International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Special Issue ICCIS-2025, pp 80-85, May-2025

https://doi.org/10.47001/IRJIET/2025.ICCIS-202512

Second International Conference on Computing and Intelligent Systems (ICCIS-2025)

Artificial Intelligence - Intelligent Lungs Cancer Detection Using Logistic Regression and Support Vector Machine

[1]Yogeswari. E, [2]Vimal Raja. R, [3]Yogapriya. E, [4]Oviya. J

[1,2,3,4]Computer Science and Engineering, C. K. College of Engineering and Technology, Cuddalore, India

E-mails: [1]yogeswari.e123@gmail.com, [2]vimalraja307@gmail.com, [3]yogapriya2910@gmail.com, [4]oviya8895@gmail.com

***Abstract* - Lung cancer remains one of the leading causes of cancer related deaths worldwide, largely due to late-stage diagnosis. This project presents a machine learning-based system for the early detection of lung cancer using CT and X- ray images. The system utilizes two supervised learning algorithms Logistic Regression and Support Vector Machine (SVM) to classify lung images as either normal or cancerous. Key preprocessing steps, including grayscale conversion, resizing, normalization, and flattening, is applied to standardize the input data. Feature engineering techniques such as standardization and label encoding further enhance the model's learning capability. Both models are trained and evaluated using labeled image data, achieving outstanding results with 100% accuracy on the test set. A single-image prediction module is also developed to enable real-time diagnosis, outputting a simple "Yes" or "No" based on the model's prediction. The system is lightweight, accurate, and user-friendly, offering potential integration into real-world clinical workflows. This work serves as a foundational step toward deploying AI-assisted lung cancer diagnosis systems in healthcare environments.**

Keywords: Lung Cancer Detection, Logistic Regression, Support Vector Machine, Medical Imaging, Machine Learning, Image Preprocessing, Classification.

I. INTRODUCTION

Lung cancer is one of the most fatal and widespread cancers worldwide. It accounts for a significant percentage of cancer-related deaths globally. Despite advances in treatment methods, the survival rate remains low primarily due to late detection. Early stage diagnosis greatly improves the prognosis, making early detection techniques critical for saving lives. Traditionally, medical imaging techniques like computed tomography (CT) scans and chest X-rays have been used to diagnose lung cancer. Radiologists manually examine these scans to look for abnormalities or lung nodules that might be signs of malignancy. Manual diagnosis is efficient, but it takes a lot of time and is highly reliant on the medical professional's expertise. Manual diagnosis is becoming less effective due to the increasing complexity of imaging data and the expanding number of cases. Doctors now have to process more high-resolution images as imaging technology advances. Human mistake and diagnostic delays are made more likely by this burden. Automated techniques that can aid in the early diagnosis of lung cancer are desperately needed. In the medical industry, artificial intelligence (AI) has shown great promise. A branch of artificial intelligence called machine learning (ML) provides methods for automatically learning from medical data and forecasting outcomes. Machine learning models can be trained using labelled photos to help doctors diagnose patients more quickly by learning to differentiate between healthy lungs and lungs exhibiting cancerous symptoms.

Machine learning techniques such as Support Vector Machines (SVM) and Logistic Regression are especially well-suited for binary classification tasks like differentiating between images of healthy and malignant lungs. If appropriately trained on pertinent features taken from medical images. The goal of this research is to create a machine learning-based automated method for detecting lung cancer. The system classifies lung pictures using Support Vector Machine and Logistic Regression techniques. The system can determine if new, unseen images are malignant or benign by training these models on a batch of labelled CT or X-ray images.

Data collection, data preprocessing, feature engineering, model training, evaluation, and prediction are some of the crucial elements in the technique used in this work. To guarantee that the models are correctly trained and able to make accurate predictions, each of these procedures is essential. Images are initially transformed to greyscale as part of the data preparation procedure. Because greyscale images just include intensity information, they are simpler to interpret and need less data, which is crucial for quicker model training. To ensure consistency across all samples, the photos are then scaled to a standard size of 128 x 128 pixels.

