IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15487

Lung Cancer Detection Using Convolutional Neural Networks

P. Sampath Kumar[1], Dr. S. Mallikharjuna Rao*[2]

UG Scholar, Department of Electronics and Communication Engineering,

Andhra Loyola Institute of Engineering and Technology, Vijayawada, India[1]

Associate Professor, Department of Electronics and Communication Engineering,

Andhra Loyola Institute of Engineering and Technology, Vijayawada, India*[2]

Corresponding Author* Dr. Mallikharjuna Rao

Abstract: Lung cancer is one of the leading causes of death worldwide, where early and accurate detection plays a critical role in improving patient survival rates. This paper presents an automated lung cancer detection system using Convolutional Neural Networks (CNN), a deep learning technique widely used for medical image analysis. The proposed system classifies lung CT scan images into three categories: Benign, Malignant, and Normal. The system incorporates image preprocessing techniques such as resizing, normalization, and noise reduction to enhance model performance. A CNN model is trained to automatically extract features and perform classification without manual intervention. Furthermore, the trained model is deployed using a Flask-based web application, enabling users to upload images and obtain real-time predictions along with confidence scores. Experimental results demonstrate that the proposed system achieves high classification accuracy and significantly reduces dependency on manual diagnosis. The system is efficient, cost-effective, and accessible, making it suitable for preliminary screening and decision support in healthcare environments. This work highlights the potential of deep learning in improving early-stage lung cancer detection and advancing medical diagnostic systems.

Keywords: Lung Cancer Detection, Convolutional Neural Network (CNN), Deep Learning, Medical Image Processing, Flask, Image Classification, Artificial Intelligence in Healthcare

1. INTRODUCTION

Artificial Intelligence (AI) has significantly transformed the healthcare industry by enabling automated analysis of complex medical data and supporting clinical decision-making. In recent years, machine learning (ML) and deep learning (DL) techniques have been widely applied in medical imaging, particularly for the detection and classification of diseases such as lung cancer. Convolutional Neural Networks (CNNs) have proven to be highly effective in extracting meaningful features from medical images, leading to improved diagnostic accuracy [3], [4].

Lung cancer is one of the most fatal diseases globally, primarily due to late diagnosis and lack of early detection systems. According to the World Health Organization, it accounts for a large percentage of cancer-related deaths each year [2]. Early detection can significantly improve survival rates; however, conventional diagnostic methods rely on manual analysis of CT scans by radiologists, which is time-consuming, costly, and subject to human error.

Several research works have attempted to address this issue using machine learning and deep learning techniques. Alzahrani [1] proposed a predictive model incorporating CTGAN-based synthetic data and tree-based learning methods, achieving high accuracy in lung cancer detection. Other studies have utilized CNN-based approaches for automated feature extraction and classification, demonstrating superior performance compared to traditional methods [5]–[7]. Additionally, techniques such as SMOTE and CTGAN have been introduced to handle class imbalance in datasets, improving model robustness [8], [9].

Despite these advancements, challenges such as limited real-time deployment, lack of user-friendly interfaces, and dependency on structured data still exist. Therefore, there is a need for an efficient, scalable, and accessible system that integrates deep learning with practical deployment.

In this paper, a CNN-based lung cancer detection system is proposed, which utilizes CT scan images for classification into Normal, Benign, and Malignant categories. The system is further integrated with a Flask-based web application to provide real-time predictions, making it accessible for users and healthcare practitioners.

1.1 Motivation of the Work

The primary motivation behind this work is to address the limitations of traditional lung cancer detection methods, which rely heavily on manual interpretation of medical images. Such methods are often time-consuming, prone to human error, and require expert-level knowledge, making them less accessible in resource-constrained environments.

With the increasing availability of medical imaging data and advancements in deep learning, there is an opportunity to develop automated systems that can assist in early detection and diagnosis. Existing approaches, such as predictive modeling using tabular data [1], provide valuable insights but lack the capability to directly analyze image-based features.

