



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

Predictive Analysis of Lung Cancer using CT Image Processing

**Dr. Meeravali Shaikh¹, Mr. Giduthuri Kishore Kumar², Mr. S Ashwanth³, Mr. D Abhilash⁴,
Mr. U Aman Kumar⁵, Mr. B Akash⁶**

Professor, Department of Computer Science, Malla Reddy University, Maisammaguda, Hyderabad, India¹

Assistant Professor, Department of Computer Science, Malla Reddy University, Maisammaguda,
Hyderabad, India²

Student, Department of Computer Science, Malla Reddy University, Maisammaguda, Hyderabad, India^{3,4,5,6}

ABSTRACT - Medical imaging tools are essential in early-stage lung cancer diagnostics and the monitoring of lung cancer during treatment. Various medical imaging modalities, such as chest X-ray, magnetic resonance imaging, positron emission tomography, computed tomography, and molecular imaging techniques, have been extensively studied for lung cancer detection. These techniques have some limitations, including not classifying cancer images automatically, which is unsuitable for patients with other pathologies. Deep learning is one of the fastest-growing topics in medical imaging, with rapidly emerging applications spanning medical image-based and textural data modalities. With the help of deep learning-based medical imaging tools, clinicians can detect and classify lung nodules more accurately and quickly. Therefore, this work implements the advanced modifications in CNN model for the detection of lung cancer from chest scan images. The proposed CNN model is able to classify the being and malignant i.e., normal, and cancerous with higher accuracy as compared to state-of-the-art machine learning approach called support vector machine (SVM) classifier.

KEYWORDS: Support Vector Machine (SVM), CT Scans, Convolutional Neural Networks (CNNs), Image Preprocessing.

I. INTRODUCTION

Lung cancer is one of the most common and deadliest forms of cancer, with a high mortality rate primarily due to late diagnosis. Early detection plays a crucial role in improving survival rates, but traditional diagnostic methods, such as radiologists manually analyzing CT scan images, can be time-consuming and prone to human error. Given the complexity of medical imaging, there is a need for a more accurate, efficient, and automated system that can assist healthcare professionals in detecting lung cancer at an early stage.

This project aims to develop an automated lung cancer detection system using deep learning and machine learning techniques. A hybrid approach is implemented, combining convolutional neural networks (CNNs) and support vector machines (SVMs) to improve classification accuracy. The CNN model, based on VGG16, is used to extract deep features from CT scan images, enabling the detection of complex patterns that may not be easily identifiable through traditional methods. Meanwhile, SVM serves as an alternative classification model, utilizing extracted features to distinguish between benign and malignant nodules effectively.

The system follows a structured workflow that begins with image preprocessing, where CT scan images are resized, normalized, and enhanced to improve model performance. Once the images are prepared, they are passed through the VGG16-based CNN,

extracts deep features that are crucial for classification. These extracted features are then fed into two separate classification models: the CNN classifier and the SVM classifier. The CNN classifier utilizes fully connected layers to predict whether the lung nodule is benign or malignant, while the SVM classifier takes the extracted features and performs a similar classification. This dual-model approach ensures robust and reliable predictions.

To make the system accessible and user-friendly, a frontend interface is integrated, allowing users to upload CT scan images and receive real-time predictions. This feature is particularly useful for medical professionals who require a quick and accurate assessment of lung nodules. By automating the detection process, this system significantly reduces diagnostic time and minimizes the risk of human error.

The combination of deep learning and traditional machine learning ensures that the system is both powerful and interpretable, leveraging the strengths of CNNs in feature extraction and SVMs in classification. This hybrid approach enhances accuracy, reliability, and efficiency, making the system a valuable tool for radiologists and healthcare professionals. Ultimately, the goal of this project is to contribute to early lung cancer detection, improving patient outcomes and increasing survival rates through advanced AI-driven diagnostics.