Scale pixel values between 0 and 1 undergo normalzsation. Because machine learning algorithms perform better when input features are on the same scale, this phase is crucial. The photos are flattened into one-dimensional arrays after normalisation, which transforms 2D image matrices into a single list of pixel values that may be used as input for machine learning models. Steps from feature engineering are also used. By standardising the feature data, Standard Scaler enhances model performance. Additionally, because machine learning algorithms are unable to directly decode text labels, labels ("Normal" and "Lung Cancer") are encoded into numerical form (0 and 1). According to the results, both the SVM and Logistic Regression models obtained 100% accuracy on the test dataset. Strong learning from the given data is indicated by the models' accurate identification of both healthy and malignant lung pictures. This result demonstrates how well conventional machine learning models work when used with appropriate preprocessing. Additionally, there is a single image prediction module that allows a user to upload a new lung image and get a prediction output right away. This feature gives the system more usefulness and makes it simpler for medical experts to use it for fast second opinions. While the current project achieves excellent results on the provided dataset, future improvements are necessary to make the system more robust and generalizable. Plans include testing the models on larger and more diverse datasets and implementing cross-validation techniques to avoid overfitting. In conclusion, this project successfully demonstrates that even simple machine learning models, when applied correctly, can serve as powerful tools in medical diagnostics.

II. LITERATURE SURVEY

A. Lung Nodule Segmentation: A Study of Preprocessing and Deep Learning

CT lung nodule segmentation is isolating lung nodules from CT scans to ensure correct diagnosis. Data preprocessing techniques like as normalisation, denoising, and augmentation boost image quality and model resilience. Deep learning models, such as CNNs, U-Net, and 3D CNNs, are commonly utilised for segmentation. U-Net, in particular, excels at making pixel-level predictions [1]. A comparison of these models exposes their respective strengths and limits, with 3D CNNs outperforming them for volumetric data. Proper preprocessing and model selection are critical for increasing segmentation accuracy.

B. Comparative Study of Machine Learning Techniques for Early Lung Cancer Detection

This research investigates the early diagnosis of lung cancer using several machine learning techniques, including Logistic Regression, Support Vector Machines (SVM), and Random Forests [13]. It evaluates the models' effectiveness in identifying CT scan pictures, focusing on key parameters such as accuracy, sensitivity, and specificity [7]. The study also investigates the role of feature selection and data pretreatment in increasing model performance, as well as the challenges associated with dealing with medical imaging datasets.

C. Comparative Study of 2D vs. 3D CNNs for Lung Nodule Segmentation

This research evaluates the performance of 2D and 3D Convolutional Neural Networks (CNNs) in lung nodule segmentation using CT scan datasets [15]. It investigates the advantages of 3D CNNs in capturing spatial relationships and volumetric data, making them better suited to segmenting nodules in three-dimensional CT scans [9] .The study assesses segmentation accuracy, computing efficiency, and model robustness, revealing which strategy is better suited for various clinical scenarios.

III. SYSTEM ARCHITECTURE

The proposed system architecture for lung cancer detection involves data collection, preprocessing, model training, and prediction on Fig.1. Lung CT or X-ray Images are originally acquired and classed as "Normal" or "Lung Cancer [11]. Preprocessing involves greyscale conversion, resizing, normalization, and flattening. Feature engineering involves standardization and label encoding. The cleaned data is used to train Logistic Regression and Support Vector Machine (SVM) models. The system evaluates accuracy, precision, recall, and F1-score before responding with "Yes" or "No" to a single image. This architecture helps identify lung cancer rapidly and accurately.