Therefore, this work is motivated by the need to:

The proposed system aims to develop an automated and accurate lung cancer detection model by leveraging Convolutional Neural Networks (CNNs) for efficient feature extraction from CT scan images. This approach significantly reduces the dependency on manual diagnosis by medical professionals, minimizing human error and saving time. Additionally, the system is designed to provide real-time predictions through a user-friendly web-based interface, enabling quick and accessible results. Furthermore, it enhances the availability of advanced diagnostic tools in remote and underserved areas, thereby improving early detection and overall healthcare outcomes.

1.2 Objectives of the Proposed System

The main objectives of the proposed system are as follows:

The objective of this project is to design and develop a CNN-based model for lung cancer detection using CT scan images. The system aims to accurately classify lung images into three categories: Normal, Benign, and Malignant. To improve model performance, various preprocessing techniques such as resizing, normalization, and noise reduction are applied. The model focuses on achieving high classification accuracy using advanced deep learning techniques. Furthermore, the trained model is integrated into a Flask-based web application to enable real-time predictions. A user-friendly interface is provided for easy image upload and clear result visualization. This approach helps reduce the time and effort required for manual diagnosis while offering a scalable and cost-effective solution for modern healthcare applications.

A. Motivation of the Research

Water is one of the most important natural resources required for the survival of all living organisms. Increasing industrial activities, urbanization, and population growth have led to severe water pollution. Contaminated water causes serious health problems and environmental damage. Traditional monitoring methods rely on manual sampling, which is time-consuming and cannot provide real-time information. These limitations necessitate a smart and automated monitoring system capable of continuously checking water quality parameters.

With the advancement of sensor technology and microcontrollers, real-time water quality monitoring has become feasible. The motivation of this work is to develop a cost-effective and reliable system that monitors water quality continuously, detects contamination at an early stage, and takes corrective action through automated filtration.

B. Objectives of the Work

The main objective of this project is to design and develop a water quality monitoring system using sensors and a microcontroller. The system measures pH, turbidity, temperature, and TDS in real time; compares values against standard water quality ranges; activates alerts and automated filtration when thresholds are violated; and transmits data to a cloud platform for remote monitoring. The overall goal is to provide a simple, efficient, and cost-effective solution for continuous industrial water quality management.

2. LITERATURE SURVEY

Recent advancements in artificial intelligence and deep learning have significantly improved the accuracy and efficiency of lung cancer detection systems. Various approaches have been proposed in the literature, ranging from traditional machine learning models to advanced deep learning architectures.

Alzahrani [1] introduced a predictive modeling framework that combines Conditional Tabular Generative Adversarial Networks (CTGAN) with tree-based learning algorithms such as Random Forest. This approach addresses the issue of class imbalance by generating synthetic data and achieves high accuracy in lung cancer prediction. However, the model primarily focuses on structured tabular data rather than direct image-based analysis.

Deep learning-based approaches, particularly Convolutional Neural Networks (CNNs), have shown remarkable performance in medical image classification tasks. Tomassini et al. [3] presented a comprehensive survey on CNN-based lung nodule diagnosis using CT images, highlighting the effectiveness of deep learning in feature extraction and

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15487

classification. Similarly, Javed et al. [4] reviewed various deep learning techniques for lung cancer detection and emphasized their superiority over traditional machine learning methods.

Mercaldo et al. [5] proposed an explainable deep learning framework for lung cancer detection using CNNs, incorporating interpretability techniques to enhance model transparency. Ren et al. [6] developed a hybrid deep learning model combining multiple architectures to improve classification performance. Shatnawi et al. [7] demonstrated the application of CNNs for lung cancer diagnosis using CT scan images, achieving high accuracy and reliability.

Handling imbalanced datasets is a critical challenge in medical applications. Chawla et al. [8] introduced the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples to balance datasets

. Similarly, Xu etal. [9] proposed CTGAN for generating realistic synthetic tabular data, improving model generalization. Ensemble learning techniques such as XGBoost [10], Random Forest [11], and Gradient Boosting [12] have also been widely used to enhance prediction performance.

