

Artificial Intelligence for Early Lung Cancer Detection: Revolutionizing Diagnosis and Prognosis

S. Ponlatha¹, D. R. Rithika², M. Vasanthi³, V. Priyadharshini⁴, A. Sangamithra⁵

¹Associate Professor, ECE, Mahendra Engineering College, Namakkal, Tamilnadu, India

^{2,3,4,5}UG students ECE, Mahendra Engineering College, Namakkal, Tamilnadu, India

¹ponlathas@mahendra.info

Abstract: *The proposed methodology employs a variety of image processing and analysis techniques to achieve accurate detection results. To begin, the acquired lung cancer images are pre-processed with a multidimensional filter and histogram equalization in order to improve their quality and subsequent analysis. Histogram equalization optimizes an image's dynamic range, enhancing visibility of structures and abnormalities. This technique proves to be valuable in medical imaging, revealing subtle features for accurate anomaly detection. Meanwhile, Multidimensional Filtering refines image analysis with intelligent filtering methods. Pre-processing, segmentation, and feature extraction from lung cancer images are all part of the method. For accurate lung cancer detection, a deep neural network is trained and tested. The proposed method achieves 99.1251% specificity, 99.1121% sensitivity, and 99.269% accuracy. MATLAB is used to run the entire simulation. The architectural representation distinctly illustrates the method's superior ability to discern true negatives and true positives in lung cancer detection. The research advances lung cancer diagnosis and has the potential for early detection and improved patient care*

Keywords: lung cancer detection, multidimensional filter, histogram equalization, thresholding, Otsu's method, morphology procedure, feature extraction, deep learning

I. INTRODUCTION

The increase in the cases of Lung Cancers these days has created an emergency to detect them at the early stage to battle the disease [1-3]. The early detection and diagnosis of lung cancer over the years, has led to improved patient outcomes and survival rates [4]. Computed Tomography (CT) scans have emerged as a powerful tool for visualizing the internal structures of the lungs with high-resolution images among these modalities [5]. Lung cancer has historically been associated with high mortality rates due to the difficulties in detecting [6]. Because CT scans can capture detailed cross-sectional images of the lungs, radiologists and clinicians can detect suspicious lesions, nodules, and tumors at an earlier stage than ever before [7]. The contribution of this study lies in the development of enhanced deep learning approaches that leverage Complexity Feature extraction and GLCM feature extraction techniques, along with CNNs, for accurate lung cancer detection and classification. The utilization of DOST as an intermediate stage further enhances the performance of these methods [8-11]. The imaging modalities such as CT and MRI have transformed the detection and characterization of lung cancer, providing vital information about tumor size, location, and extent [12-14]. Furthermore, advances in computer-aided diagnosis (CAD) systems have improved the accuracy and efficiency of lung cancer detection by assisting radiologists in interpreting complex CT images [15].

Despite these advances, detecting lung cancer accurately and precisely remains a difficult task. The sheer volume of CT scans produced in clinical settings, combined with subtle differences in tumor appearance and the need for early



detection, necessitates robust and efficient automated approaches [16-18]. Deep learning techniques have shown enormous promise in this area.

For example, in the TNM system, Stage I lung cancer generally refers to tumors that are relatively small and confined to the lung tissue, without evidence of spread to lymph nodes or distant sites. These are considered 'early stage' cancers because they are localized and have not yet advanced to a more aggressive or widespread form.

The need for a morphology-based deep learning approach for precise lung cancer detection stems from the limitations of traditional methods, which are time-consuming and prone to errors. It is possible to automatically learn and extract meaningful features from CT scans allowing for more accurate identification and classification of lung cancer.

II. RELATED WORKS

In general, the literature review section will delve into current research and studies on lung cancer diagnosis, medical imaging modalities, and the use of deep learning techniques. The review will lay down the foundation for the suggested morphology-based deep learning strategy and back up its significance in progressing lung cancer detection from lung CT scans through analyzing and synthesizing this literature. Dritsasetal.proposedRotationForestin2022, a high-Performance algorithm accessed via well-known metrics [19]. They reported an impressive 97.1% accuracy. However, one limitation of the research is that it did not specifically discuss the Rotation Forest algorithm's potential challenges or limitations.

Naseer et al. [20] used the LUNA16 dataset, for various stages of lung nodules, to implement a CNN with optimizers. While they achieved an impressive accuracy of 97.42%, one limitation of traditional CNNs, as mentioned in their study, is the need for a large amount of labeled data for training, which can be difficult to obtain and may affect the generalizability of the results. Venkadesh et al. [21] used an ensemble model in 2021 for the purpose of feature extraction. These features were then combined as input for classification. The limitation of this approach is the lack of discussion about the potential drawbacks or limitations of the ensemble model used.