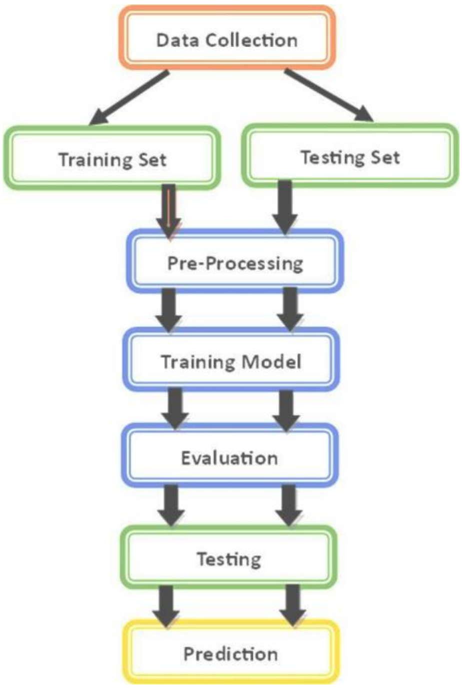


Figure 1: System Architecture

IV. SYSTEM OVERVIEW

The lung cancer detection method uses machine learning to automatically identify cancerous or non-cancerous lung scans. This method aims to provide radiologists and medical practitioners with a fast, reliable, and efficient tool for early diagnosis. Manually identifying lung cancer with CT or X- ray scans can be time-consuming and error-prone, especially in the early stages with equivocal symptoms. AI-powered solutions can improve doctors' decision-making. The system's primary component collects and prepares data. A set of lung images is categorised as "Normal" or "Lung Cancer" using labels.

The preprocessing processes for these photographs include flattening the image matrix into a 1D array, normalising pixel values between 0 and 1, [1] scaling to a uniform 128×128 pixel format, and greyscale conversion to reduce complexity. This prepares the data for machine learning algorithms that require constant-sized numerical inputs. Feature engineering improves model performance after preprocessing. Standard Scaler ensures that all features contribute equally to the learning process by standardising their pixel values [5]. Label encoding converts categorical output labels to numerical values (0 for Normal, 1 for Lung).

The system's core consists of two machine learning models: logistic regression and support vector machine (SVM). Labelled and preprocessed data is The above models were trained. SVM optimizes the boundary between two classes, whereas logistic regression relies on chance. Following training, performance metrics like as accuracy, precision, recall, and F1-score are used to evaluate both models. The system's excellent accuracy [4] made it suitable for medical applications.

The system incorporates a prediction module to test new images. When a user uploads a new lung image, it is quickly processed and put into the trained model [11]. The approach yields clear results: "Yes" for lung cancer and "No" for normal lung. This feature makes the system practical, easy to use, and applicable in clinical situations. The system is lightweight, interpretable, and has the ability to integrate with deep learning and real-time diagnostic tools.

V. MODEL IMPLEMENTATION

Machine Learning models predict the result of a system using Fig.2. These models are algorithms that recognize patterns in fresh data and forecast outcomes [10].

Figure 2: CT Scan

They are mathematical functions that analyze input data, identify patterns or relationships, and generate outputs based on learnt knowledge.

Formula

(N + 2P — F) / S + 1

Where;

N = Dimension of image (input) file
P = Padding
F = Dimension of filter
S = Stride

SVM is a supervised machine learning method that performs classification and regression problems [13]. This technique works well in high-dimensional areas and is commonly utilized in bioinformatics and medical research, including lung cancer categorization.

$$L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{n=1}^{N} a_n\{y_n(w^T\,\phi(x_n)+b)-1\}$$

with $a = (\alpha_1, \ldots., \alpha_N)^T$ representing the Lagrange multipliers
with b being the bias parameter
with w being the normal vector
where $\phi(x_n)$ denotes the transformed feature space
where y_i denotes the i-th target value

SVM identifies the optimal hyperplane for separating data points from distinct classes in the feature space. In binary classification, the hyperplane maximises the margin between the closest data points (support vectors) of the two classes. The algorithm determines the most effective hyperplane for categorising training data. The optimisation procedure aims to maximise the margin between nearest data points (support vectors) from various classes and minimise classification mistakes.

