Lung Cancer Detection using Deep Learning on CT Scans

Paripelli Venkata Shiva Sai
Computer Science Engineering
Lovely Professional University
Phagwara, Punjab
shivasaii788@gmail.com

Dr. Soni Singh
Lovely Professional University
Phagwara, Punjab
sonisingh0107@gmail.com

***Abstract*— Lung cancer has consistently been a deadly disease among the illnesses that affect the survival of humans. It is also among the most prevalent tumors, which is an aggressive and highly powerful disease. In its early stages, the condition is largely asymptomatic, which makes detection extremely difficult. For this reason, identification and figuring out the initial phase of cancer is critical to save the survival chance of humans. A tolerant person can have a better prognosis and retrieval rate with treatment beforehand. Even though CT scans are superior to other imaging methods for identifying lung cancer, viewing these images with the unaided eye can occasionally result in errors and a delay in the cancer's detection. For this reason, image processing techniques are frequently used in the preventive industry as a tool for early observation of lung cancers. Several researchers have methods that are derived based on machine learning, deep learning, and image processing to predict the severity of cancer. In this study, we will primarily list, describe, compare, and analyze numerous approaches to segmentation of images, extracting the features after the segmentation, and application of various methods and approaches to initial-stage lung cancer detection & classification.**

Keywords*— Lung Cancer, Machine learning, Image processing, Disease detection.

I. INTRODUCTION

The most serious ailment in the world, according to medical experts, is cancer. Cancer affects both men & women in many different forms. One of the most common malignancies to be detected is cancer in lungs, which is also one of the principle reasons for the demise of humans that is taking place globally. Each year, 2.20 million new people are diagnosed with lung cancer; of those, 75% pass away within five years of receiving a diagnosis [1]. India has a count of more than 70,000 cases in a year [2]. As per the study of United States, lung cancer takes the second place in most popular malignant tumors [3]. Lung nodules contain 2 components: malignant and benign [4], where malignant cells produce cells that are more dangerous and proliferate little amounts inside lungs of human beings while benign is known to be non-cancerous. For humans to survive, it is imperative that we identify malignant cancer cells as soon as possible. However, with regard of cancer cell position, shape, & arrangement - benign & malignant modules are more alike but might occasionally diverge from one another. The difficult part of this calculation is figuring out how likely it is that the cancer will be this advanced. Given these similarities, distinguishing between malignant and benign cancer remains a difficult task [5]. Nodules are categorized as malignant if their diameter exceeds 3 millimeters and as benign if it is less than 3 millimeters [3].

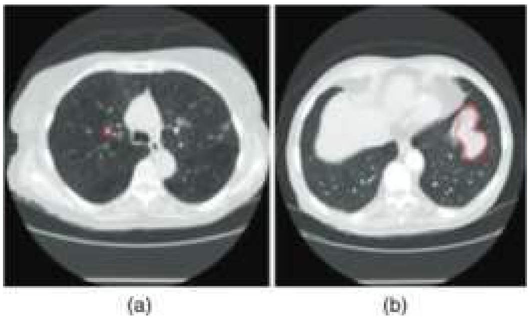

Fig. 1. 2 Different Lung nodules (a) Benign & (b) Malignant [4].

Early lung cancer identification significantly improves a person's chances of survival. It is essential to support doctors in making educated decisions during therapy to help them identify lung cancer early on. Early identification is crucial for prompt and effective therapy, which ultimately enhances the overall outcome for the patient [6]. Cancer survival rates may increase to roughly 55% if the disease was identified early enough to be diagnosed. Moreover, it has been reported that after five years, people in the early stages who receive the right care may have a 40% chance of surviving [7]. The extent of cancer dissemination is correlated with cancer stage. It is frequently ascertained by combining different imaging modalities with pertinent tissue biopsies, which aid in determining the cancer kind assistance using CT scan images [8]. Furthermore, it is advantageous to use Computer-Aided Diagnosis (CAD). Radiologists and pulmonologists began using CT (computed tomography) pictures of the chest [9].

These days, lung nodules with cancer can be diagnosed with great tools to undertake an automatic assessment of pulmonary nodules in order to help medical decision making.

The computer Aided Diagnosis (CAD) is an implementation, which is a practical approach for medical imaging in today's world and an effective medical diagnosis tool [9].