For early tumor diagnosis, Agarwal et al. [22] proposed a standard Convolutional Neural Network (CNN) with the AlexNet Network Model. Their research used a private dataset and had a 96% accuracy rate. The use of a private dataset, however, is a limitation of this study, as it may introduce bias and limit the generalizability of the findings to other datasets. Masud et al. [23] discussed of using a light CNN architecture in 2020, achieved a high classification accuracy of 97.9%. However, one limitation of this study is that it only looked at the LIDC dataset, leaving the performance on other datasets unexplored. Similarly, Al-Yasriy et al. [24] proposed a CNN technique for cancer detection and categorization using AlexNet. Despite their accuracy of 93.548%, the use of an imbalanced dataset poses a limitation, as imbalanced data can lead to biased model performance and reduced effectiveness in detecting minority classes. In 2019, Toraman et al. [25] suggested Fourier Transform Infrared (FTIR) spectroscopy signals. They attained an accuracy of 95.71%.

Nasser et al. [26] created an Artificial Neural Network (ANN) with an accuracy of 96.67% for lung nodule detection. The lack of a detailed analysis or discussion of potential limitations or challenges encountered during the ANN development and training process is a limitation of this study. Selvanambi et al. [27] demonstrated a Glow-worm swarm optimization in 2018, with an accuracy of 98%. However, the study's limitation is the lack comprehensive discussion of the potential challenges or limitations associated with the GSO algorithm and its application in lung cancer prediction. Zhao et al.[28] proposed a hybrid CNN that makes use of networks like LeNet and AlexNet. They reported an 87.7% accuracy rate. However, one limitation of this study is the comparatively lower accuracy obtained, which may be problematic and have an impact on the proposed hybrid CNN approach's dependability and effectiveness.



Table 1. Research gaps

Study	Proposed Method	Reported Accuracy	Limitations
Dritsas et al.(2022) [19]	Rotation Forest	97.10%	Lack of discussion on algorithm's challenges/limitations
Naseer et al.(2022) [20]	CNN with optimizers (using LUNA16dataset)	97.42%	Dependence on large labeled datasets for training
Venkadesh et al.(2021) [21]	Ensemble model for feature Extraction	94.5	Absence of discussion on limitations of ensemble model
Agarwal et al.(2021) [22]	CNN with AlexNet (using Private dataset)	96%	Potential bias and limited generalizability due to private dataset Use
Masudetal. (2020) [23]	Light CNN architecture (using LIDC dataset)	97.90%	Performance on other datasets unexplored
Al-Yasriy et al.(2020) [24]	CNN with AlexNet (using imbalanced dataset)	93.55%	Biased model performance and reduced effectiveness for Minority classes due to imbalanced data
Nasser et al.(2019) [20]	Artificial Neural Network (ANN)	96.67%	Lack of detailed analysis or discussion of limitations/challenges in ANN development/training
Selvanambi et al.(2018) [27]	Glow-worm swarm optimization	98%	Absence of comprehensive discussion of potential GSO algorithm challenges/limitations
Zhaoetal. (2018) [28]	Hybrid CNN (LeNet and AlexNet)	87.70%	Lower accuracy may impact dependability and effectiveness of hybrid CNN approach

Table 1 provides a clear overview of the strengths and limitations of each study. In summary, early detection of lung cancer can lead to improved outcomes, reduced mortality rates, and a better quality of life for individuals diagnosed with the disease. It also has broader societal and economic benefits, making it a crucial focus area in the fight against lung cancer.

CAD systems are specialized software tools designed to assist healthcare professionals in the interpretation of medical images, such as X-rays, CT scans, and MRIs. These systems use advanced algorithms and machine learning techniques to analyze images and highlight areas of interest that may require further examination. They aim to improve accuracy and efficiency in the diagnostic process. The steps include in CAD are image preprocessing, Feature Extraction, Classification, alert generation. While processing the limitations includes false negatives, dependence on quality if input images, lack of clinical context, limited to image analysis etc.

The research findings might have positive implications for patients in the following ways Reduced Invasive Procedures, Decreased Psychological Burden, Improved Quality of Life, Enhanced Monitoring and Surveillance, Personalized Treatment Approaches, Earlier Detection and Treatment etc.

In a nutshell while these studies have made substantial contributions to lung cancer detection and classification, it is critical to recognize the limitations of each approach. Addressing these limitations can help future research in this area improve its accuracy, generalizability, and effectiveness.

III. METHODOLOGY

Figure 1 represents the proposed process flow block diagram of the new Improved Enhanced algorithm for detecting lung cancer. The algorithm incorporates several key components. To summarize, the new Improved Enhanced algorithm combines several pre-processing techniques, including Histogram Equalization, OTSU Segmentation, Sobel Filtering, and GLCM-based feature extraction. These steps are intended to improve the image, isolate lung regions, extract meaningful features, and then use a CNN with Ranking features to detect lung cancer accurately.



The Otsu method is a thresholding technique used to segment images. It calculates an "optimal" threshold value to separate foreground and background pixels. The key parameter in the Otsu method is the threshold value, which is determined by maximizing the between-class variance. This threshold value is used to classify pixels into foreground and background based on their intensity values. The CNN architecture typically includes Input Layer, Convolutional Layers, Activation Functions, Pooling Layers, Fully Connected Layers, Output Layer, Loss Function and Optimizer, Regularization and Dropout, Number of Layers and Units.

