges such as limited datasets and the need for enhanced generalizability, the ongoing evolution of these techniques holds great promise in transforming lung cancer diagnostics and enhancing survival rates. The convergence of AI and medical imaging marks a significant frontier in healthcare, signifying a crucial shift towards more effective and precise diagnostic tools for lung cancer. The survey notably highlights diverse approaches encompassing attribute recognition, machine learning optimization, and the utilization of advanced neural networks like CNNs, DenseNet, and YOLOv5.

III. PROPOSED WORK

In this proposed work, CNN play a decisive role in the detection of lung cancer. Leveraging the potential of CNNs, this work aims to analyze medical imaging data to recognize potential cancerous nodules within the lungs. CNN is renowned for their aptitude to pull out complex properties from images, essence them absolute for jobs such as image classification and segmentation. By training a CNN model on a dataset of CT lung images, the system can learn to distinguish between cancerous or malignant and non-cancerous or benign nodules with a high degree of accuracy. Through many layers of convolutional and pooling operations, the CNN can automatically detect subtle patterns indicative of cancerous growth.

The dataset utilized in this work comprises 800 CT lung images, with 563 images designated for training purposes and 237 images for testing as shown in Figure 3. These images are crucial for training and evaluating the performance of the proposed deep learning model in detection of lung cancer. Each image in the dataset represents CT lung image data capturing various aspects of lung structures and potential nodules.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 4, April 2024

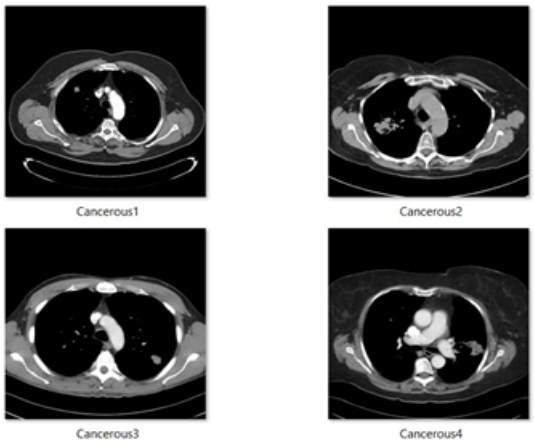


Figure 3: Sample of dataset

The watershed algorithm is employed for the segmentation of both lung structures and nodules from lungs within CT imaging data. The watershed algorithm is particularly effective in segmenting regions of interest by exploiting the gradient information existing in an image. Initially, the lungs are segmented from the CT images, ensuring the precise delineation of lung structures.

Watershed Algorithm:

- Gradient Calculation: We compute the image gradient to pinpoint areas exhibiting significant intensity changes, decisive for recognizing potential anomalies.
- Marker Selection: Initial markers or seeds are strategically chosen within the image, typically representing areas of interest or regions targeted for segmentation.
- Flood Fill: Through a progressive flooding process initiated from the markers, neighbouring pixels with akin intensity are methodically merged, aiding in delineating distinct regions.
- Basins Formation: As the flooding procedure advances, basins or catchment areas naturally evolve around each marker, effectively describing regions of uniform intensity.
- Segmentation: Ultimately, the boundaries of these basins serve as the segmentation framework, facilitating the precise delineation of segmented regions within an image.

IV. RESULTS AND DISCUSSIONS

The segmentation of lung structures and nodules, followed by the classification of nodules as cancerous or non-cancerous, yields critical insights for early identification of lung cancer. By accurate representation of lung boundaries and identifying nodules indicative of potential malignancies, the segmentation process as shown in Table 1 lays the groundwork for subsequent analysis.

In the upraisal of the training process, a confusion matrix works as a fundamental tool for upraising the execution of the classification model. In this specific scenario, where 563 training samples were utilized, comprising 296 cancerous and 267 non-cancerous instances, the confusion matrix provides a compact summary of the model's predictions. It delineates four essential metrics: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True positives represent correctly identified cancerous cases, while true negatives denote accurately classified non-cancerous instances. False positives signify non-cancerous cases erroneously classified as cancerous, and false negatives indicate cancerous instances misclassified as non-cancerous. With an accuracy of 97%, precision of 94%, and recall of 94%, the model demonstrates strong performance in accurately classifying instances of cancerous and non-cancerous cases during training.

