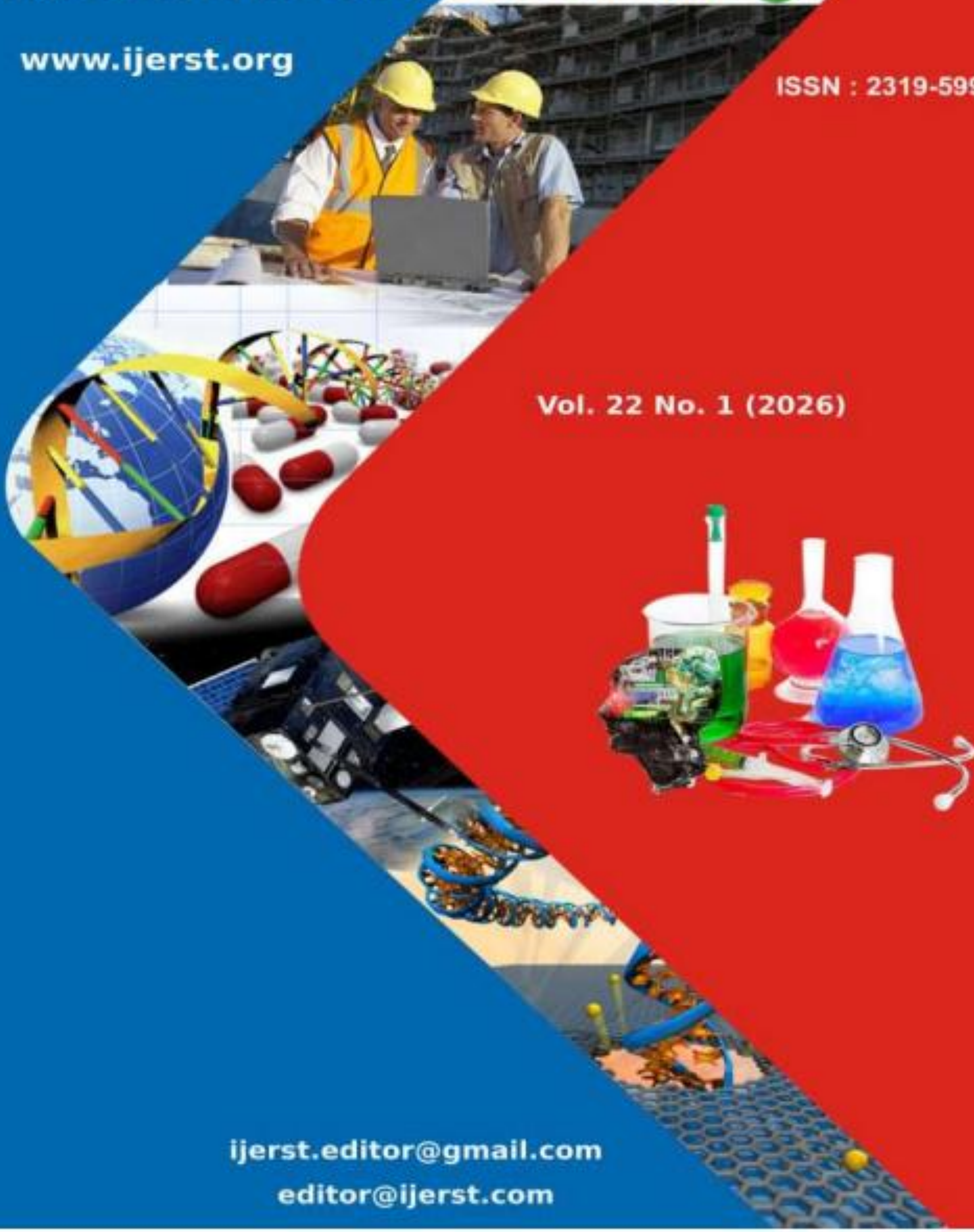
IJERST

International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 1 (2026)



ijerst.editor@gmail.com
editor@ijerst.com

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 1, 2026

Peer Reviewed, Referred & Indexed Journal

Research

DEEP LEARNING-BASED LUNG CANCER DETECTION WITH VISION TRANSFORMER-BASE

Aala Ravi Kiran[1], Bapatla Sai Kamal[2], Gurram Pavan[3], Gandupilli Panvi Tejesh[4]
[#1] *Assistant Professor in Department of Information Technology and NRI Institute of Technology, Agiripalli*
[#2#3#4] B.Tech with Specialization of Information Technology in NRI Institute Of Technology,Agiripalli

Abstract: In order to lower death rates and increase patient survival, early identification of lung cancer is essential. In this study, a Vision Transformer (ViT) model for CT scan image analysis is used to propose an automated lung cancer detection system. The suggested method classifies lung CT scans into several categories using deep learning techniques, allowing for precise and effective diagnosis help.