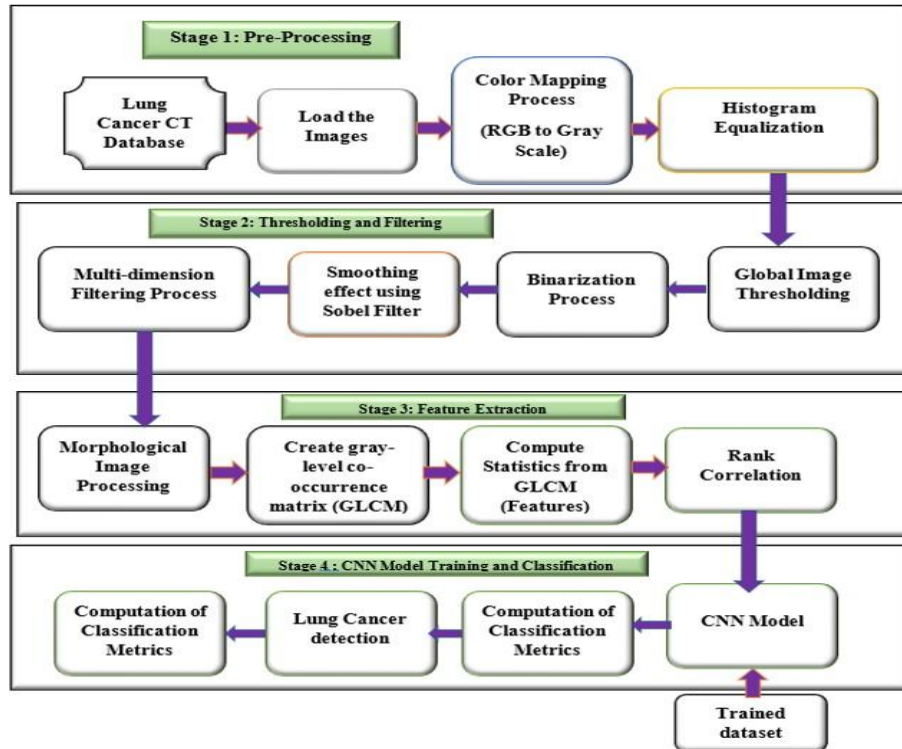


Figure 1. Proposed Lung cancer detection method

Algorithm

Stage1: Pre-Processing:

Step i. Import the Lung Cancer CT scan dataset from LIDC.

Step ii. Perform color mapping process to convert the RGB image to grayscale.

a. Colour mapping is the process of converting an RGB image to grayscale by calculating the luminance or intensity value for each pixel. The luminosity method is one of the most commonly used formulas for performing this conversion.

The luminosity method calculates the grayscale value (G) based on the RGB values of a pixel (R, G, B) using the following formula:

$$G=0.21R+0.72 G+0.07B \quad (1)$$

In this formula, the coefficients 0.21, 0.72, and 0.07 represent the perceived luminance contributions of the red, green, and blue channels, respectively.

Step iii. Apply histogram equalization to enhance the image contrast.

a. Compute the histogram of the gray scale image input. The frequency of occurrence of each intensity value in the image is represented by the histogram.

b. Determine the histogram's cumulative distribution function(CDF). The cumulative probability of occurrence for each intensity value is represented by the CDF.



c. Normalise the CDF to a scale of $[0, 255]$ (for 8-bit grayscale images). This step ensures that histogram equalization works across the entire intensity range.

$$cdf_{normalized} = \frac{cdf - cdf_{min}}{L - 1} * (M * N - cdf_{min}) \quad (2)$$

where, $cdf_{normalized}$: Normalized CDF values; cdf : Cumulative distribution function; cdf_{min} : Minimum value of the CDF; L : Number of intensity levels (typically 256 for 8-bit images); M : No. of Rows; N : No. of Columns.

d. Using the normalised CDF, apply the histogram equalisation transformation to each pixel in the grayscale image. Replace each pixel's intensity value with its corresponding value in the normalised CDF.

$$output_{pixel} = cdf_{normalized}[input_{pixel}] \quad (3)$$

e. For display purposes, round the output pixel values to the nearest integer.

f. When compared to the original grayscale image, the resulting image will have more contrast.

Stage 2: Thresholding and Filtering:

Step i. Perform global image thresholding using the Otsu method to segment the image into foreground and background.

a. Create a histogram of the grayscale image input.

Assume the grayscale image has intensity values ranging from 0 to $L-1$, where L is the number of intensity levels possible. The histogram will be a 1D array of L elements, with the element at index i representing the number of pixels in the image with intensity i .

b. Normalise the histogram in step two.

Divide each histogram element by the total number of pixels in the image. This step ensures that the histogram is transformed into a probability distribution with a sum of 1.

c. Determine the normalised histogram's cumulative distribution function (CDF).

The CDF is calculated by adding the normalized histogram values from 0 to i , where i is a number ranging from 0 to $L-1$. The CDF will be a 1D array with L elements as well.

d. Determine the cumulative and total means.

