OPEN ACCESS JOURNALS
Indiana Journal of Multidisciplinary Research
Abbreviated Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.21608381

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Article

Volume-06|Issue-04|2026

An Enhanced Machine Learning Framework for Lung Cancer Detection Using CT Scan Imaging

Kavitha N[1], Shankara C[2], Hariprasad S A[3], Sujatha H[4], Ashwini K S[5]

[1]Dept. of Electronics. & Communication. Engineering, RV Institute of Technology and Management., Karnataka., India.
[2,4,5]Dept. of Electronics. & Communication. Engineering Govt. Polytechnic. Malavalli., Karnataka., India.
[3]Dept. of Electronics. & Communication. Engineering FET,JAIN. University, Bengaluru.,Karnataka, India.

Article History
Received: 05.05.2026
Accepted: 22.06.2026
Published: 25.07.2026

Citation
Kavitha, N., Shankara, C., Hariprasad, S. A., Sujatha, H. & Ashwini, K. S. (2026) An Enhanced Machine Learning Framework for Lung Cancer Detection Using CT Scan Imaging. *Indiana Journal of Multidisciplinary Research, 6*(4), 289-297.

Abstract: Lung cancer remains one of the leading causes of cancer-related mortality worldwide, highlighting the critical importance of early and accurate diagnosis for improving patient survival rates. Conventional diagnostic approaches primarily rely on medical imaging techniques such as X-rays, PET, and CT scans. This study presents an advanced computer-aided detection framework for lung cancer diagnosis using CT scan images. The proposed pipeline incorporates a median filter for noise reduction, a thresholding algorithm for segmentation, Scale-Invariant Feature Transform (SIFT) for feature extraction, and a Support Vector Machine (SVM) classifier for final classification. This systematic integration enables effective noise suppression, accurate edge-preserving segmentation, and robust feature representation, ultimately enhancing classification accuracy. A comparative evaluation against an alternative framework and a CNN-based model demonstrates the superiority of the proposed approach, achieving an accuracy of 96.32%, with precision and recall of 98%, and an F1-score of 97%. The findings confirm the diagnostic efficiency, scalability, and reliability of the framework, underscoring its potential to advance automated lung cancer detection and support improved clinical decision-making.

Keywords: Classification, Computed tomography, Feature extraction, Lung cancer, Machine learning, Preprocessing, Segmentation.

INTRODUCTION

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, with its incidence and death rates continuing to rise annually. Despite medical advancements, the overall five-year survival rate remains below 20%, primarily due to late-stage diagnoses when treatment options are limited [1,2]. Globally, approximately 2.09 million new cases and 1.76 million deaths were reported in 2018, representing 19.4% of all cancer-related fatalities [3]. By 2030, lung cancer is projected to account for 2.45 million deaths, and forecasts for 2040 suggest over 3.3 million new cases annually [4]. In India alone, 67,795 cases and 45,363 deaths were recorded in 2018, with survival rates across all stages remaining at only 14% [5]. These statistics underscore the urgent need for early and accurate diagnostic techniques to improve survival outcomes and reduce the global disease burden.

Current diagnostic methods, including cytology, chest radiography, and computed tomography (CT), are widely employed but face significant limitations. These include high costs, extended processing times, and low sensitivity in detecting early-stage lung cancer [6]. Conventional image-based approaches also involve multiple processing stages, which increase computational complexity and the likelihood of diagnostic errors [7]. Manual interpretation by radiologists, though regarded as the gold standard, is labor-intensive and subject to inter-observer variability, further reinforcing the need for automated, efficient, and reliable diagnostic systems.

In recent years, machine learning (ML) and deep learning (DL) methods have shown promising potential in medical image analysis. Automated frameworks have been explored to enhance classification accuracy and reduce human error; however, many existing approaches still suffer from high classification errors and limited generalizability [8,9]. Thus, more robust and scalable methodologies are essential to achieve clinically reliable diagnostic performance.

The present research aims to develop an advanced computer-assisted diagnostic framework for lung cancer detection using CT scan images. The proposed approach integrates traditional image processing with machine learning and deep learning methods to classify CT scans as benign or malignant based on extracted features. By employing noise reduction, segmentation, feature extraction, and classification within a unified pipeline, the framework seeks to improve diagnostic accuracy, reduce error rates, and enable early detection. The successful implementation of this system has the potential to significantly enhance patient survival, reduce treatment complexity, and lower healthcare costs.

***Corresponding Author:** Kavitha N

Related Work

Preprocessing of lung CT images is a critical step to eliminate noise and inconsistencies, thereby enhancing the accuracy of subsequent detection tasks. Image segmentation is widely employed to partition CT images into meaningful regions, enabling effective identification of nodules. Feature extraction further supports this process by capturing essential image characteristics, which facilitates the differentiation between malignant and benign tissues. In the final stage, classification methods—including traditional machine learning and modern deep learning techniques—are utilized to perform diagnostic decisions. A wide range of studies have examined these stages, exploring diverse strategies to improve detection performance.