3.METHODOLOGY

Despite these advancements, several limitations remain. Many existing systems lack real-time deployment capabilities and user-friendly interfaces. Additionally, models based on tabular data fail to capture spatial features present in medical images. To overcome these challenges, the proposed system utilizes a CNN-based approach for direct image analysis and integrates it with a Flask-based web application for real-time prediction and accessibility.

A. System Overview

The proposed system is designed to automatically detect lung cancer using Convolutional Neural Networks (CNN) and provide real-time predictions through a web-based interface. The system processes CT scan images, performs preprocessing, extracts features using deep learning, and classifies the images into three categories: Normal, Benign, and Malignant. The overall pipeline includes image acquisition, preprocessing, CNN-based classification, and result visualization using a Flask-based web application.

B. Block Diagram

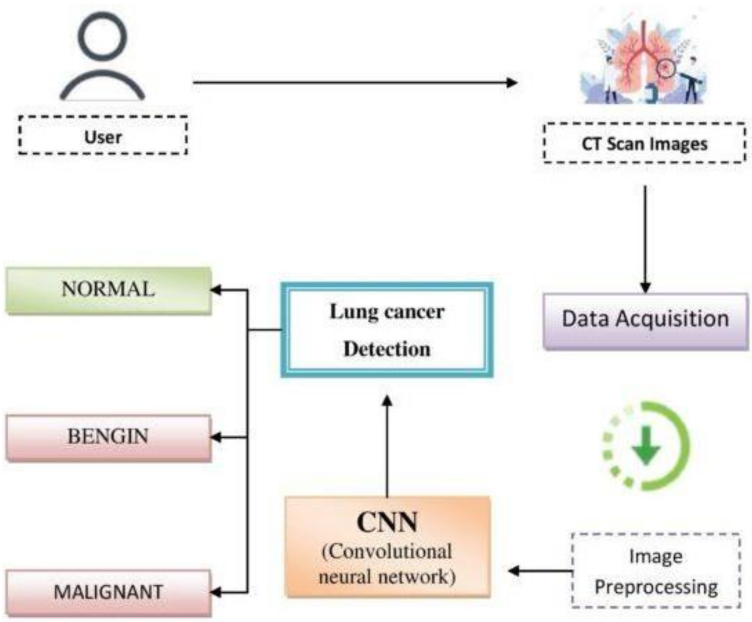


Fig 3.1 Block diagram of the lung cancer detection system using CNN.

The system begins with the user providing CT scan images, which are collected through the interface. The images undergo data acquisition and preprocessing stages, including resizing and normalization to ensure uniformity. The processed images are then fed into a CNN model, which performs feature extraction by identifying patterns and structures within the images. Based on the extracted features, the model classifies the images into three categories: Normal, Begin, and Critical, and produces the final output.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15487

C. Image Preprocessing

Image preprocessing is a crucial step to improve model performance and ensure consistency across input data. In this system, all input images are resized to a fixed dimension of 224 × 224 pixels to maintain uniformity. The pixel values are normalized to a range of 0 to 1 to enhance computational efficiency and stability during training. Noise reduction techniques are applied to minimize unwanted distortions and improve image quality. Finally, the images are converted into numerical arrays, making them suitable for processing by the CNN model. These preprocessing steps collectively enhance feature extraction and reduce computational complexity.

