International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 046-049, January, 2026
ISSN: 2756-3693

# Predictive Modelling For Lung Cancer Detection Using Machine Learning Techniques

**Sangapallavi**
B.Tech Student, Department of Electronics and Computer Engineering, J.B. Institute of Engineering and Technology, Hyderabad, India.
**Mrs. K. Pooja**
Assistant Professor, Department of Electronics and Computer Engineering, J.B. Institute of Engineering and Technology, Hyderabad, India.
pooja.kondan27@gmail.com

***Article Accepted 22nd January 2026***


**Abstract**
*Early and reliable identification of lung cancer plays a critical role in improving patient survival. This paper presents a computer-aided diagnosis framework for the detection of lung cancer from computed tomography (CT) images using digital image processing and classical machine learning techniques. The proposed pipeline consists of image preprocessing, lesion segmentation, feature extraction and supervised classification. Noise reduction and contrast enhancement are applied during preprocessing, followed by edge-based and watershed segmentation to isolate candidate regions. Both region-based and texture-based features are extracted from the segmented images and are used to train Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN) classifiers. Performance is evaluated using accuracy, precision, recall and F1-score. Experimental results indicate that the ANN classifier achieves the best overall performance for both feature categories. The proposed approach demonstrates that a structured feature-driven machine learning framework can effectively support automated lung cancer detection from CT images.*
*Keywords: Lung cancer detection, CT images, image segmentation, feature extraction, machine learning, ANN, SVM, Random Forest.*

1
. **Introduction**

Lung cancer remains one of the leading causes of cancer-related mortality worldwide. The major challenge in clinical practice is the early and reliable detection of malignant nodules from large volumes of medical imaging data. Computed tomography (CT) imaging is widely used for screening and diagnosis; however, manual inspection of scans is time-consuming and subject to inter-observer variability. Consequently, automated and semi-automated decision-support systems have gained significant attention.

Machine learning and digital image processing provide effective tools for extracting relevant patterns from medical images. Preprocessing and segmentation techniques allow the isolation of suspicious regions, while feature extraction and classification algorithms support the identification of benign and malignant nodules. This study focuses on a conventional but interpretable machine learning framework that integrates image processing with feature-based classifiers.

1.**1 Purpose of Machine Learning in Medical Imaging**

Machine learning enables computer systems to learn discriminative patterns directly from data and improve prediction accuracy without explicit rule-based programming. In medical imaging, machine learning is primarily employed to support diagnosis, reduce human error, and enhance the efficiency of clinical workflows.

1.2 Problem Statement

The growing incidence of lung cancer and the limited availability of expert radiologists motivate the development of automated detection systems. Distinguishing benign and malignant nodules in early stages remains a challenging task due to image noise, low contrast and variability in nodule size and appearance.

1.3 **Objectives**

The main objectives of this work are:

- To design a complete lung cancer detection pipeline using CT images.
- To apply preprocessing and segmentation methods for accurate region isolation.
- To extract discriminative region-based and texture-based features.
- To evaluate multiple machine learning classifiers for benign and malignant classification.

2. **Related Work**

International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 046-049, January, 2026
ISSN: 2756-3693

Several studies have explored automated lung cancer detection using machine learning and deep learning approaches. Previous research has employed texture, shape and statistical features combined with classifiers such as SVM and ANN for nodule classification. More recent works focus on convolutional neural networks and multi-modal imaging techniques. However, classical feature-based approaches remain valuable due to their lower computational requirements and higher interpretability, especially in resource-constrained environments.

## 3. **Proposed Methodology**

The proposed framework consists of four major stages: preprocessing, segmentation, feature extraction and classification. The overall workflow is illustrated in Figure 3.1 and the system architecture is shown in Figure 3.2 (figures retained from the original attachment).

### 3.1 **Preprocessing**

CT images are first enhanced using histogram equalization to improve contrast. Median filtering is applied to suppress impulse and acquisition noise while preserving structural edges. The preprocessing stage improves the reliability of subsequent segmentation.

### 3.2 **Segmentation**

Segmentation is performed to identify candidate lung lesion regions. Edge detection using the Prewitt operator is initially applied to highlight boundaries. Thresholding is then used to retain prominent edges. Finally, watershed segmentation based on gradient magnitude is employed to refine the lesion regions.