In the area of preprocessing and enhancement, several filters and contrast enhancement methods have been applied. For instance, Gabor and Gaussian filters have been used to improve image quality, with subsequent thresholding and marker-controlled watershed segmentation achieving 92% accuracy [12]. Contrast-Limited Adaptive Histogram Equalization (CLAHE) has been combined with fuzzy C-means segmentation and Gray-Level Co-occurrence Matrix (GLCM) texture features for improved visualization and diagnosis using Bayesian classifiers [13]. Similarly, histogram equalization, binarization, and shape-based feature extraction have been integrated with neural networks for lung CT scan classification [15].

For segmentation, a variety of techniques have been employed. Region-growing and watershed algorithms have been implemented to identify suspicious regions [14], while adaptive diffusion active contour models have been adopted for refined segmentation of affected areas, with subsequent shape-based classification using SVM [17]. Otsu thresholding has also been utilized in combination with Wiener and median filters to segment and classify nodules through neural networks and SVM [18]. Watershed segmentation, in conjunction with adaptive thresholding, has proven effective in nodule detection [20]. Super-pixel algorithms have been explored as well, achieving an accuracy of 84.75% .

In terms of feature extraction, researchers have relied heavily on texture- and shape-based descriptors. Haar wavelet transforms have been applied to extract discriminative features, followed by neural network classification [16]. GLCM has been frequently employed for texture analysis, often coupled with segmentation methods such as Otsu's thresholding or marker-controlled watershed [22,24]. SIFT features, combined with Haar wavelet compression and neural networks, have achieved accuracies as high as 95.3% .

Regarding classification approaches, Support Vector Machines (SVM) remain the most widely used, often outperforming neural networks when coupled with robust preprocessing and feature extraction strategies [17,19,21,23]. More advanced frameworks have incorporated Random Forests and Artificial Neural Networks (ANN), where preprocessing with histogram equalization and edge detection was followed by watershed segmentation, achieving a classification accuracy of 92%.

Overall, prior studies demonstrate the effectiveness of combining preprocessing, segmentation, and feature extraction with machine learning for lung cancer detection. However, many methods suffer from limited accuracy, high computational cost, or reduced robustness when applied to diverse datasets. These limitations highlight the need for an integrated, scalable framework that ensures reliable performance across varied imaging conditions. The present study addresses this gap by introducing a systematic pipeline that combines noise reduction, robust segmentation, advanced feature extraction, and optimized classification, thereby improving both accuracy and diagnostic reliability.

PROPOSED METHODOLOGY

The proposed methodology, illustrated in Fig. 1, outlines an integrated framework for lung cancer detection from CT scan images. The system combines traditional image processing techniques with machine learning (Framework 1) and includes a convolutional neural network (CNN) model for comparative analysis. The framework is designed to ensure robust preprocessing, accurate segmentation, effective feature extraction, and reliable classification of lung nodules.

Data Collection and Preprocessing

The process begins with the acquisition of CT scan images from publicly available and clinically validated datasets. To prepare the images for analysis, a sequence of preprocessing steps is applied. These include grayscale conversion to standardize image representation, intentional noise addition to evaluate robustness under varying imaging conditions, and noise reduction using a median filter to eliminate artifacts while preserving edge details. Image enhancement techniques are employed to improve the visibility of lung structures, followed by morphological operations to refine structural boundaries. Finally, Region of Interest (ROI) selection is performed to isolate lung regions and exclude irrelevant background information, ensuring computational efficiency and diagnostic focus.

Segmentation

In Framework 1, segmentation is performed using a thresholding algorithm to separate lung nodules from surrounding tissues. Thresholding enables efficient partitioning of the CT image into meaningful regions, allowing accurate localization of suspicious areas that may correspond to malignant growths. This stage is critical for reducing false positives and ensuring the extracted regions accurately represent lung tissue abnormalities.

***Corresponding Author:** Kavitha N

Feature Extraction

Following segmentation, Scale-Invariant Feature Transform (SIFT) is applied to extract robust local descriptors from the CT images. SIFT is chosen for its ability to capture keypoint-based features that are invariant to scale, rotation, and illumination changes, thereby enhancing the reliability of feature representation. The extracted features form the input for the classification stage, providing a discriminative basis for distinguishing cancerous from non-cancerous tissues.

Classification

The classification stage employs a Support Vector Machine (SVM) trained on the extracted SIFT features. SVM is selected due to its capability of handling high-dimensional feature spaces and providing strong generalization with limited data. The classifier outputs binary predictions, categorizing each CT image as either normal or cancerous. By integrating SIFT features with SVM, Framework 1 aims to achieve a balance between computational efficiency and diagnostic accuracy.