The cumulative mean at intensity i is calculated by multiplying the intensity value i by its corresponding normalized histogram value and adding the results from 0 to i . The sum of all cumulative means is the total mean.

e. For each possible threshold, compute the between-class variance.

f. Segment the image using the chosen threshold. Set all pixel intensities below the threshold to 0 and all pixel intensities equal to or greater than the threshold to 255.

Step ii. Apply binarization process to convert the thresholded image into a binary image.

Step iii. Apply smoothing effect using the Sobel filter to reduce noise and highlight edges.

Step iv. Perform multi-dimensional filtering process to enhance specific features.

The above discussed part can be provided as a summarised Algorithm

Stage3: Feature Extraction:

Step i. Apply morphological image processing by introducing a structuring element to extract relevant features.

Step ii. Perform dilation operation to expand the regions of interest and its Pseudocode is

Step iii. Create a gray-level co-occurrence matrix (GLCM) to capture spatial relationships between pixels.

a. Load the input image and import the necessary libraries.

b. If the input image is not already grayscale, convert it to grayscale.

c. Define the GLCM calculation's distance and angle offsets. For co-occurrence measurements, these offsets determine the neighboring pixels.



- d. Set the number of grey levels to be used in quantizing the grayscale image. The size of the GLCM matrix is determined by this.
 - e. Create an empty GLCM matrix of size
 - f. Iterate over each pixel in the grayscale image:
 - i. Determine the co-occurring pixel based on the distance and angle offsets specified.
 - ii. Based on the grayscale values of the current and co- occurring pixels, increment the corresponding element in the GLCM matrix.
 - g. Divide each element in the GLCM matrix by the sum of all elements in the matrix to normalize it. The GLCM is scale- invariant after this step.
 - h. Normalize the GLCM matrix by dividing each element by the sum of all elements in the matrix. This step ensures that the GLCM is scale-invariant.
- Step iv. Compute statistics from the GLCM, such as energy, contrast, and Entropy, as features.

$$Contrast = \sum (i-j)^2 P(i,j) \quad (5)$$

Where, (i,j) represents the element in the GLCM matrix and P(i,j) is the normalized probability.

$$Energy = \sum (P(i,j))^2 \quad (6)$$

$$Entropy = - \sum (P(i,j) \log_2(P(i,j))) \quad (7)$$

Step v. Carry out rank correlation process to select the most informative features.

- a. Assign ranks to data points in each dataset based on feature values for each computed texture feature. Take the following steps:
 - i. Sort the data points in each dataset according to their feature values.
 - ii. Give each data point a rank based on its position in the sorted list. If there are ties, give the tied data points the average rank.
- b. For the current texture feature, compute the difference in ranks for each data point in both datasets.
- c. Squaring the difference store move the sign.
- d. Add together all of the squared differences for the current texture feature to get the sum of squared differences (SSD).
- e. n is the total number of data points.
- f. Calculate the rank correlation coefficient using the formula:

$$(6SSD)$$

$$RankCorrelation = 1 - \frac{(6SSD)}{(n(n-1))} \quad (8)$$

This formula is for the Spearman's rank correlation coefficient.

Pass the resulting features to Stage 4.

Stage 4: CNN Model Training and Classification:

- Step i. Train a CNN model using the features that have been extracted from the previous stage.
- Step ii. Utilize a pre-defined CNN model architecture for training.
- Step iii. Compute classification metrics
- Step iv. Perform lung cancer detection by classifying whether the image is cancerous or not.
- Step v. Evaluate the performance of the overall system

The proposed lung cancer detection method using the Improved Enhanced algorithm has potential real-world applicability in clinical settings. However, there are several considerations and potential hurdles that need to be addressed for successful integration into a clinical workflow. The stages are Data Acquisition and Integration, Pre-Processing and Computational Resources, Integration with Existing Systems, Clinical Validation and Regulatory



Approval, Interpretability and Explainability, Continuous Monitoring and Improvement, Legal and Ethical Considerations.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

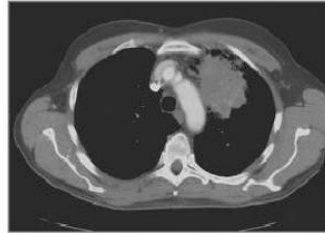


Figure 2. Original lung scan image

The image in Figure 2 is an original lung scan image obtained from the LIDC dataset via the Kaggle platform. The LIDC dataset is a well-known dataset used in medical imaging research, specifically for lung cancer analysis and detection. For this study, a subset of 32 samples from the LIDC dataset was selected for training. The original lung scan image is used to begin further processing and analysis. The resized version of the original lung scan image is depicted in Figure 3. Resizing is an important step in medical image classification tasks for a variety of reasons. For starters, it increases computational efficiency by lowering the computational load required for subsequent analysis. The computational resources and processing time required for feature extraction and classification are reduced when images are resized. Second, resizing is required to address memory constraints, particularly when working with large datasets. Memory usage can be reduced by reducing image dimensions, allowing for smoother execution of classification algorithms. Furthermore, resizing ensures consistent image sizes, allowing for compatibility across images during the classification process. It also makes other preprocessing steps easier, such as feature extraction, normalization, and data augmentation, possible.