D. CNN Model (Feature Extraction and Classification)

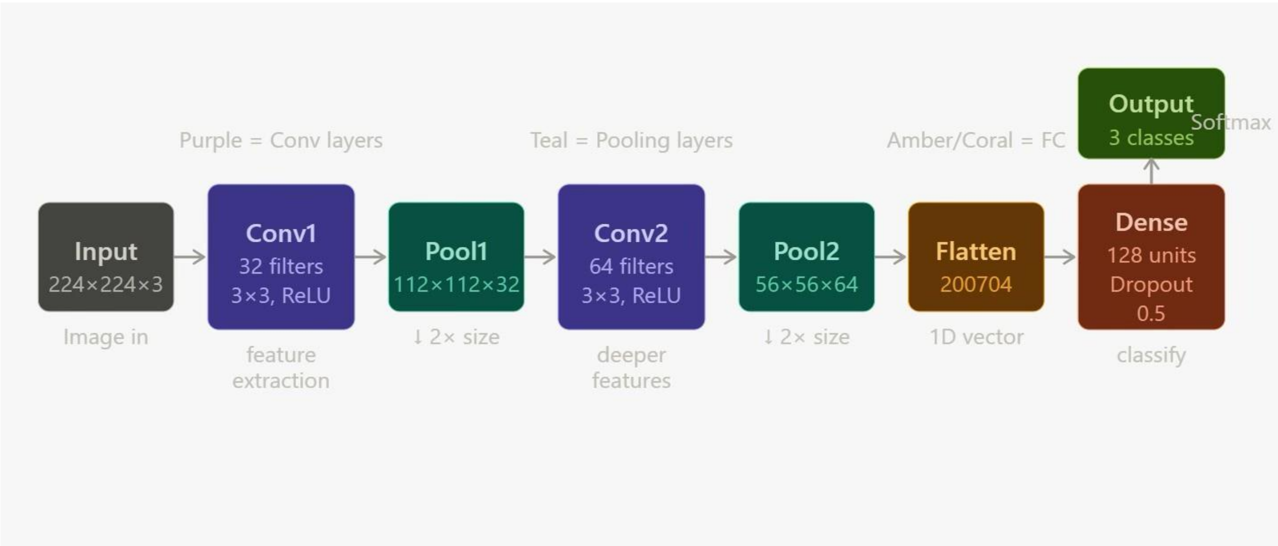


Fig 3.2 CNN layer architecture for lung cancer detection.

The CNN model is designed to automatically learn spatial features from input images without manual intervention. It consists of multiple convolution layers that extract important features such as edges, textures, and patterns from the CT scan images. Pooling layers are used to reduce the spatial dimensions of the feature maps, thereby decreasing computational complexity while retaining essential information. The extracted features are then passed through a flatten layer, which converts the multi-dimensional feature maps into a one-dimensional vector. This vector is fed into fully connected layers that perform classification based on learned features. Finally, a softmax output layer generates probability scores for each class, enabling the system to classify the input image as Normal, Benign, or Malignant.

Table3.1 Primary Functions of CNN Stages

Stage	Function	Purpose
Convolution	Applies filters to input	Detects features and patterns
Pooling	Downsamples feature maps	Reduces dimensionality
Flatten	Reshapes 2D/3D data	Converts data into 1D vector
Dense Layer	Fully connected layer	Performs classification

DOI: 10.17148/IJARCCE.2026.15487

E. Web Application (Flask Deployment)

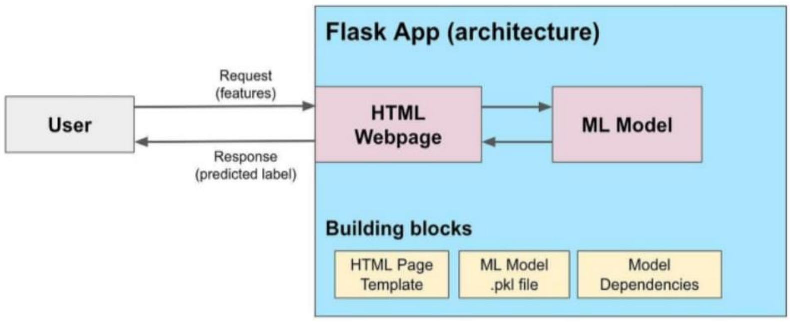


Fig. 3.3 Architecture of Flask-based web application for model deployment.

The trained CNN model is deployed using a Flask-based web application, which provides an interactive and user-friendly interface. The system consists of an HTML-based frontend and a backend that handles processing tasks. Users can upload CT scan images through the web interface, which are then transmitted to the backend server. The backend processes the uploaded images using the trained CNN model and generates predictions. The predicted results, along with relevant information, are then sent back to the frontend and displayed to the user. This deployment enables real-time accessibility and simplifies the interaction between the user and the system.