### 3.3 **Feature Extraction**

Two groups of features are extracted from the segmented regions:

Region-based features

- Area
- Perimeter
- Centroid

Texture-based statistical features

- Mean intensity
- Standard deviation
- Smoothness
- Entropy

These features provide complementary spatial and textural information for classification.

### 3.**4 Classification**

Three supervised learning algorithms are used:

- Support Vector Machine (SVM)
- Random Forest (RF)
- Artificial Neural Network (ANN)

The classifiers are trained using the extracted features to predict whether a nodule is benign or malignant.

## 4. **Dataset Description**

The experiments are conducted using publicly available lung CT datasets, including subsets of the Kaggle Data Science Bowl 2017 dataset and the LUNA16 dataset. These datasets contain three-dimensional CT scans with expert annotations. Due to computational limitations, selected subsets are used for experimentation. The image dimensions are approximately 512 × 512 per slice with 200–300 slices per scan.

## 5**. Experimental Setup and Evaluation**

The dataset is divided into training and testing sets. Performance is evaluated using the following metrics:

- Accuracy
- Precision
- Recall
- F1-score

A confusion matrix is used to analyse classification outcomes.

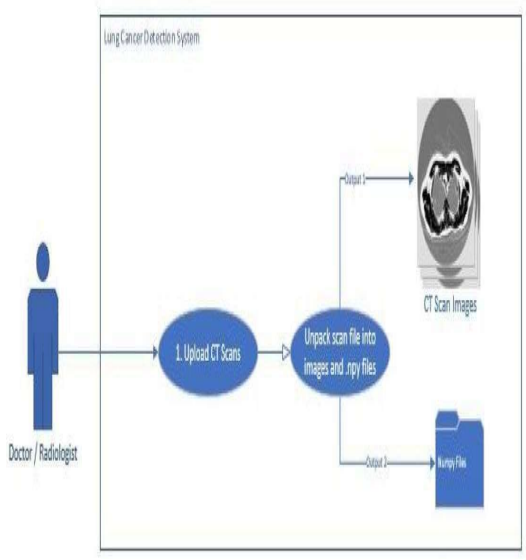
### 5.1 Performance Results

The classification performance obtained using region-based features is summarized in Table 1 (retained from the original attachment).

| Classifier | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| Random Forest | 79% | 100% | 50% | 67% |
| SVM | 86% | 100% | 67% | 80% |
| ANN | 92% | 100% | 69% | 81% |

The corresponding accuracy graph is shown in Figure 5.4.1 (retained from the original attachment). The ANN classifier consistently achieves higher accuracy and balanced performance across all metrics

**Fig Upload CT Scan**



International Journal of Food Safety and Public Health
Vol. 13 (1), pp. 046-049, January, 2026
ISSN: 2756-3693