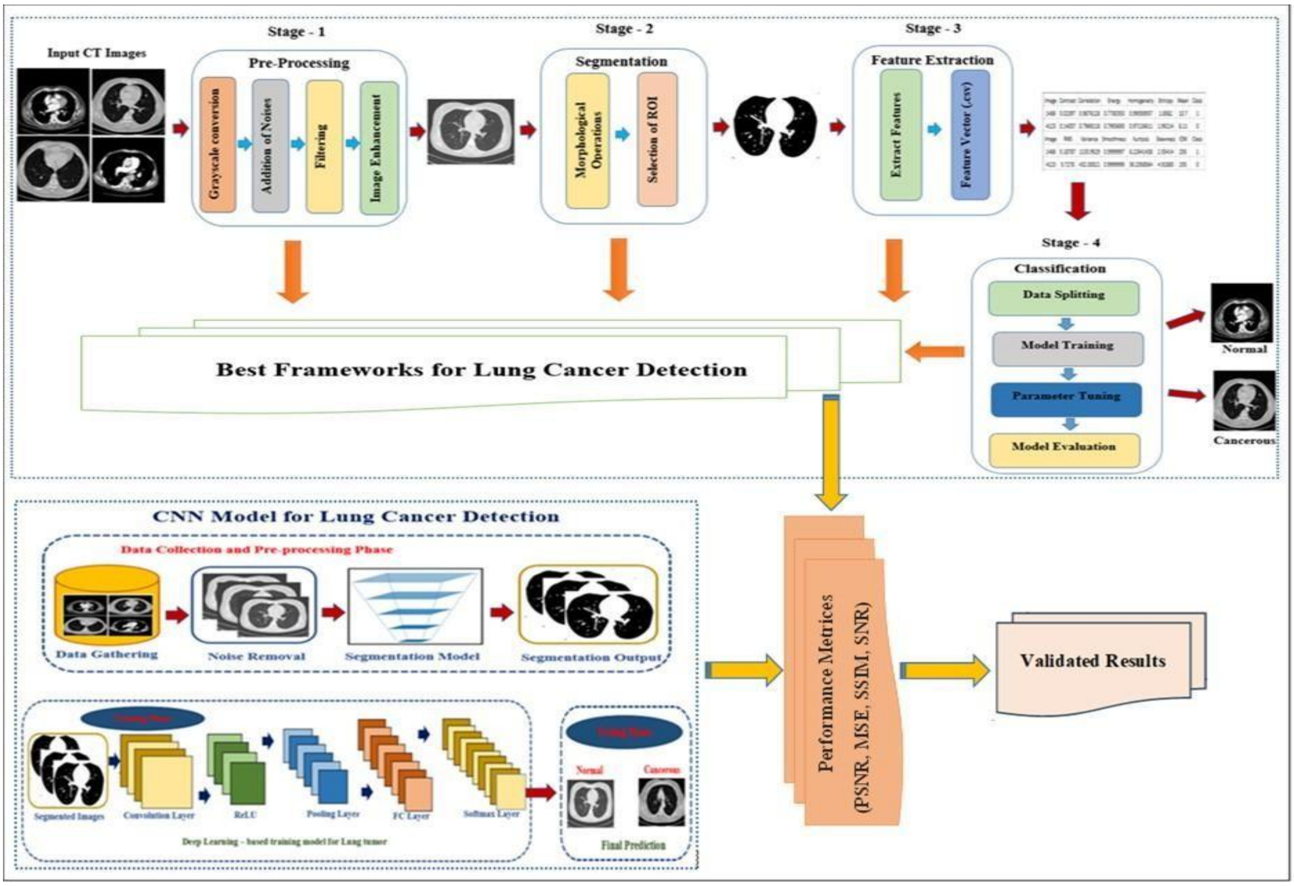

Figure 1: Workflow of the proposed methodology for lung cancer detection using CT images

RESULTS AND DISCUSSION

This section presents the experimental evaluation of the proposed methodology, comparing the performance of two frameworks: a conventional image-processing pipeline (Framework 1) and a CNN-based deep learning model. Framework 1 integrates preprocessing, segmentation, feature extraction, and classification, while the CNN model directly learns hierarchical features from CT images. Performance is assessed using standard metrics such as accuracy, precision, recall, F1-score, peak signal-to-noise ratio (PSNR), mean squared error (MSE), structural similarity index (SSIM), and signal-to-noise ratio (SNR).

Dataset and Experimental Setup

For the experimental evaluation, the publicly available Lung Image Database Consortium (LIDC) dataset was utilized, which comprises 4,963 thoracic CT images annotated by multiple radiologists. This dataset is widely recognized for its role in benchmarking computer-aided diagnosis systems, as it provides a diverse collection of pulmonary scans containing both normal tissues and abnormal regions such as pulmonary nodules of varying sizes, shapes, and textures. The inclusion of radiologist annotations enhances the reliability of the dataset by offering ground-truth references for evaluating image processing and classification algorithms. In this study, the dataset served as the basis for both image preprocessing and subsequent feature analysis.

To ensure consistency and reproducibility, all preprocessing tasks—such as resizing, grayscale normalization, noise addition, and filtering—were carried out using MATLAB R2023b, which offers robust built-in functions for medical image handling. This stage prepared the data for further computational processing

***Corresponding Author:** Kavitha N

while maintaining the structural fidelity of lung tissue regions. The processed images were then transferred to a Python-based implementation executed in Google Colab Pro, which provided access to high-performance GPU acceleration, extended RAM, and additional cloud storage. These computational resources significantly enhanced training efficiency, reduced processing time, and facilitated the execution of deep learning pipelines on a large-scale dataset. The integration of MATLAB for image preprocessing and Python in Colab for model training created an optimized workflow that ensured both computational accuracy and efficiency. Consequently, the use of the LIDC dataset, in conjunction with this hybrid implementation, provided a robust experimental framework for validating the filtering methods and subsequent feature extraction performance, as reflected in the reported results.