Four classes of labeled CT scan data were used to train and assess the model. To improve model performance and generalization, image preprocessing methods like scaling, normalization, and augmentation were used. Compared to conventional convolution-based methods, the Vision Transformer design improves classification accuracy by capturing global contextual information from medical images.A web application built on Streamlit is used to deploy the system, enabling users to register, upload CT scan images, and receive real-time forecasts and confidence scores. Metrics like accuracy, precision, recall, F1-score, and confusion matrix analysis were used to evaluate performance. The experimental findings show that the suggested ViT-based model performs reliably in predictions and achieves good classification accuracy.

This research offers a user-friendly AI-powered platform to provide clinical decision support and early lung cancer diagnosis, while also demonstrating the efficacy of transformer-based architectures in medical image processing.

Received: 22-01-2026 Accepted: 26-02-2026 Published: 06-03-2026

1. INTRODUCTION

Due to delayed diagnosis and a shortage of skilled radiologists, lung cancer is still one of the most deadly illnesses in the world and a major cause of cancer-related deaths. Patient survival rates are greatly increased by early detection of lung cancer; yet, conventional diagnostic techniques mostly rely on the subjective, time-consuming, and human error-prone manual analysis of CT scan data. Even seasoned medical professionals frequently find it difficult to make an accurate diagnosis due to subtle changes in tumor shape, size, and texture.

Deep learning methods have grown in significance in medical image analysis due to the quick development of artificial intelligence. VGG16 is one example of a Convolutional Neural Network (CNN) that has shown impressive performance in the identification of lung cancer by automatically extracting spatial information from CT scan pictures. The inability of CNN-based models to capture global contextual information and long-range relationships in medical pictures, notwithstanding their success, can compromise classification accuracy, particularly in complicated instances.

This study suggests a deep learning method for lung cancer categorization based on Vision Transformers (ViT) in order to get over these restrictions. Through the use of self-attention mechanisms and the division of images into fixed-size patches, Vision Transformers allow the model to learn both local and global links within the image. ViT is especially well-suited for medical imaging jobs because tumor patterns may be dispersed throughout several lung areas due to its capacity to capture global context.

This project's goal is to create a dependable and automated method for classifying lung cancer by employing the Vision Transformer algorithm to divide lung CT scan images into several groups, including normal tissue, squamous cell carcinoma, large cell carcinoma, and adenocarcinoma. The suggested solution seeks to lessen reliance on manual interpretation while increasing diagnostic accuracy, robustness, and consistency by utilizing transformer-based architecture. This method can help radiologists make better clinical decisions, aid in early detection, and enhance patient outcomes..

2. LITERATURE SURVEY

A. Asuntha and A. Srinivasan (2020) proposed a deep learning-based automated classification system for lung cancer using CT scan images. Their approach utilized Convolutional Neural Networks (CNNs) to automatically extract discriminative features from lung nodules and classify them effectively. The study demonstrated that CNN-based models significantly outperformed traditional machine learning techniques in terms of accuracy, robustness, and reliability, highlighting the potential of deep learning in medical image analysis.

B. Jacobs, A. A. A. Setio, E. Scholten, and B. van Ginneken (2021) evaluated deep learning algorithms for lung cancer detection on large-scale screening CT datasets. The research incorporated an observer study involving radiologists to compare model performance with human expertise. The findings indicated that deep learning systems enhanced early lung cancer detection and reduced false positives, thereby improving clinical decision-making and screening efficiency.

1. **H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang (2019)** developed an automated pulmonary nodule detection framework based on deep convolutional neural networks. Their system focused on accurate feature learning directly from CT images without relying on handcrafted features. The results showed improved diagnostic accuracy and effective identification of lung nodules, reinforcing the capability of CNNs in medical imaging tasks.