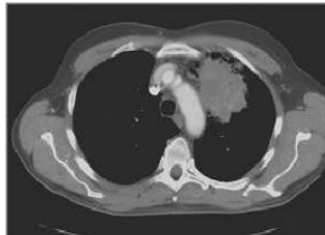


Figure 3. Image resize resultant

The image in Figure 4 shows the outcome of histogram equalisation on the resized lung scan image. Histogram equalisation is a technique used in image processing to improve contrast. Histogram equalisation aims to maximise the use of an image's dynamic range by redistributing its intensity values. This process produces a more contrasted image, making the underlying structures and abnormalities more visible. Histogram equalisation is especially useful in medical image analysis because it improves visibility of subtle features and aids in the accurate identification of abnormalities.



Figure 4. Histogram equalized image



The global thresholded image obtained using the Otsu method is shown in Figure 5. The Otsu method is a popular image segmentation thresholding technique. Its goal is to find the best threshold value for separating the image into foreground and background regions. The Otsu method determines a threshold that maximizes the separation of these two regions by calculating the between-class variance of the intensity values. The thresholded image in Figure 5 emphasises the distinct regions within the lung scan, allowing subsequent analysis and feature extraction to focus on specific areas of interest.

The image in Figure 6 is the result of applying a multi-dimensional filter to the thresholded lung scan image, specifically the Sobel filter. The Sobel filter is a popular edge detection filter that emphasises sharp intensity transitions in images. The resulting image in Figure 6 emphasises the edges and boundaries of structures within the lung scan by convolving the Sobel filter with the thresholded image. This edge data is useful for further analysis and feature extraction, assisting in the identification and characterization of important anatomical structures or abnormalities.



Figure 5. Global thresholded image



Figure 6. Sobel filter based resultant

Figure 7 depicts the morphologically dilated image produced by the multi-dimensional filter used in Figure 6. Morphological dilation is an image processing operation that expands the boundaries of regions or objects. Morphological dilation enlarges the regions of interest and connects nearby structures by convolving a structuring element with the image. Morphological dilation can be useful in the context of lung scan analysis for enhancing and consolidating the boundaries of lung structures or abnormalities. The resulting dilated image in Figure 7 is an intermediate representation that provides a clearer and more comprehensive visualisation of the relevant anatomical structures or pathologies.



Figure 7. Morphologically dilated image Similarly, 32 samples are recurrently being trained in the Lung Cancer detection process, and the features extracted using the GLCM process are tabulated in Table 2 for the corresponding 5 samples that have been processed to the proposed system. These features offer quantitative representations of specific lung image characteristics relevant for cancer detection.



Table 2. Features extracted for samples of lung cancer images

Features Extracted	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Entropy	0.7642	0.7712	0.7753	0.7572	0.7742
Contrast	2.942	2.941	2.9291	2.9748	2.9175
Energy	0.5892	0.5891	0.5873	0.5916	0.5913

Performance evaluation

The accuracy of the Lung Cancer detection can be calculated using the following formula:

$$(TP+TN)$$

$$Accuracy=(TP+TN+FP +FN) \quad (9)$$

where, TP: True Positive (identified Tumors); TN: True Negative, FP: False Positive; FN: False Negative (not identified).

Specificity is the proportion of true negatives identified correctly by the model. It indicates the model's ability to correctly classify non-tumor or non-cancer cases.

$$TN$$

$$Specificity=(TN +FP) \quad (10)$$

Sensitivity is the proportion of true positives that were identified by the model. It indicates the model's ability to correctly classify tumour or cancer cases.

$$TP$$

$$Sensitivity= \frac{TP}{TP+FN} \quad (11)$$

Table 3 provides an overview of the proposed method's accuracy performance evaluation in comparison to previous methods used for lung cancer detection. The proposed method achieved an impressive 99.269% accuracy. This accuracy rate indicates the proposed approach's ability to correctly classify lung cancer cases, making it highly effective in detecting the disease's presence. The table also compares the proposed method to others, such as Machine Learning [19], CNN Alexnet + SGD [20], Alexnet CNN [22], CNN [23], CNN with Alexnet [24], and ML with FTIR Signals [25]. This indicates that the proposed method significantly improved lung cancer detection accuracy when compared to existing techniques, which included both traditional machine learning and deep learning methods.

Table 3. Accuracy performance evaluation

Year	Techniques Used	Accuracy (%)
2022	Machine Learning [19]	97.1
2022	CNN Alexnet + SGD [20]	97.42
2021	Alexnet CNN [22]	96
2020	CNN [23]	97.9
2020	CNN with Alexnet [24]	93.458
2019	ML with FTIR Signals [25]	95.71
	Propose Proposed Method	99.269

Figure 8 shows a comparison plot for lung cancer detection accuracy. The plot depicts the accuracy performance of various methods, including the proposed morphology-based deep learning approach. Figure 8 illustrates the significant improvement in accuracy provided by the proposed approach when compared to existing techniques. These findings emphasise the proposed method's potential to improve lung cancer diagnosis and contribute to more accurate and efficient clinical decision-making processes.