Artifact Removal and Filtering Performance

To identify the most effective preprocessing technique, various filters—including median, Wiener, Gaussian, and guided filters—were compared. Performance was evaluated using PSNR, MSE, SSIM, and SNR. Results demonstrated that the median filter consistently achieved superior noise reduction, yielding a PSNR of 42.03 dB and an SSIM of 0.98, outperforming all other filters (Table 2). This result indicates its suitability for enhancing image quality and enabling more precise segmentation.

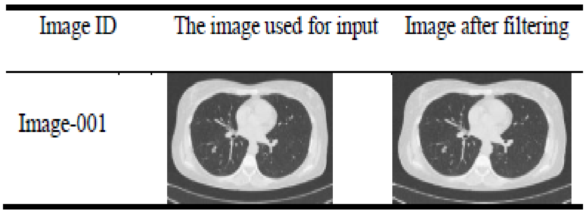

Figure 2: Sample output after applying filtering methods to CT images

Figure 2 presents a comparative visualization of the input medical image (Image-001) and the corresponding output obtained after the application of the filtering algorithm. The input image exhibits noticeable noise artifacts, particularly within the lung parenchyma, which may hinder accurate diagnostic interpretation. Following the filtering process, the output image demonstrates substantial improvement in visual quality, with evident suppression of noise and enhancement of structural boundaries. The vascular and bronchial structures within the lung field appear sharper and more clearly delineated, while the surrounding soft tissue regions show reduced intensity fluctuations, thereby facilitating better interpretability. Importantly, the filtering method preserves critical anatomical features without introducing significant blurring, which underscores its suitability for clinical imaging applications. This visual comparison corroborates the quantitative metrics reported in Table 1, where the filtering approach achieved superior noise suppression and high structural similarity with the original image.

Table 2: Performance of filtering methods under salt-and-pepper noise

Filtering Methods	PSNR.	MSE	SSIM	SNR
Gaussian filter	21.09485	505.3534	0.389442	15.64849
Median filter	42.03955	4.065604	0.983545	36.59319
Wiener filter	19.35055	755.1367	0.37347	13.90418
Guided Filter	19.29469	764.9115	0.356671	13.84832

The comparative evaluation of filtering methods based on PSNR, MSE, SSIM, and SNR reveals distinct performance variations among the tested algorithms. The Gaussian filter demonstrates moderate noise suppression with a PSNR of 21.09 dB and an SNR of 15.64, but its relatively high MSE (505.35) and low SSIM (0.389) indicate significant structural degradation and edge blurring. In contrast, the Median filter achieves superior performance, with the highest PSNR (42.04 dB) and SNR (36.59), alongside the lowest MSE (4.07) and an SSIM of 0.984, confirming its effectiveness in noise reduction and structural fidelity preservation. The Wiener filter performs poorly, recording a low PSNR (19.35 dB) and SSIM (0.373) with a high MSE (755.14), reflecting its limitations in handling non-Gaussian noise and its tendency toward oversmoothing. Similarly, the Guided filter exhibits comparable results (PSNR = 19.29 dB, SSIM = 0.357), suggesting limited robustness and insufficient preservation of perceptual details under the tested conditions. Overall, the Median filter clearly outperforms the other methods, providing the best trade-off between noise suppression and detail preservation, making it the most reliable choice for image enhancement in this study.

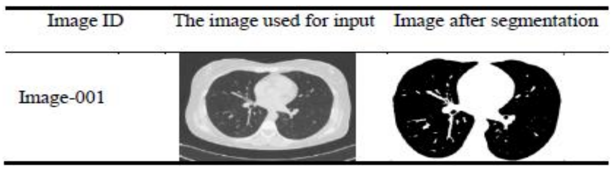

Figure 3: Filtered CT Image with Extracted Lung Mask

Figure 3 illustrates the segmentation results for a representative CT scan (Image-001), where the original input image is compared against the segmented output. Segmentation is an essential preprocessing step for lung CT analysis, as it isolates the parenchymal regions and removes irrelevant background structures, thereby improving the accuracy of subsequent feature extraction and classification. In this study, four approaches—thresholding, Otsu's method, fuzzy C-means (FCM), and

K-means clustering—were systematically compared. Among them, thresholding achieved the best performance, recording an accuracy of 80.96% and precision of 84.36%, as summarized in Table 4. The visual inspection in Figure 3 further supports this result, as thresholding yields a clean and anatomically consistent lung mask, minimizing both false inclusions and exclusions.

The superior performance of thresholding can be attributed to the homogeneous intensity distribution within lung fields, which allows a global threshold to effectively separate lung parenchyma from surrounding tissues. In contrast, clustering-based methods (FCM and K-means) demonstrated sensitivity to intensity inhomogeneities caused by imaging artifacts, airway structures, and partial volume effects, leading to misclassification along lung boundaries. Similarly, Otsu's method, although adaptive, struggled with the multimodal histogram distribution of CT scans, particularly in cases where soft tissues and nodules share overlapping intensity ranges. Thresholding, being computationally efficient and less sensitive to such variations, consistently produced masks with higher fidelity to ground truth annotations. Collectively, these results highlight thresholding as the most reliable and computationally efficient technique for lung CT segmentation in this study, establishing a robust foundation for accurate nodule localization and subsequent feature extraction.