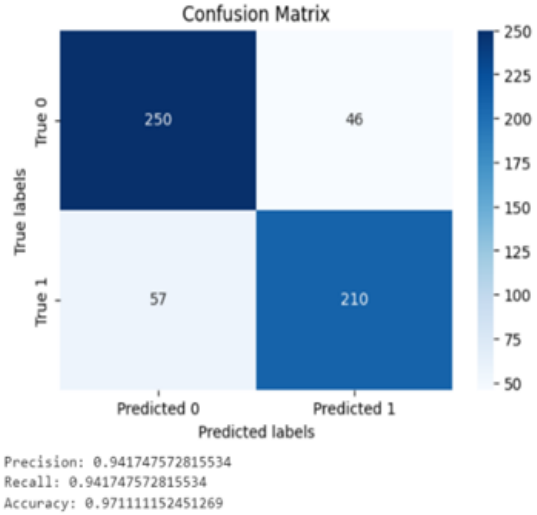


Figure 4: Confusion Matrix

On completion of segmentation and training of the model, a user interface is designed to classify CT image submitted by the user a Benign or Malignant.

In In Figure 5, we can observe the preview page designed for users to upload CT scan images for prediction within the lung nodule classification system. This page offers a user-friendly interface, allowing individuals to easily select and upload their CT image for prediction.

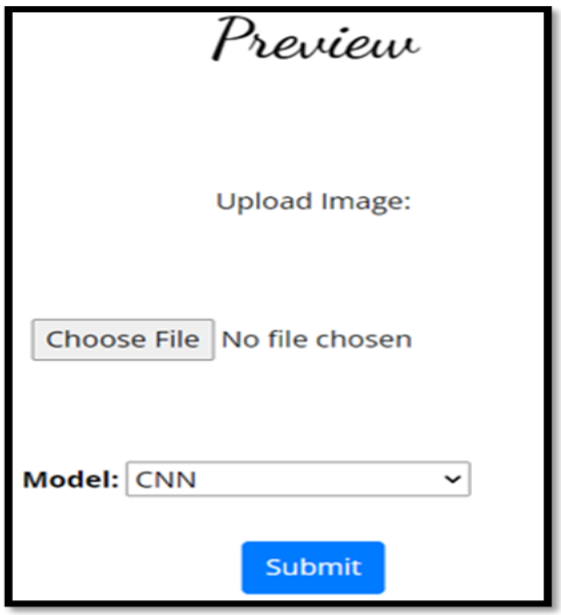


Figure 5: Interface to submit the image from the folder

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 4, April 2024

In Figure 6, we see the prediction page, where the uploaded CT scan image undergoes classification within the lung nodule classification system. The prediction page works as a pivotal resource for healthcare professionals, offering valuable insights to support clinical determining and patient management.

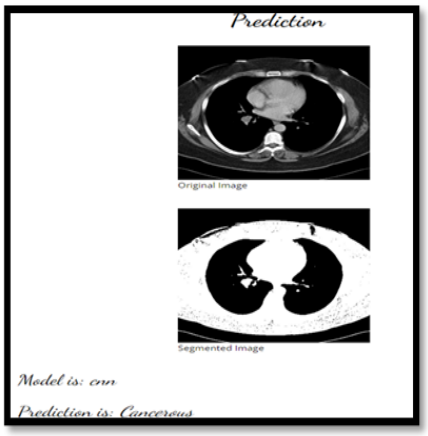
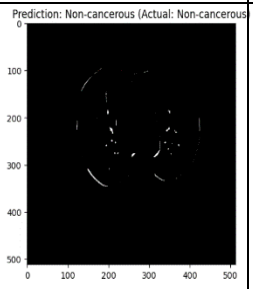
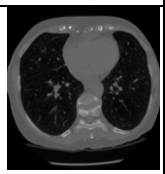
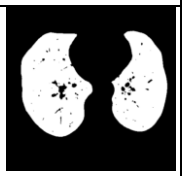
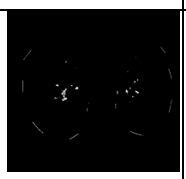
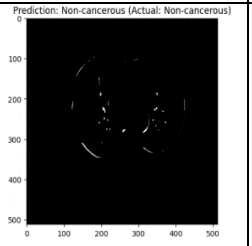
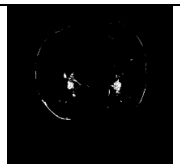
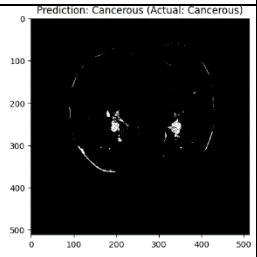
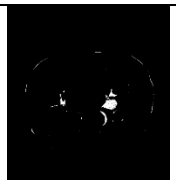
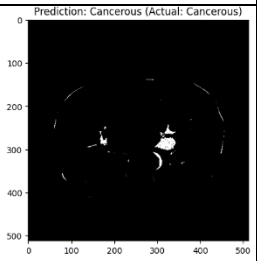


Figure 6: The results of prediction displayed on the interface

RESULTS OF LUNG AND NODULE SEGMENTATION

TABLE I.