Table 4 compares the proposed methodology to several previous methods in terms of specificity and sensitivity in lung cancer detection. The proposed method had a 99.1251% specificity and a 99.1121% sensitivity. These values represent



the proposed method's ability to accurately identify true negatives (specificity) and true positives (sensitivity) in lung cancer detection.

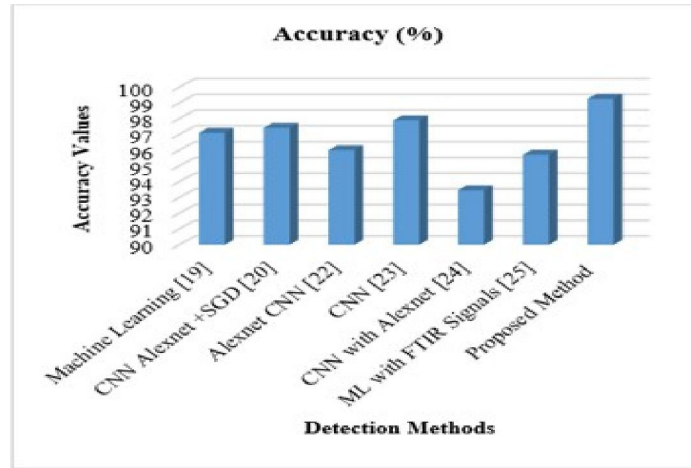


Figure 8. Comparative plot for accuracy in lung cancer detection

Table 4. Performance evaluation for specificity and sensitivity in lung cancer detection

Year	Techniques used	Specificity (%)	Sensitivity (%)
2022	CNN Alexnet + SGD [20]	97.25	97.58
2021	Deep Learning [21]	90	86
2020	CNN with Alexnet [24]	95	95.71
2019	ML with FTIR Signals [25]	97.50	93.33
	Proposed Method	99.1251	99.1121

The following table compares the proposed method's performance to CNN Alexnet + SGD [20], Deep Learning [21], CNN with Alexnet [24], and ML with FTIR Signals [25]. When compared to all previous methods, the proposed methodology has higher specificity and sensitivity values. This suggests that, when compared to existing techniques, the proposed morphology-based deep learning approach has significantly improved the specificity and sensitivity of lung cancer detection.

A comparison plot for specificity and sensitivity in lung cancer detection is shown in Figure 9. The plot depicts the performance differences between the proposed morphology-based deep learning approach and the previous methods in terms of specificity and sensitivity.



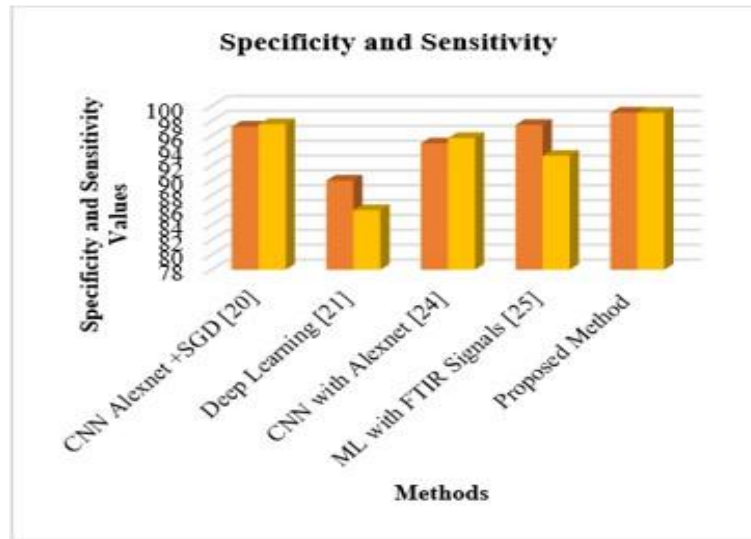


Figure 9. Comparative plot for specificity and sensitivity in lung cancer detection

The results presented in the study demonstrate a significant advancement in the field of lung cancer detection. The proposed method achieved an accuracy of 99.269%, specificity of 99.1251%, and sensitivity of 99.1121%. These results indicate an exceptionally high level of accuracy in classifying both cancerous and non-cancerous cases.

In the context of lung cancer detection, these results have several important implications:

Improved Clinical Decision-Making: The high accuracy, specificity, and sensitivity of the proposed method suggest that it could serve as a reliable tool for assisting healthcare professionals in the early detection of lung cancer. This can lead to more accurate diagnoses and treatment plans.

Early Detection and Intervention: High sensitivity means that the proposed method is effective in correctly identifying true positives (cases of lung cancer). Early detection of cancer is crucial for timely intervention and improved patient outcomes. With a sensitivity of 99.1121%, the proposed method excels in this aspect.

Reduced False Positives: The high specificity value (99.1251%) implies a low rate of false positives. This is particularly important in clinical practice, as it minimizes the chances of unnecessary follow-up tests or procedures for patients who do not have lung cancer.