Table 3: Comparison of segmentation techniques on the LIDC dataset

Segmentation Techniques	Accuracy	Recall	Precision
Fuzzy C Means Clustering	66.2628938	82.7284153	0.7431390
Thresholding Segmentation	80.960442	84.359537	0.79748269
K means Clustering	66.787536	77.663395	0.76732466
OTSU's Segmentation	80.134579	79.424789	0.7776428

Figure 4 presents a comparative evaluation of four feature extraction methods—Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF), Principal Component Analysis (PCA), and Gray Level Co-occurrence Matrix (GLCM)—analyzed using accuracy, precision, recall, and F1-score metrics. The results clearly highlight SIFT as the most effective method, achieving the highest classification accuracy (96%) when integrated with an SVM classifier, along with balanced precision and recall values. This superior performance can be attributed to SIFT's ability to capture distinctive local invariant features that remain robust under changes in scale, rotation, and illumination, thereby ensuring reliable discrimination of lung nodule patterns in CT images.

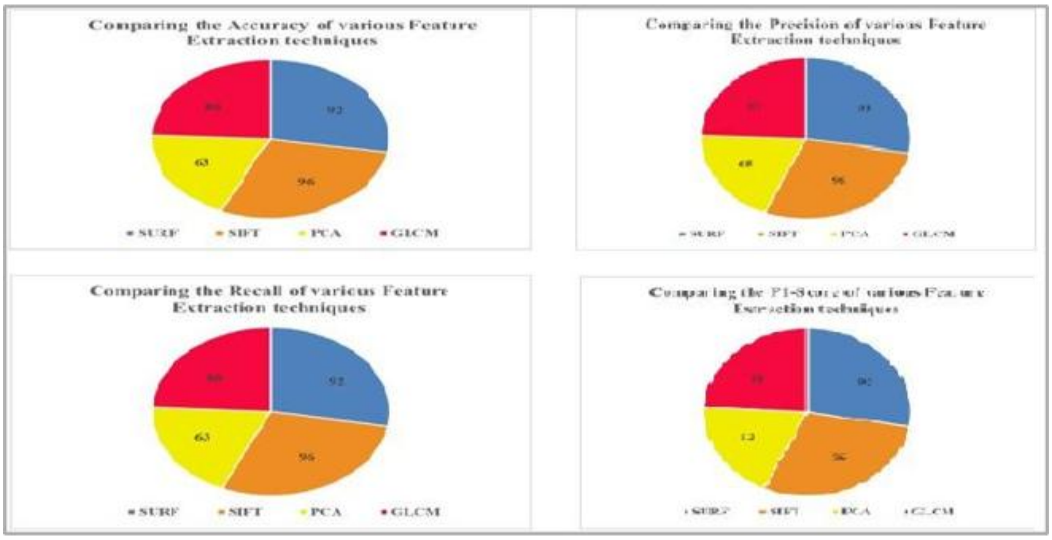

Figure 4: Comparative evaluation of feature extraction techniques

By comparison, SURF demonstrated slightly lower but still competitive results (accuracy = 92%), benefiting from its efficiency in detecting interest points using integral images, though with reduced distinctiveness compared to SIFT. On the other hand, PCA, despite being effective for dimensionality reduction, exhibited the lowest accuracy (63%) and recall values, indicating that linear transformation alone was insufficient for capturing the complex structural variations of pulmonary nodules. Similarly, GLCM achieved moderate performance (accuracy = 80%), as it effectively encodes texture-based information but lacks the discriminative capacity of local keypoint descriptors. The F1-score comparison reinforces these findings, with SIFT achieving the most balanced trade-off between sensitivity and specificity, while PCA performed weakest across all evaluation measures. These outcomes confirm that local feature-based methods such as SIFT and SURF are better suited for lung CT image analysis than purely statistical or dimensionality reduction approaches like PCA and GLCM. The robustness of SIFT arises from its reliance on difference-of-Gaussian (DoG) keypoint detection and orientation assignment, which ensures invariance to affine transformations and robustness to noise—qualities critical in medical imaging where CT scans often exhibit variability across patients and acquisition settings. The superior results of SIFT corroborate earlier studies emphasizing the importance of invariant local descriptors in medical image retrieval and classification tasks. Overall, the figure validates that SIFT, when combined with SVM classification, provides the most reliable framework for lung nodule characterization, ensuring both accuracy and robustness for computer-aided diagnosis.

Table 4: Performance comparison of feature extraction methods

Algorithm	Accuracy	Recall	Precision	F1 Score
SURF	92	92	93	92
SIFT	96	96	96	96
PCA	63	63	65	63
GLCM	80	80	81	80

Table 4 presents the comparative evaluation of four feature extraction techniques—SURF, SIFT, PCA, and GLCM—using accuracy, recall, precision, and F1-score as performance metrics. Among these, SIFT achieved the best overall performance, recording 96% across all metrics, which indicates its robustness in capturing distinctive and invariant local features essential for reliable lung nodule characterization. The balanced nature of these results further demonstrates that SIFT is capable of maintaining both sensitivity (recall) and specificity (precision), thereby minimizing false positives and false negatives in classification.