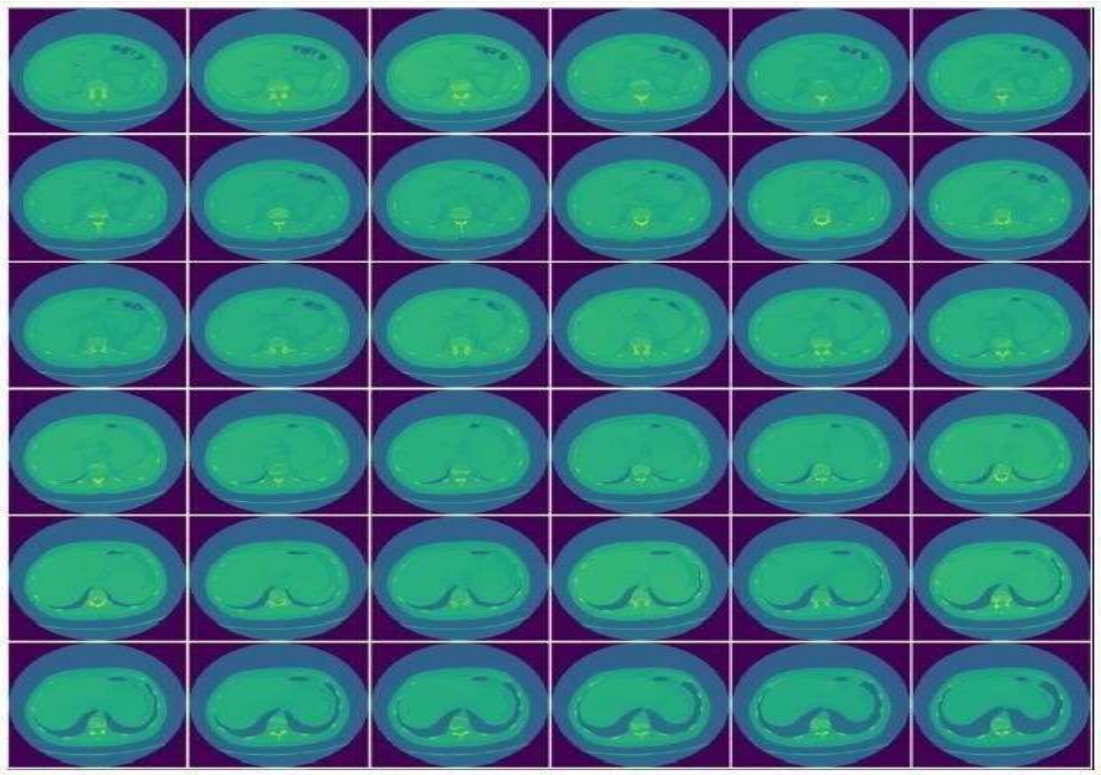


**Fig2 CT Scan Slices**

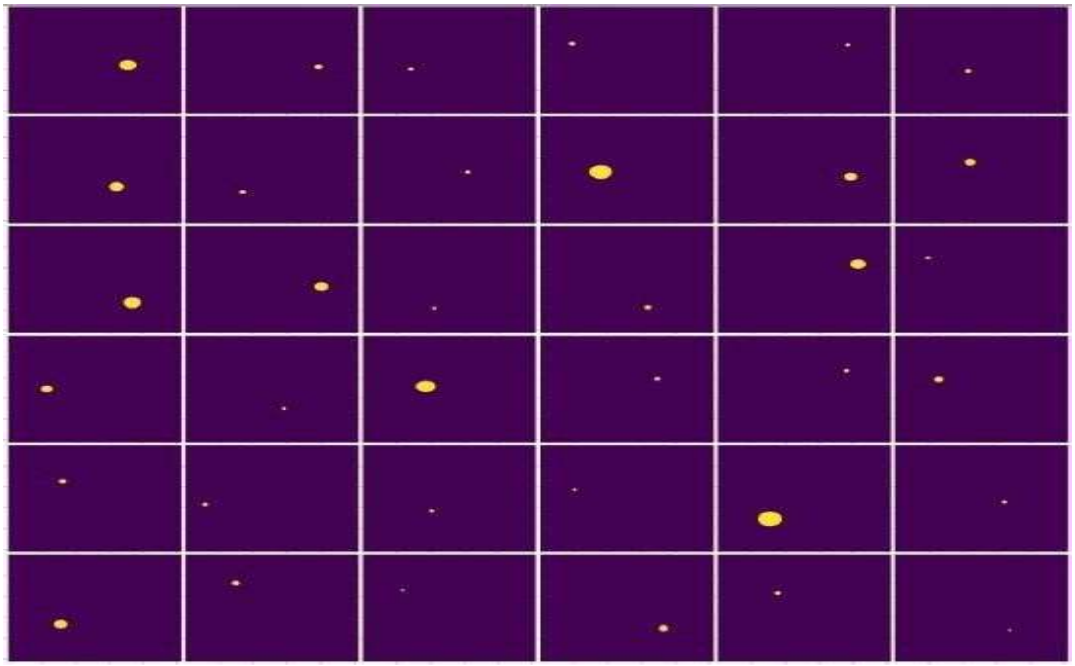


**Fig 3 Cancer Masks**

## 6. Discussion

The experimental results demonstrate that combining digital image processing with conventional machine learning techniques can effectively support lung cancer detection. Texture and region-based features jointly capture important characteristics of lung nodules. While SVM and Random Forest provide competitive results, ANN exhibits superior generalization capability for the considered feature space.