A. Halder, S. Chatterjee, D. Dey, and S. Munshi (2020) introduced a hybrid method that combined adaptive morphology-based segmentation with deep learning techniques for lung nodule detection. Their approach improved segmentation precision and minimized false positive rates in CT scan analysis. The integration of image processing and deep learning contributed to more accurate and reliable lung cancer diagnosis.

2. **R. Manickavasagam, S. Selvan, and M. Selvan (2022)** presented a computer-aided diagnosis (CAD) system utilizing CNN architectures for detecting and classifying malignant lung nodules. The framework demonstrated strong performance in identifying early-stage lung cancer, thereby supporting radiologists in clinical assessment and improving diagnostic outcomes.

A. Dosovitskiy et al. (2021) introduced the Vision Transformer (ViT) architecture, a transformer-based model that leverages self-attention mechanisms to capture global image dependencies. Unlike traditional CNNs that focus on local feature extraction, ViT models global contextual relationships within images. The study showed that transformer-based architectures could outperform CNNs in image classification tasks, making them highly suitable for advanced medical imaging applications such as lung cancer classification

.3. METHODOLOGY

a) Proposed Work:

Using CT scan pictures, the suggested method presents a deep learning architecture based on Vision Transformer (ViT) for precise and automated lung cancer categorization. The suggested method makes use of self-attention mechanisms to capture both local and global contextual information available across the entire image, in contrast to conventional CNN-based models that mainly concentrate on local feature extraction using convolution operations. This makes it possible for the system to diagnose lung cancer with greater precision, resilience, and dependability.

To guarantee high-quality input data for model training, the suggested methodology starts with a thorough picture pretreatment step. Data augmentation, noise reduction, image scaling, and normalization are examples of preprocessing procedures. These actions lessen the effects of dataset imbalance and assist in managing differences in CT scan quality. Rotation, flipping, scaling, and other data augmentation techniques are used to diversify the dataset, which enhances the model's capacity for generalization and lowers overfitting.

Each lung CT scan image is separated into fixed-size image patches in the Vision Transformer architecture. These patches are subsequently flattened and transformed into embedding vectors. In order to maintain spatial information, positional embeddings are inserted. Several transformer encoder layers using multi-head self-attention mechanisms process these embeddings. In order to detect intricate tumor patterns dispersed throughout the image, the model must be able to understand long-range dependencies and interactions between various lung regions.

In order to classify CT scan pictures into four categories—adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and normal tissue—the encoded features are further processed through fully connected layers for multi-class classification. Class probabilities are produced by the output layer using the Softmax activation function. To guarantee quicker convergence and avoid overfitting, sophisticated optimization strategies such the Adam optimizer, learning rate scheduling, dropout regularization, and early stopping are used.

A Streamlit-based web application incorporates the trained Vision Transformer model, enabling users to upload CT scan pictures and obtain real-time predictions and confidence scores. The outcomes of predictions are kept in a database for later use and examination. All things considered, the suggested approach provides an effective, scalable, and clinically applicable method for automated lung cancer identification and categorization..

b) System Architecture:

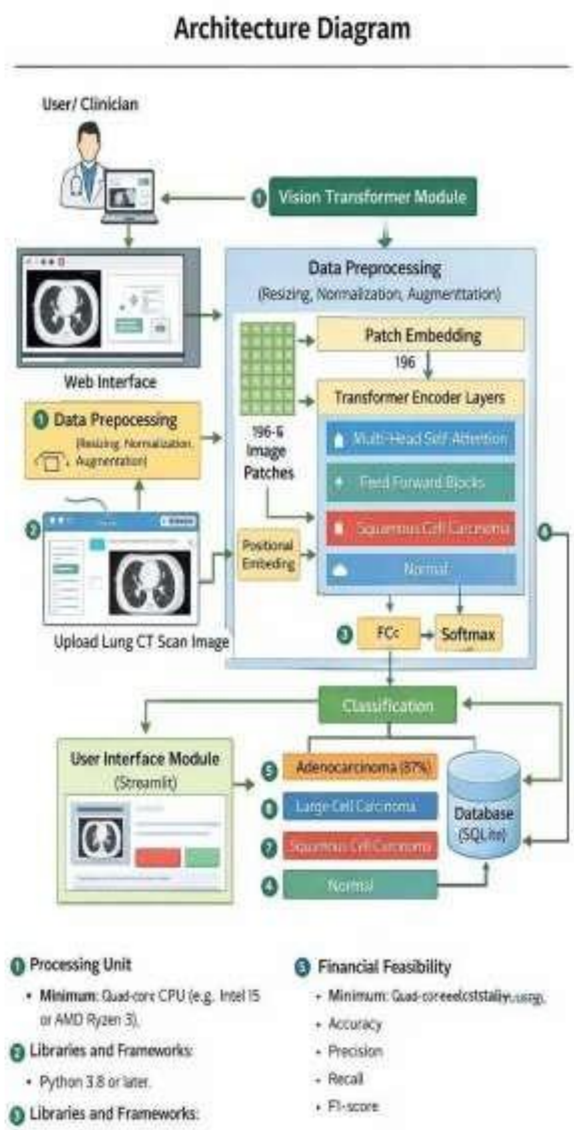
Fig 1 Proposed Architecture

he proposed architecture for Lung Cancer Classification using the Vision Transformer (ViT-Base) model represents an end-to-end automated system that assists clinicians in diagnosing lung cancer from CT scan images. The process begins when a user uploads a lung CT image through a web-based interface (Streamlit), after which the image undergoes preprocessing steps including resizing, normalization, and augmentation to enhance model performance. The preprocessed image is divided into fixed-size patches, which are converted into vector representations through patch embedding and enriched with positional encoding to preserve spatial information. These embeddings are passed through multiple Transformer encoder layers consisting of multi-head self-attention and feed-forward networks, enabling the model to capture global contextual relationships within the image. The extracted deep features are then forwarded to fully connected layers followed by a Softmax activation function to classify the image into one of the categories: Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, or Normal. The final prediction along with confidence score is displayed to the user and stored in an SQLite database for record management. Overall, this architecture provides an efficient, accurate, and clinically supportive deep learning framework for automated lung cancer classification.

c) MODULES:

1. Data Collection and Integration

The Data Collection and Integration phase focuses on gathering and organizing lung CT scan images required for training and testing the lung cancer classification system. In this stage, lung CT images are collected from standard public medical imaging datasets to ensure reliability and diversity. The images are carefully labeled into appropriate categories such as cancerous and non-cancerous (or multi-class cancer types) to maintain dataset



integrity. The collected data is then structured into organized folders for training, validation, and testing to support systematic model development. This phase primarily utilizes public medical datasets, file system management techniques, and Python libraries for handling and organizing image data efficiently.

2. Data Preprocessing

The Data Preprocessing stage prepares the collected lung CT scan images in a suitable format for training the Vision Transformer (ViT) model. All images are resized to a fixed resolution (e.g., 224 × 224) to ensure uniform input dimensions. Pixel intensity values are normalized to improve convergence during training, and noise reduction or image enhancement techniques are applied to improve image quality. Data augmentation methods such as rotation and flipping are used to increase dataset diversity and reduce overfitting. Finally, the dataset is split into training, validation, and test sets. This stage employs Python, NumPy, OpenCV, PIL, and torchvision transforms for efficient preprocessing operations.

3. Model Development

The Model Development phase involves designing and implementing the Vision Transformer–based deep learning architecture for lung cancer classification. In this stage, the ViT model is defined with patch embedding layers that divide images into fixed-size patches and convert them into vector representations. Transformer encoder blocks incorporating multi-head self-attention mechanisms are implemented to capture global contextual relationships within the CT images. Fully connected layers are added at the final stage to perform multi-class classification. Additionally, suitable loss functions and optimization strategies are configured to enhance learning performance. Technologies such as PyTorch, Vision Transformer architecture, and optionally CUDA for GPU acceleration are used in this phase.

4. Training Pipeline

The Training Pipeline automates and manages the complete model training process. Preprocessed lung CT scan images are loaded into the training framework, and the Vision Transformer model is trained using predefined hyperparameters such as learning rate and batch size. Optimization techniques, including the Adam optimizer and learning rate scheduling, are applied to improve convergence. The system continuously monitors training and validation performance to detect overfitting or underfitting. Model checkpoints are saved periodically to preserve the best-performing model for future use. This stage is implemented using PyTorch, torch optimizers, and GPU acceleration where available.