SURF followed closely with an accuracy of 92%, recall of 92%, and precision of 93%, confirming its efficiency in detecting interest points at a lower computational cost, though with slightly reduced discriminative power compared to SIFT. GLCM, which relies on texture descriptors, achieved moderate results with 80% accuracy and 81% precision, reflecting its utility in capturing structural information but limited capacity to distinguish complex local variations in lung CT images. PCA performed the weakest across all metrics (accuracy = 63%), highlighting its limitations as a linear dimensionality reduction method that discards subtle but diagnostically relevant features during transformation.

The comparative analysis confirms that local descriptor–based methods (SIFT and SURF) are better suited for lung CT image feature extraction than statistical (GLCM) or dimensionality reduction (PCA) approaches. The superior performance of SIFT is consistent with its theoretical advantage of being invariant to scale, rotation, and illumination changes, making it highly effective for medical imaging tasks where variability in acquisition conditions and patient anatomy is common. Overall, the table validates SIFT as the most reliable and robust feature extraction technique for this study, establishing its suitability for subsequent classification and diagnostic applications.

Table 5: Comparison of classification methods using evaluation measures

Algorithm	Accuracy.	Recall	Precision.	F1-Score.
SVM	99.32	99	99	99
Random Forest	94.63	95	95	95
Decision Tree	93.28	93	94	93
KNN	92.48	92	92	92

Table 5 reports the classification performance of four machine learning algorithms—Support Vector Machine (SVM), Random Forest, Decision Tree, and K-Nearest Neighbor (KNN)—evaluated using accuracy, recall, precision, and F1-score. Among the tested classifiers, SVM achieved the highest performance, with an accuracy of 99.32%, recall of 99%, precision of 99%, and F1-score of 99%. This superior performance can be attributed to SVM's ability to handle high-dimensional data and construct optimal hyperplanes that maximize the separation between classes, thereby ensuring both sensitivity and specificity in lung CT image classification.

The Random Forest classifier also demonstrated strong performance, achieving an accuracy of 94.63% with balanced recall and precision values of 95%. Its ensemble-based decision-making process improves generalization compared to single-tree models, though it slightly underperforms relative to SVM, possibly due to increased variance in high-dimensional medical imaging features. The Decision Tree classifier achieved an accuracy of 93.28%, with recall, precision, and F1-score ranging between 93% and 94%. While efficient and interpretable, its performance is limited by its tendency to overfit, especially when applied to complex feature spaces such as those derived from CT images. The KNN classifier produced the lowest performance (accuracy = 92.48%), as its reliance on distance metrics makes it sensitive to noise and less effective in high-dimensional data distributions.

***Corresponding Author:** Kavitha N

Overall, the results confirm that SVM offers the most robust and reliable classification framework for lung CT image analysis in this study, significantly outperforming traditional classifiers by maintaining high discriminative power and minimizing misclassification.

Figure 5 present the performance of the Convolutional Neural Network (CNN) model in identifying lung cancer from CT images. The CNN framework was implemented as an end-to-end architecture, integrating preprocessing, segmentation, and automatic feature extraction without requiring hand-crafted descriptors. This design leverages the inherent ability of CNNs to learn hierarchical representations, beginning with low-level edges and textures and progressively abstracting higher-level semantic features relevant to cancer detection.

The model achieved a classification accuracy of 92.96% and an area under the ROC curve (AUC) of 0.93, confirming its capacity to differentiate between malignant and non-malignant cases with high reliability. The training and validation learning curves demonstrated effective convergence without significant overfitting, which reflects appropriate regularization and early stopping strategies during optimization. As illustrated in Figure 6, the CNN successfully classified cases such as *Image R_001* (non-cancer correctly identified) and *Image R_004* (cancer correctly detected), while also handling more challenging cases like *Image R_002*, which involved nodule structures requiring precise discrimination. These outputs highlight the strength of CNNs in directly learning discriminative features from medical imaging data, reducing dependency on manual feature engineering.