The segmentation quality directly influences classification performance. Errors in edge detection or watershed segmentation may lead to incomplete or inaccurate lesion regions, which in turn affect feature reliability.

## 7. Conclusion

This paper presented a feature-based machine learning framework for lung cancer detection using CT images. The system integrates preprocessing, segmentation, feature extraction and supervised classification. Experimental evaluation shows that the ANN classifier provides the highest accuracy among the evaluated models. The proposed approach offers an efficient and interpretable solution suitable for clinical decision-support systems.

## 8. Future Work

Future work will focus on integrating deep learning-based segmentation and classification models to further improve detection accuracy. Hybrid models combining handcrafted features with convolutional neural network features will also be explored. In addition, larger and more diverse datasets will be employed to improve model robustness and clinical reliability.

## REFERENCES:

[1] Krishnaiah, V., G. Narsimha, and Dr N. Subhash Chandra. "Diagnosis of lung cancer prediction system using data mining classification techniques." International Journal of Computer Science and Information Technologies 4.1 (2013): 39-45.

[2] Zhang, Junjie, et al. "Pulmonary nodule detection in medical images: a survey." Biomedical Signal Processing and Control 43 (2018): 138- 147.

[3] Fenwa, Olusayo D., Funmilola A. Ajala, and A. Adigun. "Classification of cancer of the lungs using SVM and ANN." Int. J. Comput. Technol. 15.1 (2016): 6418-6426.

[4] Daoud, Maisa, and Michael Mayo. "A survey of neural network-based cancer prediction models from microarray data." Artificial intelligence in medicine (2019).

[5] Palani, D., and K. Venkatalakshmi. "An IoT based predictive modelling for predicting lung cancer using fuzzy cluster based segmentation and classification." Journal of medical systems 43.2 (2019): 21.

[6] Lynch, Chip M., et al. "Prediction of lung cancer patient survival via supervised machine learning classification techniques." International journal of medical informatics 108 (2017): 1-8.

[7] Öztürk, Şaban, and Bayram Akdemir. "Application of feature extraction and classification methods for histopathological image using GLCM, LBP, LBGLCM, GLRLM and SFTA." Procedia computer science 132 (2018): 40-46.

[8] Jin, Xin-Yu, Yu-Chen Zhang, and Qi-Liang Jin. "Pulmonary nodule detection based on CT images using convolution neural network." 2016 9th International symposium on computational intelligence and design (ISCID). Vol. 1. IEEE, 2016.

[9] Sumathipala, Yohan, et al. "Machine learning to predict lung nodule biopsy method using CT image features: A pilot study." Computerized Medical Imaging and Graphics 71 (2019): 1-8.

[10] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," CA: A Cancer Journal for Clinicians, vol. 61, no. 2, pp. 69–90, 2013.

[11] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," CA: A Cancer Journal for Clinicians, vol. 68, no. 1, pp. 7–30, 2018

[12] L. A. Jemal, R. L. Siegel, and A. Jemal, "Lung cancer statistics," in Lung Cancer and Personalized Medicine, vol. 893, pp. 1–19, Springer, Berlin, Germany, 2016.

[13] C. I. Henschke, D. I. Mccauley, D. F. Yankelevitz et al., "Early lung cancer action project: overall design and findings from baseline screening," The Lancet, vol. 354, no. 9173, pp. 99–105, 1999.

[14] A. K. Alzubaidi, F. B. Sideseq, A. Faeq, and M. Basil, "Computer aided diagnosis in digital pathology application: review and perspective approach in lung cancer classification," in Proceedings of the New Trends in Information & Communications Technology Applications, pp. 219–224, IEEE, Baghdad, Iraq, March 2017.

[15] W. Sun, B. Zheng, and Q. Wei, "Computer aided lung cancer diagnosis with deep learning algorithms," in Proceedings of the Medical Imaging: Computer-Aided Diagnosis, vol. 9785, p. 97850Z, San Diego, CA, USA, March 2016.

[16] K. Kuan, M. Ravaut, G. Manek et al., "Deep learning for lung cancer detection: tackling the kaggle data science bowl 2017 challenge," 2017, https://arxiv.org/abs/1705.09435.