5. Model Evaluation

The Model Evaluation phase assesses the performance, accuracy, and reliability of the trained lung cancer classification model. Evaluation metrics such as accuracy, precision, recall, and F1-score are computed to measure classification effectiveness. Confidence scores for predictions are analyzed to understand the certainty of model outputs. The trained model is validated using unseen test data to ensure generalization capability. Tools such as Scikit-Learn and Matplotlib are used for computing metrics and visualizing performance results.

6. Prediction and Result Management

The Prediction and Result Management module provides real-time lung cancer prediction and manages diagnostic results efficiently. The system accepts CT scan images through the application interface and processes them using the trained Vision Transformer model. It generates predictions along with confidence scores and displays the results to the user in an understandable format. Additionally, prediction results along with timestamps are stored in a database for future reference and medical record management, ensuring traceability and systematic result storage

4. EXPERIMENTAL RESULTS

.**Accuracy:** The capacity of a test to accurately distinguish between healthy and sick cases is a measure of its accuracy. We can determine a test's accuracy by calculating the proportion of reviewed cases with true positives and true negatives. In mathematical terms, this looks like:

Accuracy = TP + TN TP + TN + FP + FN.

$$Accuracy = \frac{(TN + TP)}{T}$$

F1-Score: One way to evaluate a model's performance in machine learning is via its F1 score. It takes a model's recall and precision and mixes them. Using the whole dataset, the accuracy metric determines the frequency with which a model produced accurate predictions.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

Precision: The proportion of occurrences or samples that are correctly classified out of all the ones that are classified as positive is known as precision. Consequently, the following is the formula for determining the accuracy:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: The capacity of a model to detect all significant instances of a given class is measured by recall, a metric in machine learning. It sheds light on how well a model captures instances of a certain class and is