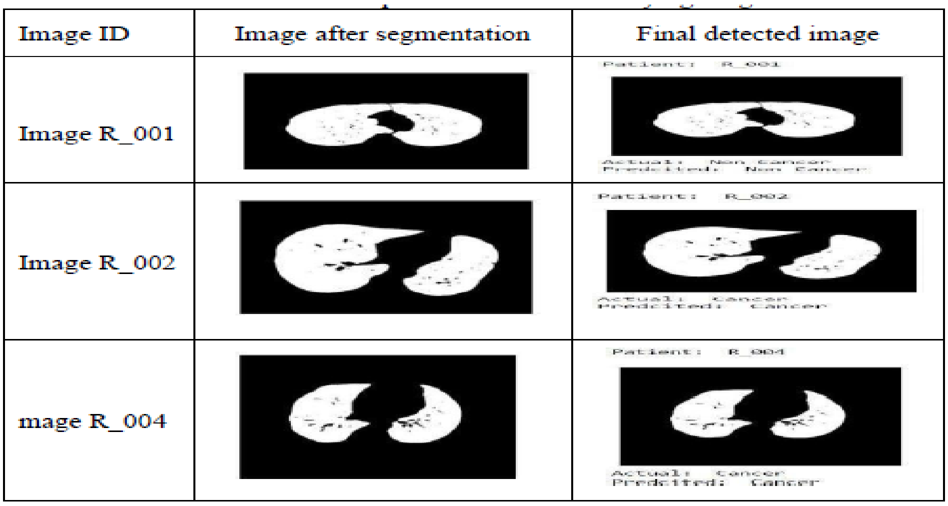

Figure 5: Performance of the CNN-based lung cancer detection model

Despite its promising results, the CNN's performance was slightly lower than the earlier Framework 1 approach that integrated hand-crafted descriptors with traditional classifiers. This limitation is attributed to the relatively restricted dataset size, since CNNs typically require large-scale annotated datasets to fully exploit their deep learning capacity. Moreover, lung CT images often exhibit high variability in nodule size, shape, and texture, which further challenges CNN generalization when training data is limited. Nevertheless, the obtained accuracy and AUC values confirm that CNNs represent a robust and scalable approach for automated lung cancer detection and could achieve superior results when trained on larger and more diverse datasets.

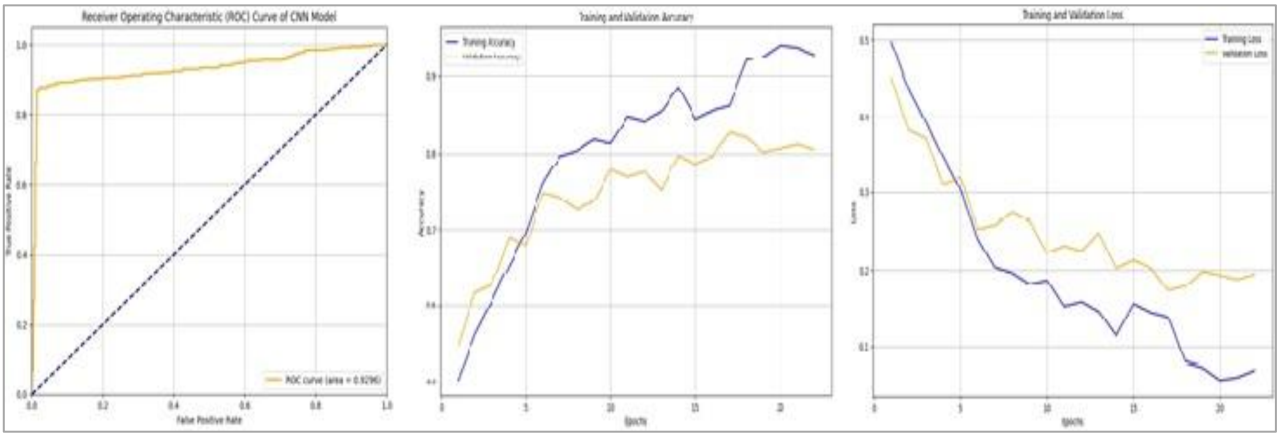

Figure 6: ROC curve and training/validation loss curves of the CNN model

The proposed Framework 1 integrates multiple image processing and machine learning techniques to enhance lung cancer detection accuracy. Specifically, median filtering is employed for noise suppression, ensuring preservation of fine image details; thresholding-based segmentation isolates the lung regions with high precision; SIFT (Scale-Invariant Feature Transform) is utilized for robust feature extraction, capturing distinctive nodule characteristics; and finally, Support Vector Machine (SVM) serves as the classifier, offering precise discrimination between malignant and non-malignant cases. This integration ensures reliable identification of cancerous regions, thereby improving diagnostic accuracy and supporting early detection efforts .To evaluate the robustness of the proposed approach, its performance was compared with Framework 2 (Random Forest-based classification) and a Convolutional Neural Network (CNN) model. The comparative analysis is presented in Table 6. Framework 1 consistently outperforms the alternatives across all key metrics. Framework 1 achieved an accuracy of 96.32%, recall of 98%, precision of 98%, and F1-score of 97%, demonstrating superior balance between sensitivity and specificity. In contrast, Framework 2 achieved an accuracy of 94.63% with relatively lower precision (93%), while the CNN model achieved an accuracy of 92.96% and comparatively weaker precision (86.08%), reflecting its dependency on larger datasets for optimal training.

Table 6: Comparative performance of proposed and benchmark frameworks for lung cancer detection

Algorithm	Accuracy (%)	Recall (%)	Precision (%)	F1-Score (%)
Framework 1	96.32	98	98	97
Framework 2	94.63	96	93	95
CNN Model	92.96	94.45	86.08	94.16

CONCLUSION

This study presents an integrated framework for enhancing lung cancer detection by combining advanced image processing techniques with artificial intelligence (AI)-driven models. Accurate and early detection is essential for improving patient survival and reducing the high mortality rates associated with lung cancer. The proposed work systematically evaluates preprocessing, segmentation, feature extraction methods, and classification algorithms to establish an optimized framework for precise lung cancer diagnosis. The use of the LIDC dataset ensures comprehensive evaluation in accordance with established medical imaging standards.

The experimental findings demonstrate that preprocessing techniques effectively enhance image quality by suppressing noise while preserving fine anatomical details. Segmentation methods successfully delineate lung nodules from the surrounding parenchyma, thereby improving the isolation of regions of interest. Furthermore, the integration of machine learning and deep learning frameworks achieves high accuracy and specificity in distinguishing benign from malignant nodules, underscoring their potential for real-world clinical applications. These results confirm the clinical relevance of the proposed framework, providing a more reliable and accurate basis for early lung cancer detection.