Test Case	Input	Segmented Lung	Nodule Segmented	Expected Output	Output	Status
1				Non-Cancerous		CNN classification as non-cancerous- Pass
2				Non-Cancerous		CNN classification as non-cancerous - Pass
3				Cancerous		CNN classification as cancerous Pass
4				Cancerous		CNN classification as cancerous -Pass

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 13 Issue 4, April 2024

V. CONCLUSION

The proposed work constitutes a significant improvement in the province of medical image processing and early identification of lung cancer. By leveraging deep learning techniques and a dataset comprising 800 images, including both training and testing subsets, this work demonstrates a comprehensive approach to automated lung cancer detection. Through the blending of CNN and sophisticated image processing methods such as segmentation and feature extraction, the system effectively analyzes affected area within lung images to identify early-stage cancerous nodules. The implications of this work extend to human health, offering a sophisticated decision support system that can facilitate timely medical interventions and in time boost patient issues. By automating the job of classifying lung nodules as cancerous or non-cancerous, streamlines the diagnostic process and enhances accessibility to early cancer detection. The forming of a user-friendly interface further enhances the usability of the system. With an accuracy of 97%, precision of 94%, and recall of 94%, the model demonstrates strong performance in accurately classifying instances of cancerous and non-cancerous cases during training. In future, various other deep learning models can be explored to have still more accuracy.

REFERENCES

[1] Eali Stephen Neal Joshua, Midhun Chakkravarthy, Debnath Bhattacharyya, An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques: A Systematic Study (Revue d'Intelligence Artificielle, Vol. 34, No. 3, 2020).

[2] Hamdalla F. Al-Yasriy, Muayed S. AL-Husieny, Furat Y. Mohsen, Enam A. Khalil, Zainab S. Hassan, Diagnosis of Lung Cancer Based on CT scans Using CNN (ISCAU doi:10.1088/1757-899X/928/2/022035, 2020).

[3] Amjad Khana, Zahid Ansari, Identification of Lung Cancer Using Convolutional Neural Networks Based Classification (Turkish Journal of Computer and Mathematics Education, Vol.12 No.10, 2021).

[4] Gagan Thakral, Sapna Gambhir, Nagender Aneja, Proposed methodology for Early Detection of Lung ancer with low-dose CT scan using Machine Learning (International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON, 2022).

[5] Sharmila Nageswaran, G. Arunkumar, Anil Kumar Bisht, Shivlal Mewada, J. N. V. R. Swarup Kumar, Malik Jawarneh, and Evans Asenso, Lung Cancer Classification and Prediction Using Machine Learning and Image Processing (BioMed Research International Volume 2022, Article ID 1755460, 2022).

[6] Gayathri Devi Nagalapuram, Varshashree, Vansika Singh, Dheeraj, Donal Jovian Nazareth, Dr. Savitha Hiremath, A Web Application for Lung Cancer Analysis and Detection using Deep Learning Approach (International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10 Issue VI, 2022).

[7] Soundari D V, Poongodi C, Sanjay Kumar V G, Srijith Raj S, Srikanth V, Lung Cancer Detection using 3D-Convolution Neural Network (Smart Technologies, Communication and Robotics (STCR) IEEE DOI: 10.1109/STCR55312.2022.10009146, 2022).

[8] Mattakoyya Aharonu, R Lokesh Kumar, CNN based Framework for Automatic Lung Cancer Detection from Lung CT Images (International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), 2022).

[9] M. Praveena, A. Ravi, T. Srikanth, B. Hari Praveen, h B. Sai Krishna, A. Sunil Mallik, Lung Cancer Detection using DL Approach CNN (International Conference on Communication and Electronics Systems (ICCES), 2022).

[10] AsgharAli Shah, HafzAbid Mahmood Malik, AbdulHafeez Muhammad, AbdullahAlourani & Zaeem Arif Butt, Deep learning ensemble 2D CNN approach towards the detection of lung cancer (https://doi.org/10.1038/s41598-023-29656-z, 2023).