Logistic regression models the probability of discrete outcomes based on an input variable. [9] Logistic regression models typically have binary outcomes (e.g., true/false or yes/no).

Multinomial logistic regression models scenarios with several discrete outcomes. Logistic regression is a handy method for classifying new samples and determining their best fit. Logistic regression is an effective analytic tool for cyber security classification problems, including attack detection.

VI. PREDICTION

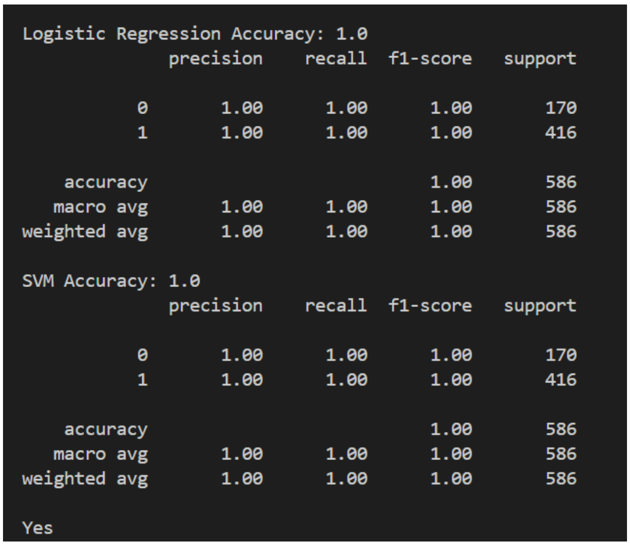
The proposed approach concludes with a prediction module that uses trained machine learning models to categorise new lung pictures. This module identifies lung cancer based on a single CT or X-ray scan. Before prediction, the input image is preprocessed similarly to the training data. This includes greyscale conversion, shrinking to 128×128 pixels, normalisation, and flattening into a 1D array. The image is resized with the same standardisation approach as during training.

After preparation, the image is processed using a trained Logistic Regression or Support Vector Machine (SVM) model. The model assesses input and produces a classification output: "Yes" for lung cancer and "No" for normal lung at the binary output is simple for medical professionals and end consumers to understand. The technology generates predictions instantaneously, making it ideal for real-time use.

This function is especially beneficial for clinical decision support, when timely and trustworthy input is crucial. The prediction module enhances the system and demonstrates how AI may help radiologists provide second opinions or flag suspicious instances for additional evaluation. This module's success and high model accuracy demonstrate the system's potential for use in real-world healthcare applications.