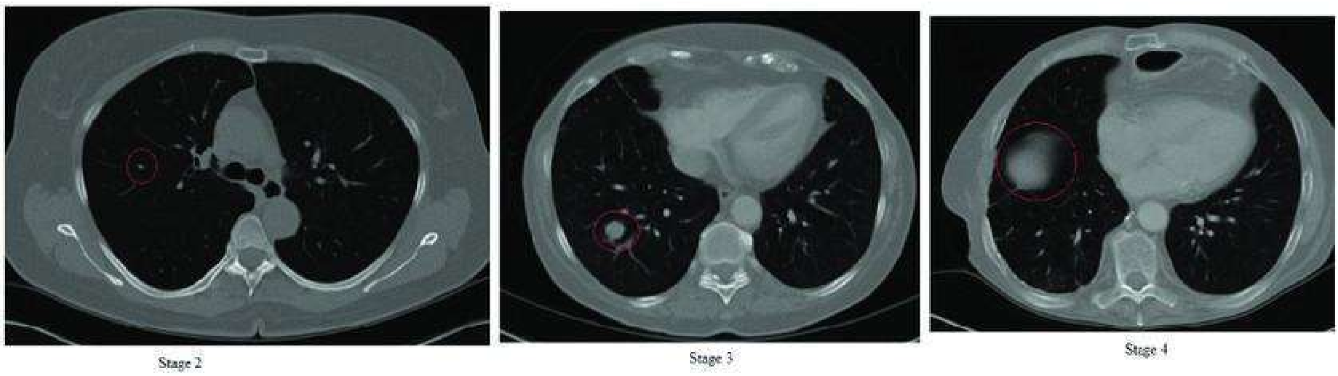

Fig. 2. Different stages of cancer growth [10].

Sophisticated X-ray technology is used during a CT scan to obtain images of the human body from various perspectives. After that, a computer is fed the images and given instructions to analyze them in a way that creates a fragmentary depiction of a person's body, inner organs, and twills [11].

II. LEAD OF MACHINE LEARNING IN LUNG DISEASE DETECTION

Machine Learning (ML) algorithms offer new methods for analyzing complicated medical data and identifying patterns that humans would miss in the fight against lung cancer, which gives hope to those fighting the disease. These algorithms comb through enormous volumes of data from imaging devices, such CT scans, and X-rays, to find minute anomalies that might point to lung cancer. We can get a follow to detect and treat lung cancer early using these potentials. Some of the major Deep Learning algorithms like CNN, DNN, SAE etc. can be used widely for detection [12].

Image processing is a main root in the process of predicting and detecting lung cancer, because it gives valuable information about the data, which is more helpful in the medical field for further studies. Using modern technologies like CT scans, MRIs, and X-rays, we can gather all the information relevant to detection, from the detailed images in which image processing plays a major role. But on the other side these techniques also have some drawbacks like failed detection pf lesions and high false results [13]. Image processing techniques are able to identify minute alterations, aberrations, or growth patterns in lung tissues that could indicate the existence of cancer cells in lungs. Healthcare practitioners can contribute for the improvement of patient outcomes by identifying and evaluating lesions early on by utilizing advanced image analysis techniques.

This will help with timely interventions. In an effort to better understand and classify lung cancer, scientists and researchers have been experimenting with various models and methodologies. They have developed a number of

techniques to recognize and anticipate lung cancer. Using deep learning, which reduces computing complexity to increase the efficiency of cancer identification in lungs from CT images, is one possible way.

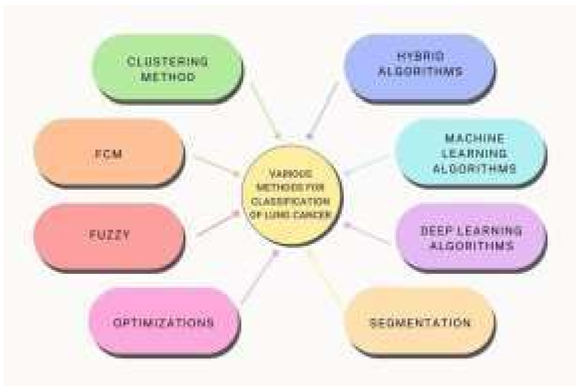

Fig. 3. Various algorithms for cancer prediction.