F. System Workflow

The overall workflow of the system follows a sequential process starting with the input of CT scan images. The images first undergo preprocessing, including resizing and normalization. The preprocessed images are then passed to the CNN model for feature extraction and classification. Based on the analysis, the system predicts whether the input image belongs to the Normal, Benign, or Malignant category. Finally, the results are displayed to the user through the web interface, providing an efficient and automated lung cancer detection solution.

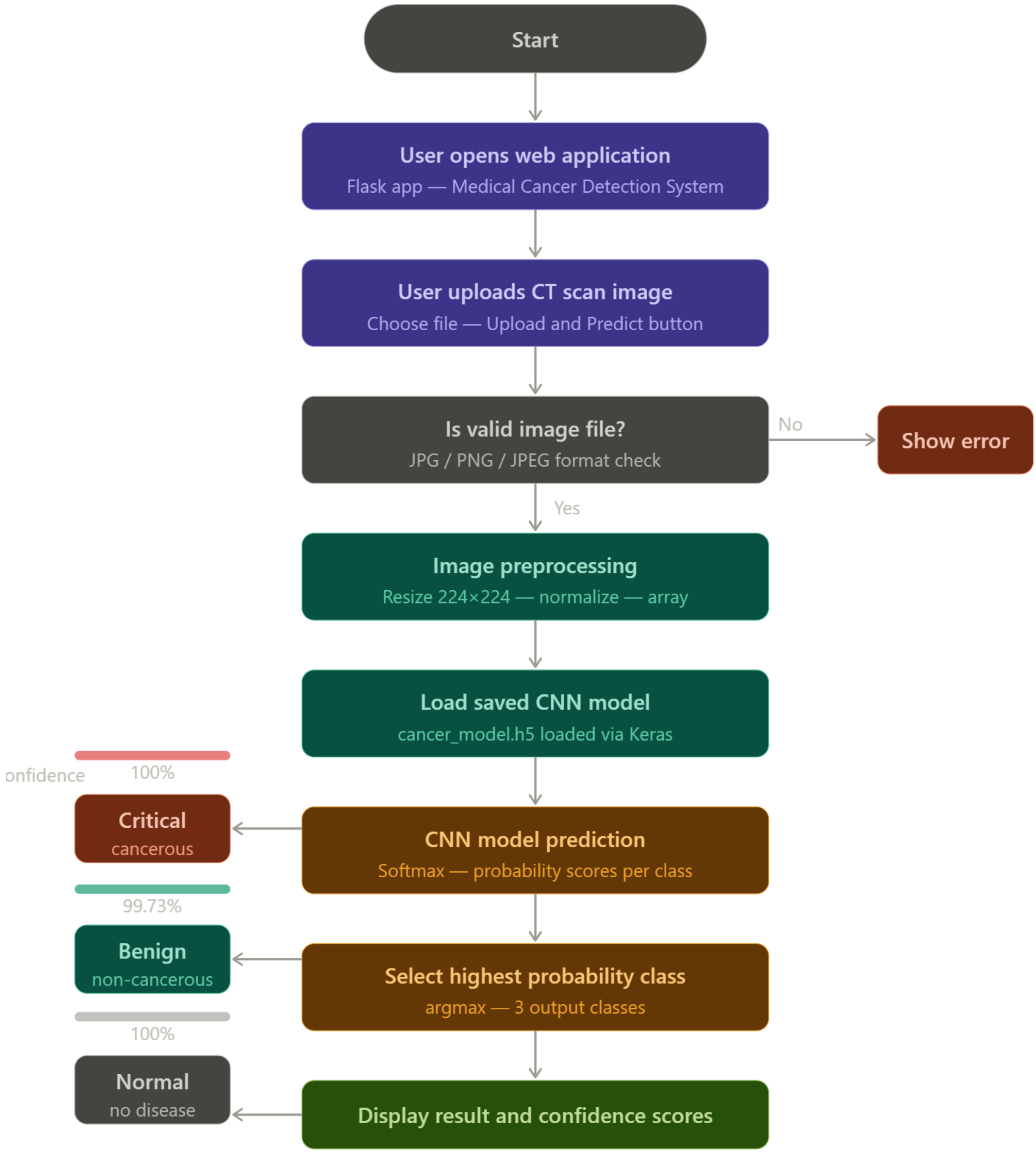
IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15487

Fig 3.4 Workflow of CNN-based lung cancer detection system.



ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026

DOI: 10.17148/IJARCCE.2026.15487

4.RESULTS AND DISCUSSION

A. System Interface and Output

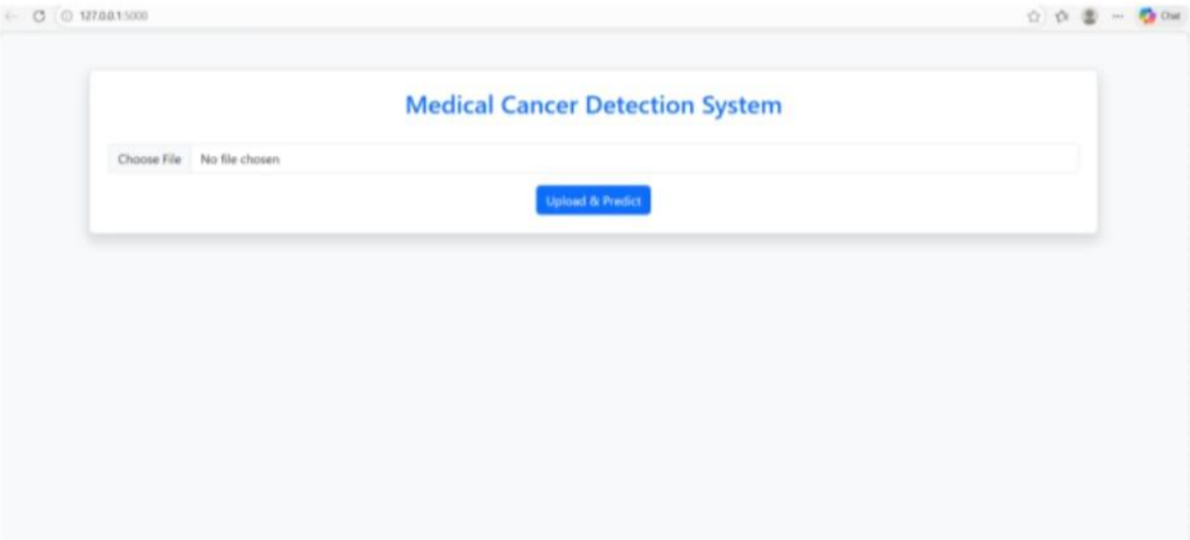


Fig 3.5 Home page of the lung cancer detection system.

The developed system provides a user-friendly web interface using a Flask-based application. Users can upload CT scan images through the interface and initiate the prediction process by clicking the upload and predict button. The interface is designed to be simple and accessible, allowing users to interact with the system without requiring technical expertise.