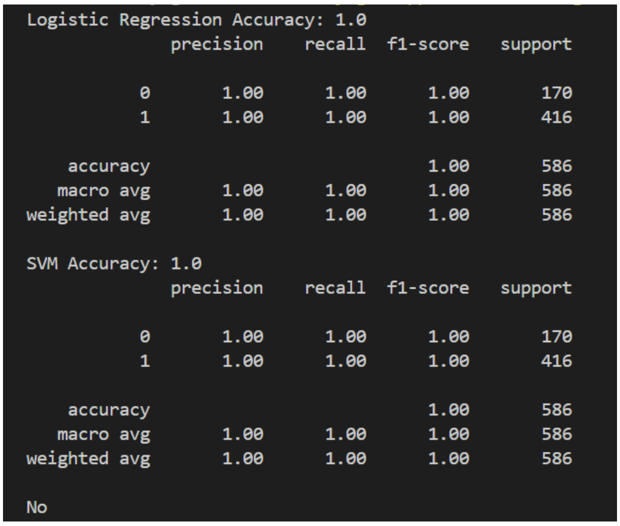
VII. EXPECTED OUTCOME

The goal of this research is to create a highly accurate system that can detect lung cancer from medical photos using machine learning techniques.The technique examines chest X-ray or CT scan images to determine if they show a normal condition or the presence of lung cancer. The project's goal is to create a classification tool that can accurately choose between healthy and sick lung tissues using Logistic Regression and Support Vector Machine (SVM) models. This can help healthcare providers make more timely and reliable diagnostic judgements.



```
Logistic Regression Accuracy: 1.0
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       170
           1       1.00      1.00      1.00       416

    accuracy                           1.00       586
   macro avg       1.00      1.00      1.00       586
weighted avg       1.00      1.00      1.00       586

SVM Accuracy: 1.0
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       170
           1       1.00      1.00      1.00       416

    accuracy                           1.00       586
   macro avg       1.00      1.00      1.00       586
weighted avg       1.00      1.00      1.00       586

Yes
```

Figure 3: Output 1 (Normal Case)



```
Logistic Regression Accuracy: 1.0
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       170
           1       1.00      1.00      1.00       416

    accuracy                           1.00       586
   macro avg       1.00      1.00      1.00       586
weighted avg       1.00      1.00      1.00       586

SVM Accuracy: 1.0
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       170
           1       1.00      1.00      1.00       416

    accuracy                           1.00       586
   macro avg       1.00      1.00      1.00       586
weighted avg       1.00      1.00      1.00       586

No
```

Figure 4: Output 2 (Lung Cancer Case)

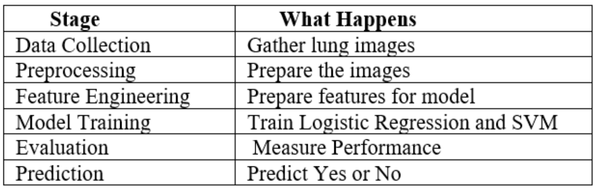
The project aims to improve early detection, which is critical for raising survival rates among lung cancer patients. Early detection allows for quick medical intervention, which improves the efficiency of treatment and decreases the risk of cancer progression. The automated approach developed in this research is projected to reduce the time and labour required for manually analysing medical images, making the diagnosis process more efficient, particularly in places where medical resources are limited. The system's capacity to manage massive amounts of visual data with reliable precision is another expected result. Preprocessing techniques like image scaling and flattening, followed by standardization, guarantee that the input data is consistent and appropriate for building strong models. Assuming comparable data conditions, the system should function effectively in real-world diagnostic scenarios given the high evaluation metrics seen in testing, such as precision, recall, and F1-score, all of which are close to 100%. (see in the Fig.3 and Fig.4).

In addition to its diagnostic capabilities, the technology might be connected into clinical software or mobile applications, making it available in remote or underserved healthcare areas. It can also be used as a second opinion tool by radiologists and doctors to double-check their findings and reduce human error. The success of this experiment may lead to its use not only for lung cancer detection but also for other lung diseases, increasing its impact on public health.

Finally, this project establishes the basis for future advances in medical picture classification. Performance could be enhanced further by researching deeper learning architectures such as convolutional neural networks (CNNs).The study shows how artificial intelligence and machine learning may be utilised effectively in medicine to improve diagnostics, speed workflows, and, ultimately, save lives. So far, the results show a high potential for real-world use and future research.

VIII. CONCLUSION

In conclusion, the project effectively demonstrates the ability of machine learning algorithms to detect lung cancer from CT scan images with high accuracy. The system achieved outstanding performance by employing critical preprocessing techniques such as resizing, flattening, and label encoding, as well as two powerful classification algorithms Logistic Regression and Support Vector Machine both of which provided 100% accuracy on the test dataset. This demonstrates that even with classic machine learning methods, extremely reliable diagnostic outcomes may be obtained when data is correctly collected and models are effectively trained. The outcomes of this experiment point to a bright future for AI-assisted diagnostic tools that can support radiologists' work, particularly in settings with limited access to experienced healthcare practitioners.