II. LITERATURE REVIEW

Lung cancer is the primary cause of cancer death worldwide, with 2.09 million new cases and 1.76 million people dying from lung cancer in 2018 [1]. Four case-controlled studies from Japan reported in the early 2000s that the combined use of chest radiographs and sputum cytology in screening was effective for reducing lung cancer mortality. In contrast, two randomized controlled trials conducted from 1980 to 1990 concluded that screening with chest radiographs was not effective in reducing mortality in lung cancer [2, 3]. Although the efficacy of chest radiographs in lung cancer screening remains controversial, chest radiographs are more cost-effective, easier to access, and deliver lower radiation dose compared with low dose computed tomography (CT). A further disadvantage of chest CT is excessive false positive (FP) results. It has been reported that 96% of nodules detected by low-dose CT screening are FPs, which commonly leads to unnecessary follow-up and invasive examinations. Chest radiography is inferior to chest CT in terms of sensitivity but superior in terms of specificity. Taking these characteristics into consideration, the development of a computer-aided diagnosis (CAD) model for chest radiograph would have value by improving sensitivity while maintaining low FP results [4].

Many computer-aided detection (CAD) systems have been extensively studied for lung cancer detection and classification [5, 6]. Compared to trained radiologists, CAD systems provide better lung nodules and cancer detection performance in medical images. Generally, the CAD-based lung cancer detection system includes four steps: image processing, extraction of the region of interest (ROI), feature selection, and classification. Among these steps, feature selection and classification play the most critical roles in improving the accuracy and sensitivity of the CAD system, which relies on image processing to capture reliable features. However, benign, and malignant nodule classification is a challenge. Therefore, a rapid, cost-effective, and highly sensitive deep learning-based CAD system for lung cancer prediction is urgently needed.

The development of malignant cells in the lungs is known as lung cancer. Overall men and women's mortality rates have increased as a result of growing cancer incidence. Lung cancer is a disease wherein the cells in the lungs quickly multiply. Lung cancer cannot be eradicated, but it can be reduced [7]. The number of people affected with lung cancer is precisely equal to the number of people who smoke continuously. Lung cancer treatment was evaluated using classification approaches such as Naive Bayes, SVM, Decision Tree, and Logistic Regression. Pradhan et al. [8] conduct an empirical evaluation of multiple machine learning methods that can be used to identify lung cancer using IoT devices. A survey of roughly 65 papers employing machine learning techniques to forecast various diseases was conducted in this study. The study focuses on a variety of machine learning methods for detecting a variety of diseases in order to identify a gap in prospective lung cancer detection in medical IoT.

Deep residual learning is used by Bhatia et al. [9] to identify lung cancer from CT scans. With the UNet and ResNet algorithms, we propose a series of pre-processing approaches for emphasising cancer-prone lung regions and retrieving characteristics. The extracted features are fed through several classifiers, namely Adaboost and Random Forest, and the individual predictions are ensembled to calculate the likelihood of a CT scan becoming cancerous.

Without learning inadequate human information, Shin et al. [10, 11] use deep learning to investigate the characteristics of cell exosomes and determine the similarities in human plasma extracellular vesicles. The deep learning classifier was tested with exosome SERS data from regular and lung cancer cell lines and was able to categorise them with 95% efficiency. The deep learning algorithm projected that 90.7% of patients' plasma exosomes were more similar to lung

cancer cell extracellular vesicles than the mean of healthy controls in 43 patients, encompassing stage I and II cancer patients.

III. METHODOLOGY

A deep CNN model for lung cancer classification from CT scan images is a powerful approach that leverages the capabilities of deep learning to automatically learn and extract relevant features from raw image data. Here is an overview of how a deep CNN model can be used for classifying CT scan images into normal and malignant categories:

i. Data Collection and Preprocessing:

- A dataset of CT scan images is collected, where each image is labeled as normal (non-cancerous) or malignant (cancerous).
- Preprocessing steps are applied to the images, including resizing them to a consistent resolution, normalizing pixel values, and enhancing image quality.

ii. Dataset Splitting:

The dataset is divided into two subsets: training, and testing sets. This division allows for training the model, tuning hyperparameters, and evaluating its performance independently.

iii. Architecture Design:

- The deep CNN architecture is designed to learn hierarchical features from the CT scan images. A typical architecture consists of multiple convolutional layers followed by pooling layers and fully connected layers.
- Convolutional layers use learnable filters to detect features such as edges, textures, and shapes at different spatial scales.
- Pooling layers reduce the spatial dimensions of the feature maps while retaining the most important information.
- Fully connected layers combine the extracted features and make the final classification decision.

iv. Model Training:

- The CNN model is trained using the training dataset. During training, the model learns to optimize its internal parameters (weights and biases) to minimize a specified loss function (e.g., binary cross-entropy).
- Backpropagation and gradient descent techniques are used to update the model's parameters iteratively.

v. Hyperparameter Tuning:

Hyperparameters, such as learning rate, batch size, and the number of layers and neurons, are tuned using the validation set to optimize the model's performance.

vi. Regularization Techniques:

Regularization methods like dropout and batch normalization may be incorporated to prevent overfitting and enhance model robustness.

vii. Evaluation:

- The model's performance is evaluated using the testing dataset. Common evaluation metrics include accuracy, precision, recall, F1-score, and the confusion matrix.
- The model's ability to correctly classify lung cancer CT scan images into normal and malignant categories is assessed. The deep CNN models have demonstrated remarkable success in medical image classification tasks, including lung cancer classification. Their ability to automatically learn relevant features from raw image data makes them a valuable tool for improving diagnostic accuracy and efficiency in healthcare.

IV. SYSTEM ANALYSIS

Existing System:

SVM is a supervised machine learning algorithm that can be used for both binary and multiclass classification tasks. It aims to find a hyperplane that best separates different classes of data. In the context of lung cancer classification, SVM can be trained on features extracted from CT scan images to distinguish between cancerous and non-cancerous regions. SVM typically relies on handcrafted features extracted from the images, such as texture, intensity, and shape

descriptors. These features need to be carefully designed and selected, which can be a challenging and time-consuming process. It can utilize the kernel trick to map the data into a higher-dimensional space, making it possible to capture complex relationships between features.

Limitations:

- Limited Feature Extraction: SVM's performance heavily depends on the quality and relevance of handcrafted features. In the case of lung cancer classification from CT scans, designing effective features that capture subtle patterns indicative of cancer can be difficult.
- Data Size and Imbalance: SVM may not perform well when dealing with large datasets, and it can be sensitive to class imbalance. In medical imaging, collecting a balanced dataset with a sufficient number of cancerous and non-cancerous cases can be challenging.
- Scalability: SVMs can become computationally expensive when dealing with high-dimensional Feature spaces, as is often the case with image data. Training on large datasets or using complex kernels can lead to longer training times.
- Lack of Spatial Information: SVM does not inherently capture spatial relationships within the images, which can be crucial for identifying tumors in CT scans. CNNs, on the other hand, are designed to automatically learn hierarchical features, including spatial features, from raw image data.

Proposed System:

Our proposed Lung Cancer Detection System aims to address the limitations of existing systems by leveraging advanced Deep learning techniques, comprehensive datasets, and user-friendly interfaces. It classifies lung cancer images as benign or malignant using deep learning (CNN with VGG16) and machine learning (SVM). The system utilizes a dual-model approach, where the CNN model, based on VGG16, provides highly accurate predictions, while the SVM model serves as an alternative for feature-based classification. The Tkinter-based GUI allows users to upload images and receive instant predictions, making the system user-friendly and accessible.

Advantages:

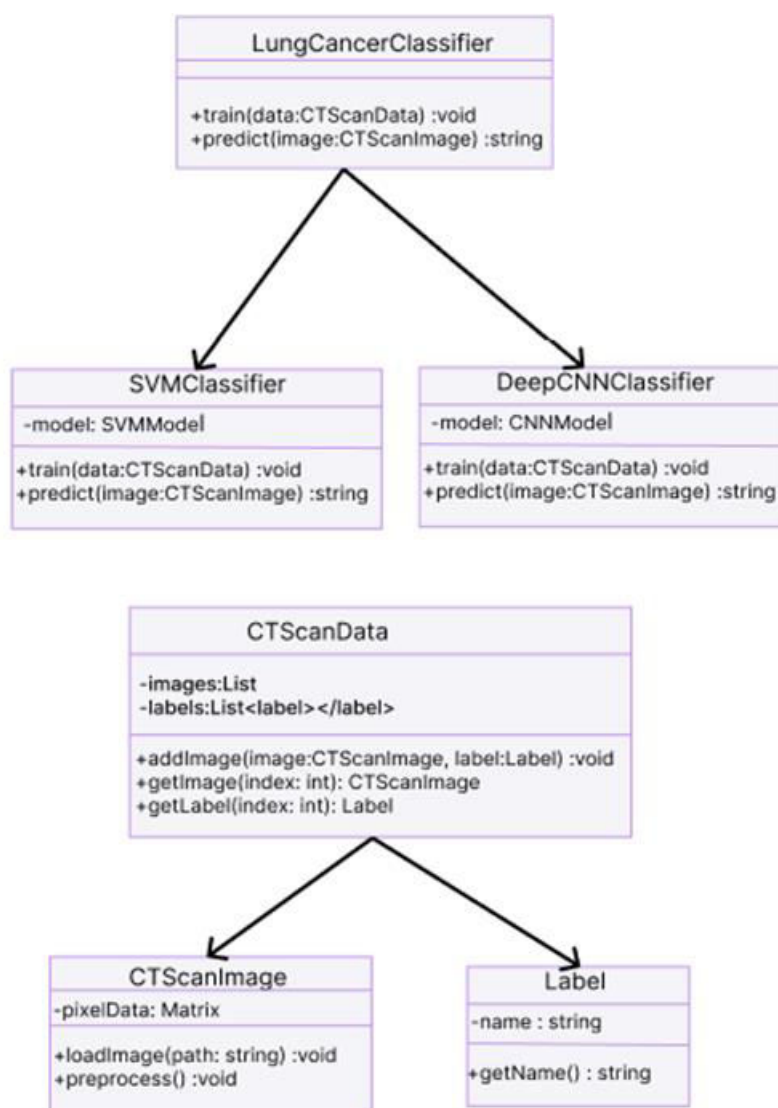
- Improved Accuracy with Hybrid Approach: The system leverages deep learning (CNN with VGG16) for feature extraction and classification. It also uses SVM with PCA to enhance classification, providing an alternative model for robustness. The hybrid approach helps mitigate the limitations of a single model, leading to better accuracy.
- Faster and More Efficient Processing: CNN (VGG16) efficiently extracts spatial features from images, reducing manual feature engineering. PCA in SVM reduces dimensionality, improving computational efficiency and model performance.
- Automated and User-Friendly GUI: The Tkinter-based GUI allows users to upload an image and receive predictions instantly. The GUI is intuitive, visually appealing, and easy to use for both medical professionals and non-experts.
- Model Generalization and Scalability: The pretrained VGG16 model ensures good generalization on unseen data. The system can be expanded and fine-tuned to support more lung cancer subtypes.

V. IMPLEMENTATION**A. Execution Flow:**

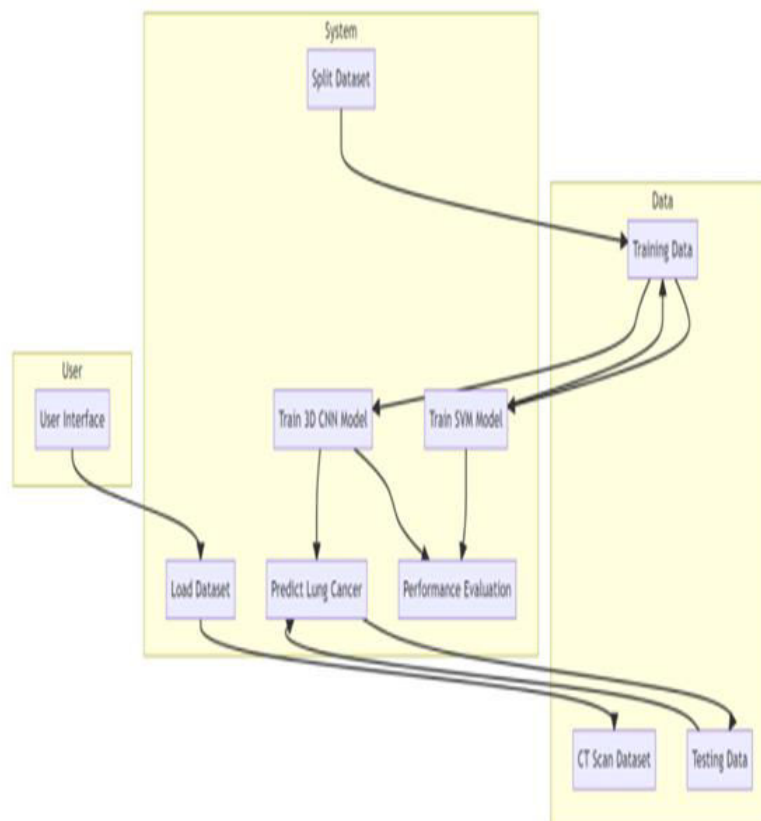
- Start Execution: The execution of the system begins.
- Home Page: Users land on the home interface of the application, where they are greeted with a title and an option to upload CT scan images for lung cancer prediction.
- Image Upload: Users click on the "Upload Image" button to select a CT scan image from their device for diagnosis.