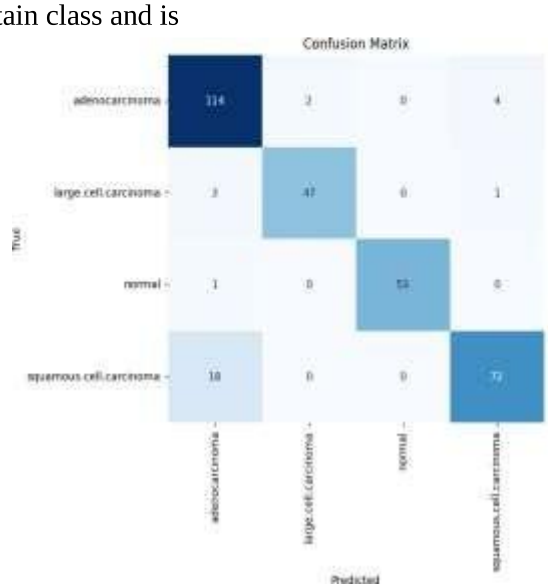


Fig 1:Confusion Matrix

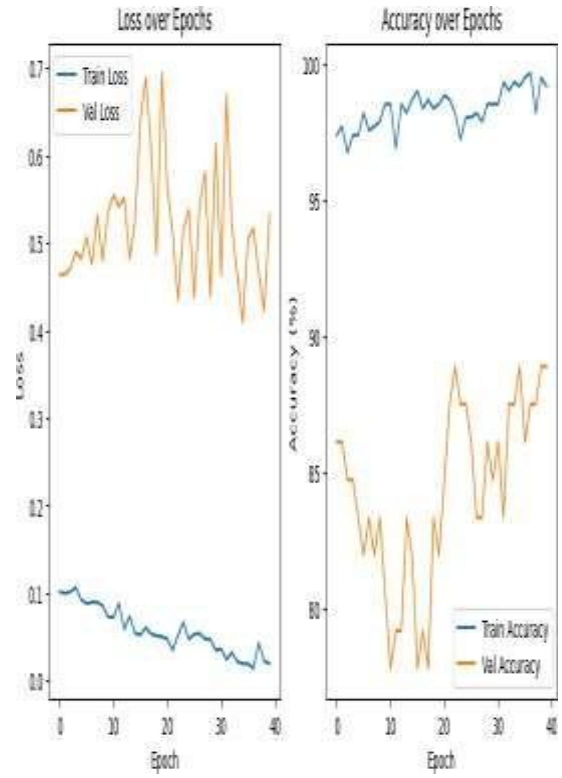


Fig 2:Training Graphs

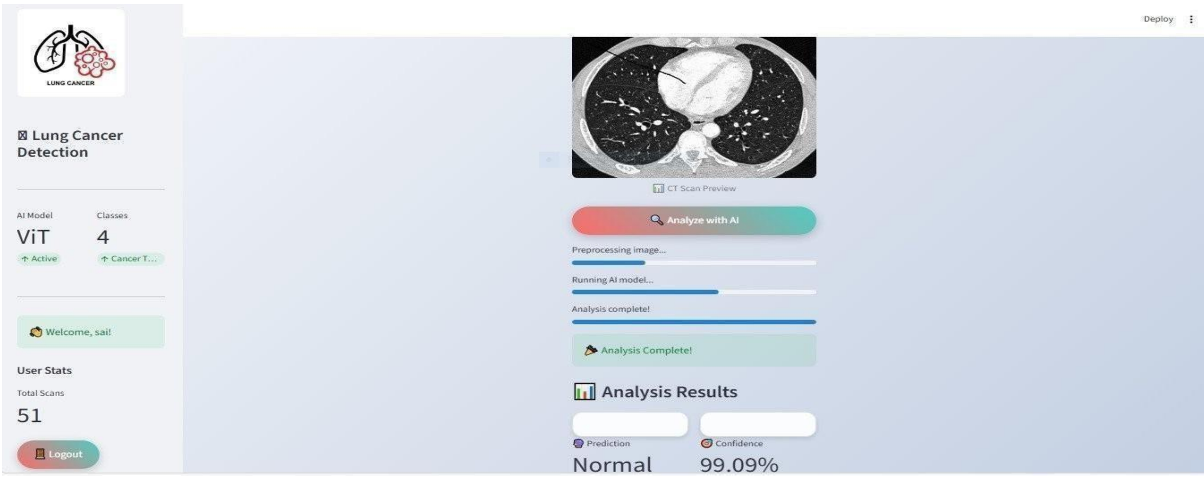


Fig 3: After uploading image click on Analyze with AI.

Fig 4:By clicking on predict button we can get know whether patient have lung cancer or not.

5. CONCLUSION

To sum up, the Vision Transformer–based Lung Cancer Classification System shows how cutting-edge deep learning methods may be applied successfully to medical picture analysis. The suggested approach uses a transformer-based architecture that captures global contextual linkages within the image to accurately classify lung CT scan images into many categories: Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, and Normal.

In contrast to conventional Convolutional Neural Networks (CNNs), the Vision Transformer uses self-attention techniques to learn long-range dependencies while processing images as patch sequences. This enhances the model's capacity for generalization and classification by allowing it to extract richer feature representations from CT scan pictures.

A comprehensive end-to-end pipeline is formed by the system's integration of several elements, such as picture preprocessing, Vision Transformer inference, a Streamlit-based user interface, and SQLite database storage. Compatibility with the trained model is ensured by preprocessing procedures including tensor conversion, normalization, and scaling. Streamlit's implementation makes the system interactive and user-friendly by enabling users to contribute photographs and receive real-time forecasts along with confidence scores.

According to experimental evaluation, the model performs consistently and with high classification accuracy across several lung cancer classifications. By giving confidence levels for every prediction, the use of softmax probability scoring improves interpretability. PyTorch's GPU support also makes efficient inference possible, cutting down on processing time and enhancing system responsiveness.

In addition to showcasing excellent technical performance, the built system also exhibits the potential of transformer-based designs in healthcare applications. The method can eliminate human error, speed up clinical decision-making, and help medical professionals with preliminary diagnosis by automating the classification of lung cancer.

All things considered, this research effectively demonstrates the application of Vision Transformer architecture to medical image categorization tasks. The system offers an effective, adaptable, and scalable framework that may be expanded for more extensive healthcare uses. The suggested method has a great chance of being incorporated into actual diagnostic procedures with further advancements and bigger datasets..