Certain models incorporate the previously described hybrid approaches, which have exertion of a single method. In order to improve our demonstrated superior performance compared to the understanding of this crucial field, we explore segmentation, fuzzy systems, clustering, machine learning, and new advances in deep learning as they relate to lung cancer diagnosis.

This research inspects the impact of techniques in machine learning specifically created for detecting lung cancer. We will examine a stretch of machine learning techniques, like assembling the methods, deep learning, and feature selection algorithms, to investigate their potential to disrupt existing diagnostic methodologies. The main motive is to demonstrate how ML algorithms could enhance current diagnostic techniques, leading to better outcomes for lung cancer patients, by examining recent developments and upcoming difficulties.

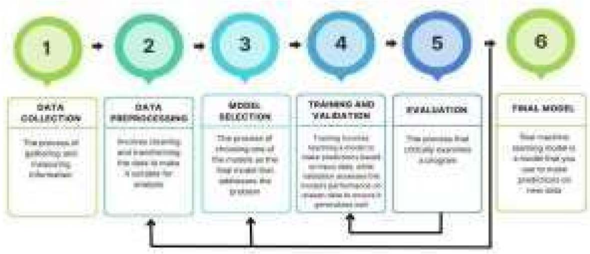

Fig. 4. Work flux of picture processing.

These test cases cover various functionalities, scenarios, and aspects of the deep fake video detection system, ensuring comprehensive testing to validate its effectiveness, reliability, and adherence to ethical and security standards. Test cases can be expanded and customized based on specific project requirements and potential edge cases.

III. LITERATURE SURVEY

REFERENCE	YEAR	METHODS	METRICS
[3]	2024	Deep convolutional neural network (EDNN)	Accuracy – 97% Sensitivity – 98% Precision – 98%
[12]	2024	Stacked Neural Network (SNN)	Accuracy – 96%
[4]	2024	CapsNets CN-1-32, CN-1-64, CN-2-64	Specificity – 98.37% Sensitivity – 97.47% Precision – 97.92%
[6]	2024	XGBoost	Accuracy – 96.92%
[14]	2023	Deep Learning, 2D CNN	Accuracy – 95%
[8]	2023	Efficient Net From CT-scan	Accuracy – 98.64% Precision – 98.63%
[15]	2023	hybrid WOA_APSO Algorithm embedding LDA, CNN	Accuracy – 97.18% Sensitivity – 97% Specificity – 98.66%
[16]	2024	LNDC-HDL CDSO IFB HDE-NN	Accuracy – 97.42% Sensitivity – 96.12% Specificity – 95.25%
[17]	2023	AlexNet Architecture and Support Vector Machine	Accuracy – 97.42% Sensitivity – 96.37% Specificity – 96.37%
[13]	2022	CNN, DCNN, RNN	Accuracy – 97% Sensitivity – 96.2% Specificity – 97.5%
[1]	2022	Convolutional Neural Networks (CNN) RA, CT Scanning	Accuracy – 83%
[11]	2024	Convolutional Neural Networks (CNN) KNN, ANN	Accuracy – 98.8% Sensitivity – 97% Specificity – 98%
[18]	2022	SVM, RBF, Polynomial Kernels	Accuracy – 99.89% Sensitivity – 97% Specificity – 96%
[19]	2014	Nelson Trail	Specificity – 88.35% Sensitivity – 89.85% Accuracy – 90.10%
[20]	2019	3D Deep Learning	Accuracy – 95.92%
[21]	2018	Nelson Randomized	Specificity – 87.35% Sensitivity – 85.85% Accuracy – 91.10%
[22]	2017	CNN	Accuracy – 94.97% Sensitivity – 96.68% Specificity – 95.89%
[23]	2016	TNM Classification	Specificity – 88.32% Sensitivity – 90.85%