- Preprocessing: The uploaded image is resized, normalized, and reshaped to match the input format required by the models.
- Model Prediction:
- The preprocessed image is passed through the CNN model (VGG16-based) to classify it as Benign or Malignant.
- Result Display: The GUI displays CNN results on the screen. The uploaded image is also shown for reference along with the prediction label.
- User Review: Users analyze the result and take necessary action or consult medical professionals based on the prediction outcome.
- End Execution: Once the user closes the application or completes their prediction task, the system concludes the execution.

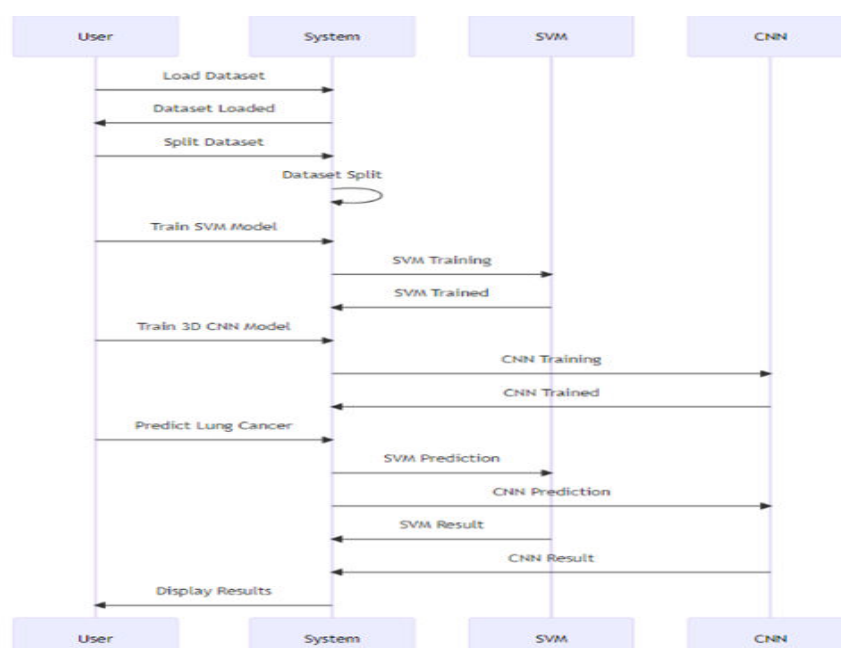
B. Activity Diagram:



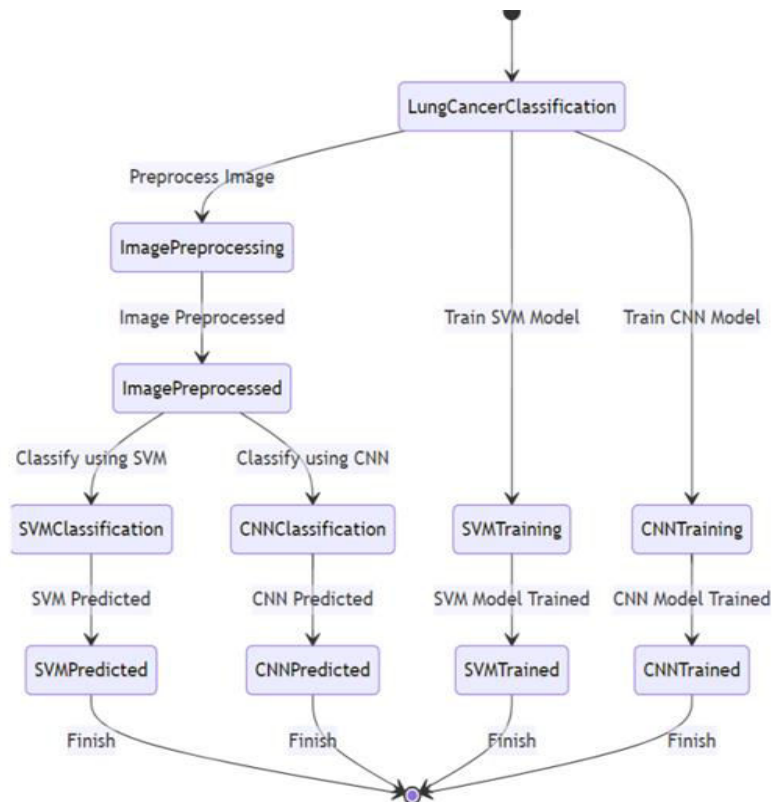
C.Data Flow Diagram:



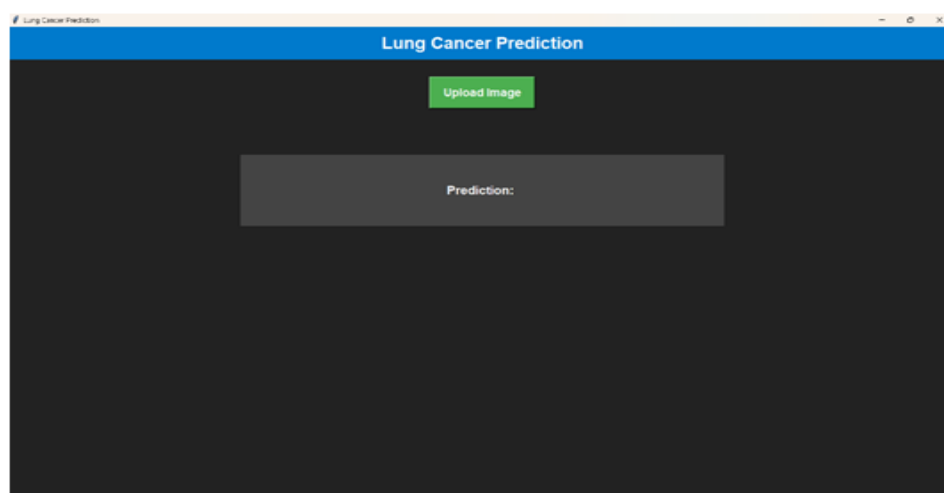
D. Sequence Diagram:

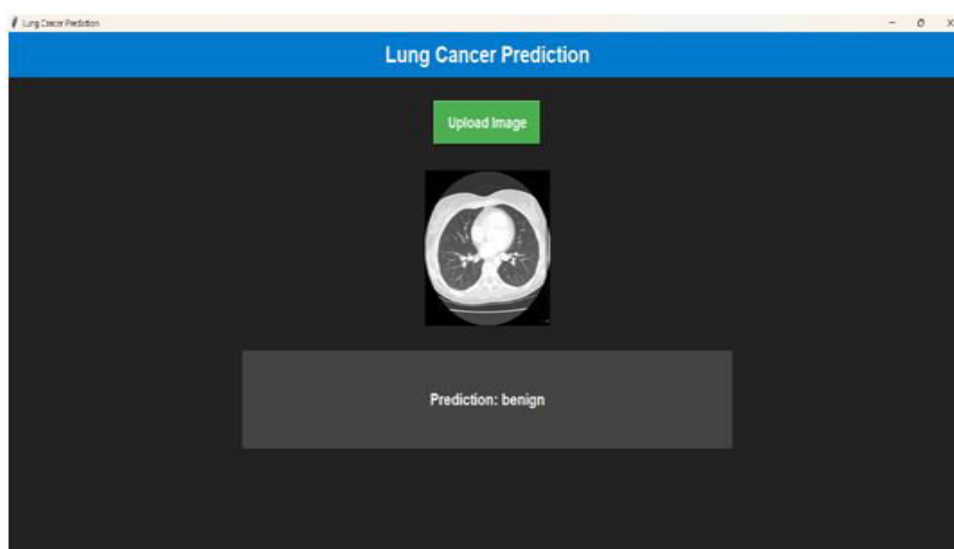
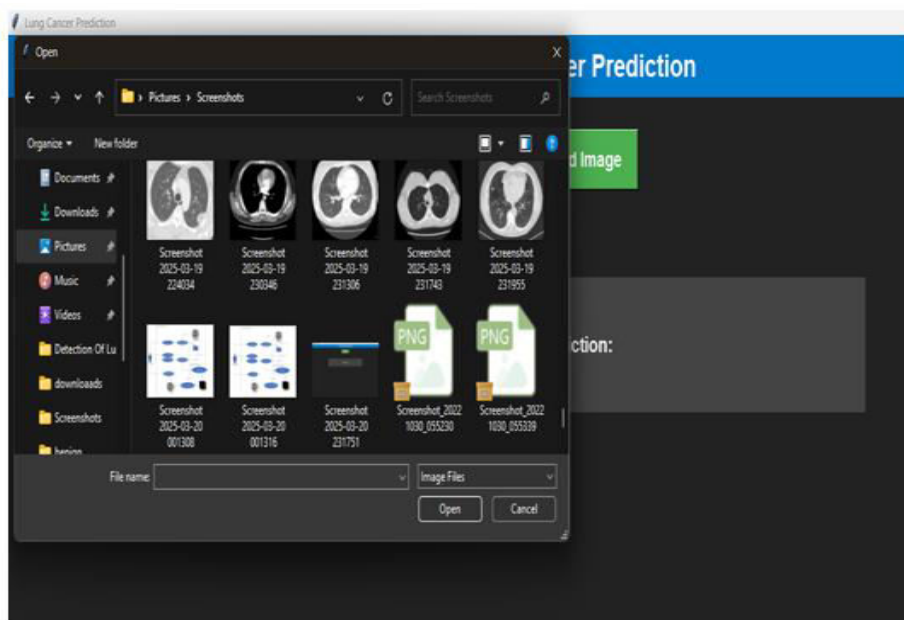