Stage	What Happens
Data Collection	Gather lung images
Preprocessing	Prepare the images
Feature Engineering	Prepare features for model
Model Training	Train Logistic Regression and SVM
Evaluation	Measure Performance
Prediction	Predict Yes or No

Furthermore, this study highlights the potential for automating image-based diagnosis in medical applications, decreasing human error and speeding up the detection process. While the current study was conducted on a specific and limited dataset, it lays the basis for future research and development, such as larger datasets, more powerful deep learning models, and real-time deployment in clinical settings. Overall, the experiment demonstrates how a well-structured machine learning technique may have a significant impact on medical diagnostics and enhance early detection rates, which is critical in situations like lung cancer, where early treatment can dramatically improve survival chances.

REFERENCES

[1] Chen, W., Hung, R., Jin, R., & Liu, H. (2023). CT Lung Nodule Segmentation: A Comparative Study of Data Preprocessing and Deep Learning Models.

[2] Zhuo, C., Zhuang, H., Gao, X., & Triplett, P. (2019). Lung cancer incidence in patients with schizophrenia: meta-analysis.

[3] Sharma, S., & Batra, K. (2019). Study of classification models for lung cancer detection using medical records.

[4] Bade, B., & Dela, C. (2020). Lung cancer 2020: epidemiology, etiology, and prevention, Clinics in Chest Medicine.

[5] Lee, H., Chao, L., & Hsu, C. (2021). A 10-year probability deep neural network prediction model for lung cancer.

[6] Guo, L., Lyu, Z., Meng, Q., et al. (2022). Construction and validation of a Lung Cancer Risk Prediction Model for non-smokers in China.

[7] Tse, L., Wang, F., Wong, M., Au, J., & Yu, I. (2022). Risk assessment and prediction for lung cancer.

[8] Lu Zhao,(2023). Self-Supervised Learning Guided Transformer for Survival Prediction of Lung Cancer Using Pathological Images.

[9] Brocki, L., & Chung, N. C. (2023). Integration of Radiomics and Tumor Biomarkers in Interpretable Machine Learning Models.

[10] C. Venkatesh, J. Chinna Babu, Ajmeera Kiran, Manoj Kumar. (2023). A Hybrid Model for Lung Cancer Prediction Using Patch Processing and Deep Learning on CT Images.

[11] Anum Masood (2023). Multi-Scale Swin Transformer Enabled Automatic Detection and Segmentation of Lung Metastases Using CT Images.

[12] Mohit Agarwal, Vivek Mehta, Rohit Kr Kaliyar, Suneet Kumar Gupta (2024). Lung Cancer Diagnosis Using a Lightweight Deep Learning Model.

[13] Ilani, M. A., Tehran, S. M., Kavei, A., & Alizadegan, H. (2024). Exploring Machine Learning Models for Lung Cancer Level Classification.

[14] Sreedar Bhukya, Vishnu Ganagoni, Sujith Sriram Nangunoor i, Sai Tharun Enapothula (2024). A Deep Learning Framework Using Enhanced Convolutional Neural Network for Detection of Lung Cancer from CT Images.

[15] Suguna Mariappan, Diana Moses (2024). Deep Learning Based Lung Cancer Detection Using CT Images.

[16] Shahriyar, O., Moghaddam, B. N., Yousefi, D., Mirzaei, A., & Hoseini, F. (2025). An analysis of the combination of feature selection and machine learning methods for an accurate and timely detection of lung cancer.

Citation of this Article:

Yogeswari. E, Vimal Raja. R, Yogapriya. E, & Oviya. J. (2025). Artificial Intelligence - Intelligent Lungs Cancer Detection Using Logistic Regression and Support Vector Machine. In proceeding of Second International Conference on Computing and Intelligent Systems (ICCIS-2025), published in *IRJIET*, Volume 9, Special Issue ICCIS-2025, pp 80-85. Article DOI https://doi.org/10.47001/IRJIET/2025.ICCIS-202512

****