			Accuracy – 89.90%
[24]	2020	K-Nearest	Accuracy – 72.24% Sensitivity – 94.12% Specificity – 53.78%
[25]	2020	Artificial neural networks	Accuracy – 90.05%
[26]	2023	Support Vector Machine	Accuracy – 96.6% Sensitivity – 100.0% Specificity – 94.2%
[27]	2020	RF Algorithm	Accuracy – 90.47% Sensitivity – 100.0% Specificity – 93.33%
[28]	2021	Gaussian Naïve Bayes	Confidence interval – 0.31-0.62 Positive predictive value – 0.90
[9]	2019	CNN, RNN-LSTM	Accuracy – 97%
[29]	2019	3D-CNN, RNN	Accuracy – 82.39%
[30]	2019	Delta Radiomics, Support Vector Machine (SVM)	Specificity – 70% Sensitivity – 85% Accuracy – 90.9%
[31]	2019	Median and Gaussian filter, Watershed segmentation algorithm, Random forest(RF) classification model	Specificity – 88.32% Sensitivity – 90.85% Accuracy – 89.90%
[32]	2019	Binarization, CapNets, Convolution neural network (CNN)	Specificity – 95.7% Sensitivity – 91.75% Precision – 91.75% Accuracy – 94.34%
[33]	2016	fuzzy auto-seed cluster	Accuracy – 94%

Table.1. Different methods and their results.

IV. PROPOSED METHODOLOGY

In this work, we provide a compound approach which smoothly merges the machine learning & image processing practice in the identification of lung cancer. First, high-quality lung images are obtained from imaging devices or medical databases. Next, common pre-processing procedures are carried out to better the image quality and lower noise. We then use sophisticated image segmentation methods to precisely separate the lung areas from the surrounding tissues. We investigate and validate both conventional and deep learning-based segmentation methods using either hand annotations or pre-existing ground truth data. The next step in the procedure is extracting the features, which is a very crucial step, where we extract the subtractions from the segmented lung area which relates to malignant cells. To do this we adopt an approach using different factors like texture, shape etc.

The Fusion of clinical data, along with medical history, increases & improves the detection of the problem more effectively. Being a hybrid model of machine learning along with conventional image processing algorithms, it has different integrations and advantages like stacking, fusion of various components for the better outcome. We have taken different factors like sensitivity, specificity, accuracy, and AUC-ROC to measure the performance of this hybrid model. Furthermore, different strategies like interpretability and explainability were added to improve the strategy of decision making as well.

Fig. 5. Flow chart of hybrid model.

We have implemented this hybrid algorithm using different datasets both from internal and external resources to prove that we can apply it on a wide range of datasets.

V. CONCLUSION

The Ambition of performing this examination is to portray a meticulous analysis of the development along with historical data and information in lung cancer identification and classification algorithms. Being more precise, this paper will emphasize different approaches including machine learning, segmentation, clustering, deep learning, and different fuzzy approaches, which plays a key role in grading lung cancer. This study will also cover the information of these methods in detail along with the examination of models which are used in this practice. Along with, we have discussed a hybrid methodology for best feature extraction approaches like CNN, Fuzzy logic, SVM, KNN, RNN & RNN which is a breakthrough in recent times for the pre-detection of lung cancer. This method is working fabulously with accurate precision and accuracy in terms of molecular size and visual data. Important benchmarks like accuracy and precision are included in the valuation of these models in order to find out the lung cancer cells. In addition, this study also opens a way for further hybridizations to keep the goals of strengthening the robustness of the system with different datasets in hand.

REFERENCES

[1] Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, "Machine Learning for Lung Cancer Diagnosis, Treatment, and Prognosis," *Genomics, Proteomics Bioinforma.*, vol. 20, no. 5, pp. 850–866, 2022, doi: 10.1016/j.gpb.2022.11.003.

[2] P. Chaturvedi, A. Jhamb, M. Vanani, and V. Nemade, "Prediction and Classification of Lung Cancer Using Machine Learning Techniques," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1099, no. 1, p. 012059, 2021, doi: 10.1088/1757-899x/1099/1/012059.

[3] S. Poonkodi and M. Kanchana, "Lung cancer segmentation from CT scan images using modified mayfly optimization and particle swarm optimization algorithm," *Multimed. Tools Appl.*, vol. 83, no. 2, pp. 3567–3584, 2024, doi: 10.1007/s11042-023-15688-0.

[4] A. R. Bushara, R. S. Vinod Kumar, and S. S. Kumar, "Classification of Benign and Malignancy in Lung Cancer Using Capsule Networks with Dynamic Routing Algorithm on Computed Tomography Images," *J. Artif. Intell. Technol.*, vol. 4, no. 1, pp. 40–48, 2024, doi: 10.37965/jait.2023.0218.