E. Activity Diagram:



VI.RESULTS





VII. CONCLUSION

This work implements an advanced CNN modifications for the detection of lung cancer from CT scan images. The higher accuracy of the 3D CNN can be attributed to its ability to automatically learn relevant features from the images through multiple convolutional layers, which is especially advantageous for complex and high-dimensional data like CT scans. On the other hand, SVM relies on handcrafted features and may struggle with capturing complex patterns in the data. Based on the obtained results, it is clear that the proposed 3D CNN model significantly outperformed the SVM classifier in the task of detecting lung cancer from CT images. In conclusion, the promising results obtained by the 3D CNN model in lung cancer detection highlight the potential of deep learning in medical image analysis. Future endeavors should focus on refining model performance, scalability, and seamless integration into clinical workflows, with the ultimate goal of positively impacting patient care and outcomes across various medical domains.

REFERENCES

- [1] Y. Xu, A. Hosny, R. Zeleznik, C. Parmar, T. Coroller, I. Franco, H.J. Aerts, Deep learning predicts lung cancer treatment response from serial medical imaging, *Clin. Cancer Res.* 25 (11) (2019) 3266–3275.
- [2] M.I. Faisal, S. Bashir, Z.S. Khan, F.H. Khan, An evaluation of machine learning classifiers and ensembles for early stage prediction of lung cancer, December, in: 2018 3rd International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST), IEEE, 2018, pp. 1–4.
- [3] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyo, A. Tsirigos, Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning, *Nat. Med.* 24 (10) (2018) 1559–1567.
- [4] D.M. Ibrahim, N.M. Elshennawy, A.M. Sarhan, Deep-chest: multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases, *Comput. Biol. Med.* 132 (2021), 104348. [5] Avanzo, J. Stancanella, G. Pirrone, G. Sartor, Radiomics and deep learning in lung cancer, *Strahlenther. Onkol.* 196 (10) (2020) 879–887.
- [6] S.H. Hyun, M.S. Ahn, Y.W. Koh, S.J. Lee, A machine-learning approach using PET- based radiomics to predict the histological subtypes of lung cancer, *Clin. Nucl. Med.* 44 (12) (2019) 956–960.
- [7] P. R. Radhika, R.A. Nair, G. Veena, A comparative study of lung cancer detection using machine learning algorithms, 2019, February, in: IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), IEEE, 2019, pp. 1–4.
- [8] K. Pradhan, P. Chawla, Medical Internet of things using machine learning algorithms for lung cancer detection, *J. Manag. Anal.* 7 (4) (2020) 591–623.
- [9] S. Bhatia, Y. Sinha, L. Goel, Lung cancer detection: a deep learning approach, in: *Soft Computing for Problem Solving*, Springer, Singapore, 2019, pp. 699–705.
- [10] H. Shin, S. Oh, S. Hong, M. Kang, D. Kang, Y.G. Ji, Y. Choi, Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes, *ACS Nano* 14 (5) (2020) 5435–5444.
- [11] V. Rajasekar, B. Predic, M. Saracevic, M. Elhoseny, D. Karabasevic, D. Stanujkic, P. Jayapaul, Enhanced multimodal biometric recognition approach for smart cities based on an optimized fuzzy genetic algorithm, *Sci. Rep.* 12 (1) (2022) 1–11.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details