6. FUTURE SCOPE

Future developments of the suggested system could concentrate on enhancing clinical applicability, performance, and dependability. Incorporating imaging data with multi-modal information, such as laboratory results, biomarkers, and patient medical history, can greatly improve model accuracy and offer a more thorough diagnostic evaluation. Time-sensitive conditions like Lung cancerrequire prompt clinical decision-making, which would be made possible by implementing real-time stroke detection during MRI or CT scanning processes. By enhancing model interpretability and assisting clinicians in comprehending the rationale behind predictions, Explainable AI (XAI) strategies can boost adoption and trust in medical practice. Additionally, increasing the training dataset's diversity to encompass a wider range of demographics, imaging conditions, and medical variances would enhance the model's robustness and generalization, guaranteeing reliable performance in actual healthcare settings..

REFERENCES

[1] C. Jacobs, A. A. A. Setio, E. T. Scholten, P. K. Gerke, H. Bhattacharya, F. A. M. Hoesein, M. Brink, E. Ranschaert, P. A. de Jong, M. Silva, B. Geurts, K. Chung, S. Schalekamp, J. Meersschaert, A. Devaraj, P. F. Pinsky, S. C. Lam, B. van Ginneken, and K. Farahani, "Deep learning for lung cancer detection on screening CT scans: Results of a large-scale public competition and an observer study with 11 radiologists," Radiology: Artificial Intelligence, vol. 3, no. 6, Art. no. e210027, Nov. 2021.

[2] P. Chaturvedi, A. Jhamb, M. Vanani, and V. Nemade, "Prediction and classification of lung cancer using machine learning techniques," IOP Conference Series: Materials Science and Engineering, vol. 1099, no. 1, Art. no. 012059, Mar. 2021.

[3] M. Šarić, M. Russo, M. Stella, and M. Sikora, "CNN-based method for lung cancer detection in whole slide histopathology images," in Proc. 4th Int. Conf. Smart Sustain. Technol. (SpliTech), Jun. 2019, pp. 1–4.

[4] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in CT images using deep convolutional neural networks," Pattern Recognition, vol. 85, pp. 109–119, Jan. 2019.

[5] C. C. Nguyen, G. S. Tran, V. T. Nguyen, J.-C. Burie, and T. P. Nghiem, "Pulmonary nodule detection based on Faster R-CNN with adaptive anchor box," IEEE Access, vol. 9, pp. 154740–154751, 2021.

[6] Y. Li, X. Wu, P. Yang, G. Jiang, and Y. Luo, "Machine learning for lung cancer diagnosis, treatment, and prognosis," Genomics, Proteomics & Bioinformatics, vol. 20, no. 5, pp. 850–866, Oct. 2022.

[7] Ganji, M. (2025). Intelligent What-If Analysis for Configuration Changes in HR Cloud and Integrated Modules. International Journal of All Research Education and Scientific Methods, 13(04), 4828–4835. https://doi.org/10.56025/ijaresm.2025.1304254828.

[8] S. Wang, D. M. Yang, R. Rong, X. Zhan, J. Fujimoto, H. Liu, J. Minna, I. I. Wistuba, Y. Xie, and G. Xiao, "Artificial intelligence in lung cancer pathology image analysis," Cancers, vol. 11, no. 11, p. 1673, Oct. 2019.

[9] N. Banerjee and S. Das, "Prediction of lung cancer in machine learning perspective," in Proc. Int. Conf. Computer Science, Engineering and Applications (ICCSEA), Mar. 2020, pp. 1–5.

[10] Erukude, S. T. (2025, September). FedOnco-Bench: A Reproducible Benchmark for Privacy-Aware Federated Tumor Segmentation with Synthetic CT Data. In 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA) (pp. 870-876). IEEE.

[11] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification," Multimedia Tools and Applications, vol. 79, nos. 11–12, pp. 7731–7762, Mar. 2020.

[12] S. Garud and S. Dhage, "Lung cancer detection using CT images and CNN algorithm," in Proc. Int. Conf. Advanced Computing, Communication and Control (ICAC3), Dec. 2021, pp. 1–6.

[13] A. Halder, S. Chatterjee, D. Dey, S. Kole, and S. Munshi, "An adaptive morphology-based segmentation technique for lung nodule detection in thoracic CT image," Computer Methods and Programs in Biomedicine, vol. 197, Art. no. 105720, Dec. 2020.

[14] R. Y. Bhalerao, H. P. Jani, R. K. Gaitonde, and V. Raut, "A novel approach for detection of lung cancer using digital image processing and convolution neural networks," in Proc. 5th Int. Conf. Advanced Computing and Communication Systems (ICACCS), Mar. 2019, pp. 577–583.