[5] A. Saha and R. K. Yadav, "Study on segmentation and prediction of lung cancer based on machine learning approaches," *Int. J. Exp. Res. Rev.*, vol. 30, pp. 1–14, 2023, doi: 10.52756/ijerr.2023.v30.001.

[6] Mohammad Shafiquzzaman Bhuiyan *et al.*, "Advancements in Early Detection of Lung Cancer in Public Health: A Comprehensive Study Utilizing Machine Learning Algorithms and Predictive Models," *J. Comput. Sci. Technol. Stud.*, vol. 6, no. 1, pp. 113–121, 2024, doi: 10.32996/jcsts.2024.6.1.12.

[7] Y. Xie *et al.*, "Early lung cancer diagnostic biomarker discovery by machine learning methods," *Transl. Oncol.*, vol. 14, no. 1, 2021, doi: 10.1016/j.tranon.2020.100907.

[8] R. Raza *et al.*, "Lung-EffNet: Lung cancer classification using EfficientNet from CT-scan images," *Eng. Appl. Artif. Intell.*, vol. 126, no. PB, p. 106902, 2023, doi: 10.1016/j.engappai.2023.106902.

[9] D. Mhaske, K. Rajeswari, and R. Tekade, "Deep learning algorithm for classification and prediction of lung cancer using CT scan images," *Proc. - 2019 5th Int. Conf. Comput. Commun. Control Autom. ICCUBEA 2019*, 2019, doi: 10.1109/ICCUBEA47591.2019.9128479.

[10] G. Jakimovski and D. Davcev, "Using double convolution Neural Network for lung cancer stage detection," *Appl. Sci.*, vol. 9, no. 3, 2019, doi: 10.3390/app9030427.

[11] B. M. R. International, "Retracted: Lung Cancer Classification and Prediction Using Machine Learning and Image Processing," *Biomed Res. Int.*, vol. 2024, p. 9851527, 2024, doi: 10.1155/2024/9851527.

[12] S. R. R. B R, S. Sen, R. Bhatt, M. L. Dhanetwal, M. Sharma, and R. Naaz, "Stacked neural nets for increased accuracy on classification on lung cancer," *Meas. Sensors*, vol. 32, no. February, p. 101052, 2024, doi: 10.1016/j.measen.2024.101052.

[13] L. Wang, "Deep Learning Techniques to Diagnose Lung Cancer," *Cancers (Basel).*, vol. 14, no. 22, 2022, doi: 10.3390/cancers14225569.

[14] A. A. Shah, H. A. M. Malik, A. H. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," *Sci. Rep.*, vol. 13, no. 1, pp. 1–15, 2023, doi: 10.1038/s41598-023-29656-z.

[15] S. Vijh, P. Gaurav, and H. M. Pandey, "Hybrid bio-inspired algorithm and convolutional neural network for automatic lung tumor detection," *Neural Comput. Appl.*, vol. 35, no. 33, pp. 23711–23724, 2023, doi: 10.1007/s00521-020-05362-z.

[16] V. K. Gugulothu and S. Balaji, "An early prediction and classification of lung nodule diagnosis on CT images based on hybrid deep learning techniques," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 1041–1061, 2024, doi: 10.1007/s11042-023-15802-2.

[17] I. Naseer, T. Masood, S. Akram, A. Jaffar, M. Rashid, and M. A. Iqbal, "Lung Cancer Detection Using Modified AlexNet Architecture and Support Vector Machine," *Comput. Mater. Contin.*, vol. 74, no. 1, pp. 2039–2054, 2023, doi: 10.32604/cmc.2023.032927.

[18] L. Hussain *et al.*, "Lung Cancer Prediction Using Robust Machine Learning and Image Enhancement Methods on Extracted Gray-Level Co-Occurrence Matrix Features," *Appl. Sci.*, vol. 12, no. 13, 2022, doi: 10.3390/app12136517.

[19] N. Horeweg *et al.*, "Lung cancer probability in patients with CT-detected pulmonary nodules: A prespecified analysis of data from the NELSON trial of low-dose CT screening," *Lancet Oncol.*, vol. 15, no. 12, pp. 1332–1341, 2014, doi: 10.1016/S1470-2045(14)70389-4.

[20] D. Ardila *et al.*, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nat. Med.*, vol. 25, no. 6, pp. 954–961, 2019, doi: 10.1038/s41591-019-0447-x