Potential for Screening Programs: The high accuracy of the proposed method makes it a promising candidate for use in large-scale lung cancer screening programs. Such programs can be instrumental in identifying cases at an early, more treatable stage.

Resource Optimization: The reduction in false positives and negatives, as indicated by the high specificity and sensitivity, respectively, can lead to more efficient allocation of healthcare resources. It can help in prioritizing cases that require immediate attention.

Enhanced Patient Outcomes: Accurate and timely diagnosis of lung cancer can significantly improve patient outcomes. It can lead to earlier treatment initiation, potentially increasing survival rates and overall quality of life for patients.

Research and Development: The high performance of the proposed method may encourage further research and development in the field of medical image analysis for lung cancer detection. This could lead to continuous advancements in detection techniques and tools.

V. CONCLUSION

We presented a methodology for detecting lung cancer using lung cancer images obtained from the LIDC Database in this paper. To achieve accurate detection results, the proposed method combines various image processing and analysis techniques. Relevant features are extracted from lung cancer images using pre-processing, segmentation, and feature



extraction. These characteristics are then fed into a deep neural network architecture. The experimental results show that the proposed methodology is effective, with a high specificity of 99.1251%, sensitivity of 99.1121%, and overall accuracy of 99.269%.

The study's findings represent a significant advancement in both the accuracy of lung cancer detection and the potential application of deep learning techniques in medical imaging. This not only has direct implications for patient care but also contributes to the broader landscape of medical research and practice. The proposed method's high accuracy, specificity, and sensitivity have several direct implications for clinical practice by taking the following parameters Enhanced Diagnoses, Reduction in False Positives, Early Intervention and Treatment, Resource Allocation, Potential for Screening Programs, Improved Patient Experience. The findings of this study have broader implications for the field of lung cancer detection and medical image analysis by the following techniques like Advancement in Detection Techniques, Integration of Deep Learning in Medical Imaging, Potential for Transferable Techniques, Contribution to Research and Clinical Practice, Impetus for Further Innovation.

The use of MATLAB as a computing tool ensured efficient implementation and consistent results. Finally, this study demonstrated the utility of image processing and deep learning techniques in the detection of lung cancer. The proposed methodology achieves high accuracy, specificity, and sensitivity, indicating that it is effective in identifying lung cancer cases. The approach's overall performance and applicability in clinical practice will be improved through further refinement of the methodology, validation through extensive clinical trials, and integration with complementary data sources in the future.

REFERENCES

- [1] Jones, P.A., Baylin, S.B. (2007). The epigenomics of cancer. *Cell*, 128(4):683-692. <https://doi.org/10.1016/j.cell.2007.01.029>
- [2] What Is Cancer? National Cancer Institute. <https://www.cancer.gov/about-cancer/understanding/what-is-cancer>.
- [3] Zheng, R.S., Sun, K.X., Zhang, S.W., Zeng, H.M., Zou, X.N., Chen, R., He, J. (2019). Report of cancer epidemiology in China, 2015. *Chinese Journal of Oncology*, 41(1): 19-28. <https://doi.org/10.3760/cma.j.issn.0253-3766.2019.01.005>
- [4] Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 71(3): 209-249. <https://doi.org/10.3322/caac.21660>
- [5] National Cancer Institute. Surveillance, Epidemiology, and End Results Program, Cancer Stat Facts: Lung and Bronchus Cancer [cited 2021 Dec 19]. <https://seer.cancer.gov/statfacts/html/lungb.html>.
- [6] Ferlay, J., Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., Bray, F. (2021). Cancer statistics for the year 2020: An overview. *International Journal of Cancer*, 149(4):778-789. <https://doi.org/10.1002/ijc.33588>
- [7] Worley, S. (2014). Lung cancer research is taking on new challenges: knowledge of tumors' molecular diversity is opening new pathways to treatment. *Pharmacy and Therapeutics*, 39(10): 698-714.
- [8] Rami-Porta, R., Call, S., Doms, C., Obiols, C., Sánchez, M., Travis, W.D., Vollmer, I. (2018). Lung cancer staging: a concise update. *European Respiratory Journal*, 51(5):1800190. <https://doi.org/10.1183/13993003.00190-2018>
- [9] Hegde, P.S., Chen, D.S. (2020). Top 10 challenges in cancer immunotherapy. *Immunity*, 52(1): 17-35.
- [10] Kourou, K., Exarchos, T.P., Exarchos, K.P., Karamouzis, M.V., Fotiadis, D.I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13: 8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- [11] Iqbal, M.J., Javed, Z., Sadia, H., Qureshi, I.A., Irshad, A., Ahmed, R., Malik, K., Raza, S., Abbas, A., Pezzani, R., Sharifi-Rad, J. (2021). Clinical applications of artificial intelligence and machine learning in cancer diagnosis: looking into the future. *Cancer Cell International*, 21(1): 1-11. <https://doi.org/10.1186/s12935-021-01981-1>
- [12] Gang, P., Zhen, W., Zeng, W., Gordienko, Y., Kochura, Y., Alienin, O., Stirenko, S. (2018). Dimensionality