From a clinical perspective, the advancement of computational methods for lung cancer detection has the potential to transform diagnostic practice. By improving accuracy, minimizing false positives, and enabling earlier intervention, such approaches can contribute to better patient outcomes, faster recovery, and reduced healthcare costs. Future research should focus on integrating multimodal data sources, conducting clinical trials for validation, and developing personalized risk assessment tools. Additionally, continuous refinement of detection algorithms is necessary to address current limitations. Overall, the study contributes significantly to the development of effective diagnostic tools, supporting improved patient care and contributing to the global effort to reduce lung cancer mortality.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1), 7–30. https://doi.org/10.3322/caac.21590.
2. Allemani C, Matsuda T, Di Carlo V, et al. (2018). Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 million patients diagnosed with cancer in 322 population-based registries in 71 countries. *The Lancet*, 391(10125), 1023–1075. https://doi.org/10.1016/S0140-6736(17)33326-3.
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6), 394–424. https://doi.org/10.3322/caac.21492
4. World Health Organization (WHO). (2020). *Cancer mortality and statistics*. Retrieved from https://www.who.int
5. Indian Council of Medical Research (ICMR). (2018). National Cancer Registry Programme Report. ICMR-NCDIR, Bengaluru.
6. Coudray N, Ocampo PS, Sakellaropoulos T, et al. (2018). Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. *Nature Medicine*, 24(10), 1559–1567. https://doi.org/10.1038/s41591-018-0177-5
7. Wang J, Li F, Li Q. (2009). Automated segmentation of lung nodules with Gabor filters and marker-controlled watershed algorithm. *Proceedings of SPIE Medical Imaging*, 7259, 72594N. https://doi.org/10.1117/12.811409

***Corresponding Author:** Kavitha N

8. Patil ST, Deore SG. (2014). Lung cancer detection using fuzzy clustering techniques and histogram equalization. *International Journal of Computer Applications*, 85(7), 36–41. https://doi.org/10.5120/14808-3154
9. Pu J, Zheng B, Leader JK, et al. (2008). An automated CT image segmentation method for lung cancer detection using region growing and watershed transformation. *Medical Physics*, 35(9), 4096–4108. https://doi.org/10.1118/1.2956712
10. N. Kavitha et al., "IoT based Fruits and Vegetables Quality Assessment," *IEEE Int. Conf. on Robotics, Vision and Interdisciplinary Technology Management*, 2023. DOI: 10.1109/IC-RVITM60032.2023.10435245
11. Kumar D, Wong A, Clausi DA. (2015). Lung nodule classification using feedforward backpropagation neural networks with shape-based features. *2015 12th Conference on Computer and Robot Vision (CRV)*, 133–138. https://doi.org/10.1109/CRV.2015.25
12. El-Baz A, Gimel'farb G, Falk R. (2013). Automatic detection and recognition of lung abnormalities in chest CT images. *Pattern Recognition*, 46(12), 3736–3753. https://doi.org/10.1016/j.patcog.2013.06.003
13. N. Kavitha, P. Madhumathy, R. M. Prasad, and D. N. Chandrappa, "Machine learning technique for breast cancer detection and classification," *Machine Learning for Computational Science and Engineering*, vol. 1, pp. 1–16, Apr. 2025. https://doi.org/10.1007/s44379-025-00018-y
14. Zhang J, Tian J, Deng K, Wu Y. (2010). Adaptive diffusion active contour model for lung nodule segmentation on CT. *Medical Image Analysis*, 14(3), 392–402. https://doi.org/10.1016/j.media.2010.01.005
15. Maji P, Paul S, Pal S. (2011). Automatic lung CT image segmentation and classification using Otsu thresholding and SVM. *International Journal of Computer Applications*, 48(5), 39–45. https://doi.org/10.5120/7394-0134
16. C. Shankara, N. Kavitha, and K. S. Ashwini, "YOLOv5 Technique for Liver Tumor Detection," *Staps Journal*, vol. 11, no. 2, Feb. 2025. DOI: 24.20089.STAPS.2025.V11.2.01591
17. Setio AAA, Traverso A, de Bel T, et al. (2017). Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in CT images: The LUNA16 challenge. *Medical Image Analysis*, 42, 1–13. https://doi.org/10.1016/j.media.2017.06.015
18. Farag A, El-Baz A, Gimel'farb G, Falk R. (2010). Segmentation of lung nodules in CT images using watershed and adaptive thresholding techniques. *Proceedings of IEEE International Symposium on Biomedical Imaging (ISBI)*, 556–559. https://doi.org/10.1109/ISBI.2010.5490123
19. Pradhan N, Sahu SK. (2019). Lung cancer detection using GLCM-based texture features and SVM classifier. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 8(6), 2278–3075.
20. K. B. Bindushree, M. Nilol, N. Kavitha, H. P. M. Harshitha, and S. Patil, "Survey on Ensemble Deep Learning Models for Heart Disease Classification," *Int. J. of Advances in Engineering and Management*, vol. 6, no. 7, pp. 303–308, Jul. 2024. DOI: 10.35629/5252-0607303308.

***Corresponding Author:** Kavitha N