[15] G. Perez and P. Arbelaez, "Automated lung cancer diagnosis using three-dimensional convolutional neural networks," Medical & Biological Engineering & Computing, vol. 58, no. 8, pp. 1803–1815, Aug. 2020.

[16] Sai Maneesh Kumar Prodduturi. (2025). EFFICIENT DEBUGGING METHODS AND TOOLS FOR IOS APPLICATIONS USING XCODE. International Journal of Data Science and IoT Management System, 4(4), 1–6. https://doi.org/10.64751/ijdim.2025.v4.n4.pp1-6.

[17] J. Xu, H. Ren, S. Cai, and X. Zhang, "An improved Faster R-CNN algorithm for assisted detection of lung nodules," Computers in Biology and Medicine, vol. 153, Art. o. 106470, Feb. 2023.

[18] R. Manickavasagam, S. Selvan, and M. Selvan, "CAD system for lung nodule detection using deep learning with CNN," Medical & Biological Engineering & Computing, vol. 60, no. 1, pp. 221–228, Jan. 2022.

[19] A. Masood, B. Sheng, P. Li, X. Hou, X. Wei, J. Qin, and D. Feng, "Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images," Journal of Biomedical Informatics, vol. 79, pp. 117–128, Mar. 2018.

[20] W. J. Sori, J. Feng, A. W. Godana, S. Liu, and D. J. Gelmecha, "DFDNet: Lung cancer detection from denoised CT scan image using deep learning," Frontiers of Computer Science, vol. 15, no. 2, pp. 1–13, Apr. 2021.

Author profiles

Ravikiran Aala is presently working as Assistant Professor in the department of Information Technology at NRI Institute of Technology, Vijayawada. He received his M.Tech degree from Jawaharlal Nehru Technological University, Kakinada (JNTUK). He has published over 7 research paper in international journals. He has more than 9 years of experience in teaching

Email- ravikiran.aala@gmail.com

BAPATLA SAI KAMAL is a B. Tech student specializing in Information Technology at NRI Institute of Technology. He has a strong interest in Artificial Intelligence, Machine Learning, and Internet of Things (IoT). He has developed solid technical skills in Python programming, Deep Learning, and Full-Stack Web Development.He completed internships as an AI Intern at Infosys Springboard, where he developed DermalScan: an AI-enabled Facial Skin Aging Detection System using deep learning techniques. He worked on end-to-end AI pipeline development including data preprocessing, model training, and performance evaluation. He also completed an AI internship at Infosys Springboard, strengthening his fundamentals in Machine Learning, Neural Networks, and real-world AI applications.

EMAIL-saikamal1213@gmail.com

Gurram Pavan is a B .Tech student in Information Technology with interests in Artificial Intelligence, Machine Learning, and Data Analytics . He has experience working as a Data Annotator at iMerit Technology Services, contributing to data labeling and quality validation for AI/ML datasets . He also completed a Generative AI internship at Codegnan IT Solutions. He was a team member in the AI Based Lung Cancer Classification project , where deep learning techniques were used for medical image classification. He also has experience in UI/UX design using Figma for mobile applications. Gurram Pavan has completed certifications in Artificial Intelligence, Data Analytics , Data Science, and UI/UX Design.

EMAIL: pa156v@gmail.com

GANDUPILLI PANVI TEJESH is currently pursuing a Bachelor of Technology (B.Tech) in Computer Science and Information Technology at NRI Institute of Technology,He has a strong foundation in Python programming, web development (HTML and CSS), and Artificial Intelligence concepts, with keen interest in Full Stack Development, Software Engineering, and AI-driven applications. He has successfully developed projects such as a Bank Management System, an AI-based Lung Cancer Detection System using Deep Learning, and an AI Stock Market Chatbot, demonstrating his problem-solving abilities and practical implementation skills. As part of his academic growth and self-learning journey, he has actively enhanced his technical knowledge through independent project development and continuous skill improvement in AI and software technologies. He is a self-motivated, quick learner with strong logical thinking and adaptability, aiming to build a professional career as a Full Stack and Software Developer while contributing to innovative and intelligent technology solutions.

EMAIL: panvitejeshgandupilli@gmail.com