- reduction in deep learning for chest X-ray analysis of lung cancer. In 2018 Tenth International Conference on Advanced Computational Intelligence(ICACI), Xiamen, China, pp.878-883. <https://doi.org/10.1109/ICACI.2018.8377579>
- [13] Li, X., Shen, L., Xie, X., Huang, S., Xie, Z., Hong, X., Yu, J. (2020). Multi-resolution convolutional networks for chest X-ray radiograph-based lung nodule detection. *ArtificialIntelligenceinMedicine*,103:101744. <https://doi.org/10.1016/j.artmed.2019.101744>
- [14] Weinstein, J.N.,Collisson,E.A.,Mills,G.B.,Shaw,K.R., Ozenberger,B.A.,Ellrott,K.,Shmulevich,I.,Sander,C., Stuart, J.M. (2013).The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10): 1113-1120. <https://doi.org/10.1038/ng.2764>
- [15] Clark, K., Vendt, B., Smith,K., Freymann, J., Kirby, J., Koppel, P., Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository. *Journal of Digital Imaging*, 26: 1045-1057. <https://doi.org/10.1007/s10278-013-9622-7>
- [16] Pavlopoulou, A., Spandidos, D.A., Michalopoulos, I. (2014). Human cancer databases(Review).2015 *OncologyReports*,33:3-18.
- [17] Luo, Y., Wang, F., Szolovits, P. (2017). Tensor factorization toward precision medicine. *Briefings in Bioinformatics*,18(3):511-514. <https://doi.org/10.1093/bib/bbw026>
- [18] Ausawalaitong, W., Thirach, A., Marukatat, S., Wilaiprasitporn, T. (2018). Automatic lung cancer prediction from chest X-ray images using the deep learning approach. In 2018 11th Biomedical Engineering International Conference (BMEiCON), Chiang Mai, Thailand, pp.1-5.. <https://doi.org/10.1109/BMEiCON.2018.8609997>
- [19] Dritsas,E.,Trigka,M.(2022).Lung cancer risk prediction with machine learning models. *Big Data and Cognitive Computing*, 6(4): 139. <https://doi.org/10.3390/bdcc6040139>
- [20] Naseer, I., Akram, S., Masood, T., Jaffar, A., Khan, M. A.,Mosavi,A. (2022). Performance analysis of state-of-the-art CNN architectures for luna16. *Sensors*, 22(12): 4426. <https://doi.org/10.3390/s22124426>
- [21] Venkadesh, K.V., Setio, A.A., Schreuder, A., Scholten, E.T., Chung, K., W. Wille, M.M., Jacobs, C. (2021). Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT. *Radiology*, 300(2): 438-447. <https://doi.org/10.1148/radiol.2021204433>
- [22] Agarwal, A., Patni, K., Rajeswari, D. (2021). Lung cancer detection and classification based on alexnet CNN. In 2021 6th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 1390-1397. <https://doi.org/10.1109/ICCES51350.2021.9489033>
- [23] Masud, M., Muhammad, G., Hossain, M.S., Alhumyani, H., Alshamrani, S.S., Cheikhrouhou, O., Ibrahim, S. (2020). Light deep model for pulmonary nodule detection from CT scan images for mobile devices. *WirelessCommunicationsandMobileComputing*,2020: 8893494. <https://doi.org/10.1155/2020/8893494>
- [24] Al-Yasriy, H.F.,Al-Husieny,M.S.,Mohsen,F.Y.,Khalil, E.A., Hassan, Z.S. (2020). Diagnosis of lung cancer based on CT scans using CNN. In *IOP Conference Series: Materials Science and Engineering*, 928(2): 022035. <https://doi.org/10.1088/1757-899X/928/2/022035>
- [25] Toraman, S., Girgin, M., Üstündağ, B., Türkoğlu, İ. (2019). Classification of the likelihood of colon cancer with machine learning techniques using FTIR signals obtained from plasma. *Turkish Journal of Electrical Engineering and Computer Sciences*, 27(3): 1765-1779. <https://doi.org/10.3906/elk-1801-259>
- [26] Nasser, I.M., Abu-Naser, S.S. (2019). Lung cancer detection using artificial neural network. *International Journal of Engineering and Information Systems (IJEAIS)*, 3(3): 17-23.
- [27] Selvanambi,R., Natarajan,J.,Karuppiyah,M.,Islam,S.H., Hassan, M.M., Fortino, G. (2020). Lung cancer prediction using higher-order recurrent neural network based on glowworm swarm optimization. *Neural ComputingandApplications*,32:4373-4386. <https://doi.org/10.1007/s00521-018-3824-3>
- [28] Zhao,X.,Liu,L.,Qi,S.,Teng,Y.,Li,J.,Qian,W.(2018). Agile convolutional neural network for pulmonary nodule classification using CT images. *International Journal of Computer Assisted Radiology and Surgery*,13: 585-595. <https://doi.org/10.1007/s11548-017-1696-0>