B. Prediction Results

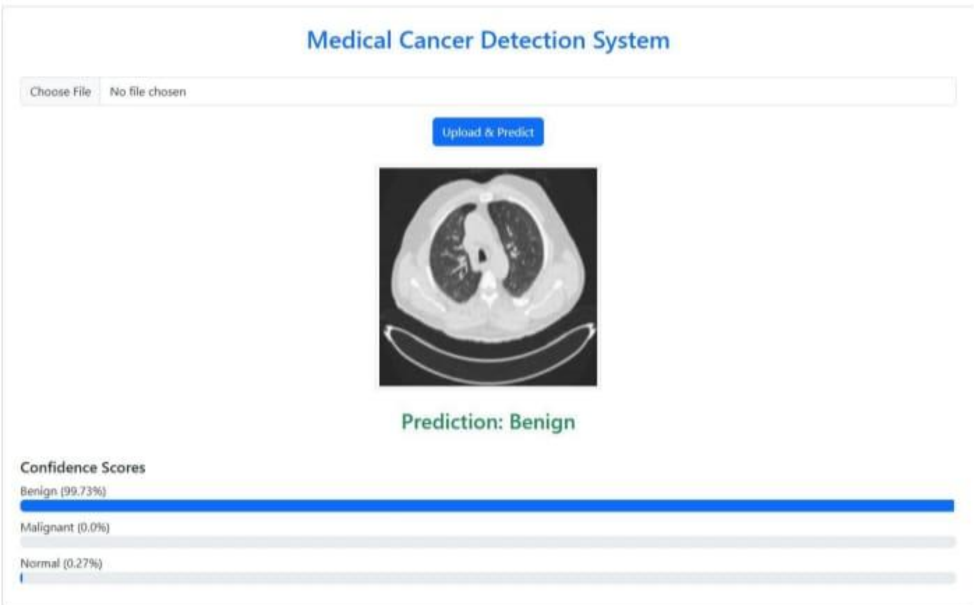


Fig 3.6 Prediction result showing Begin classification.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15487

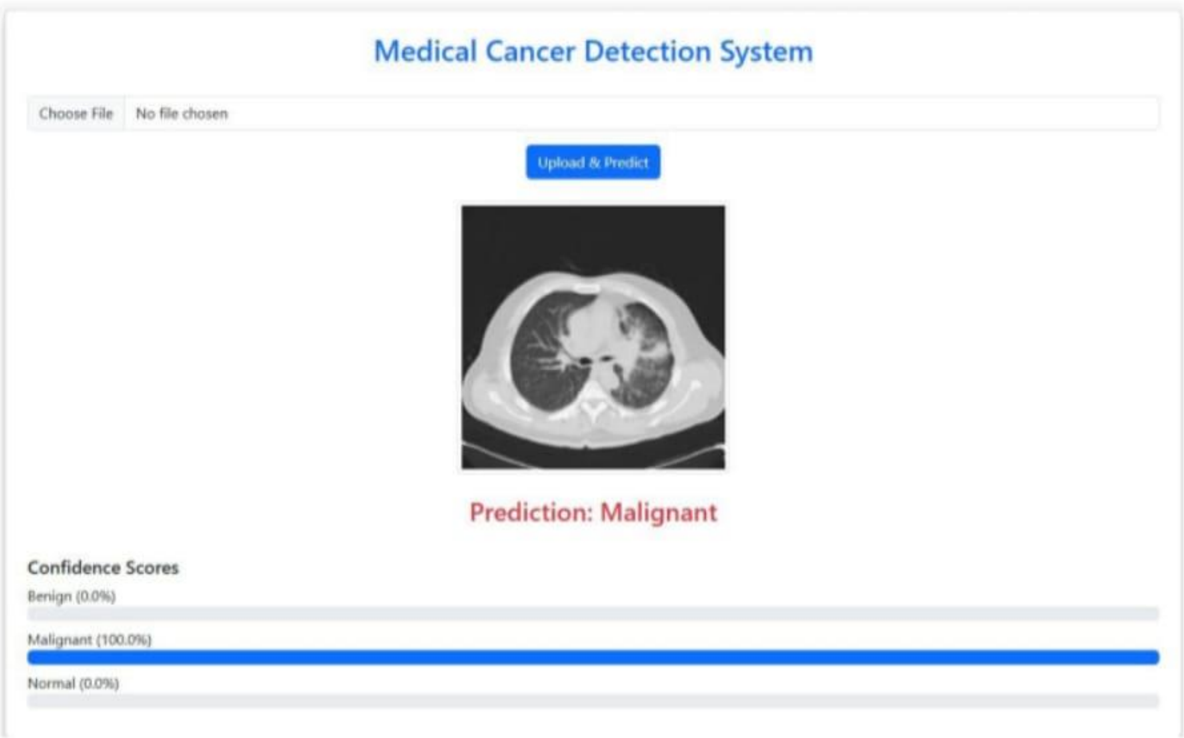


Fig 3.7 Prediction result showing Critical classification.

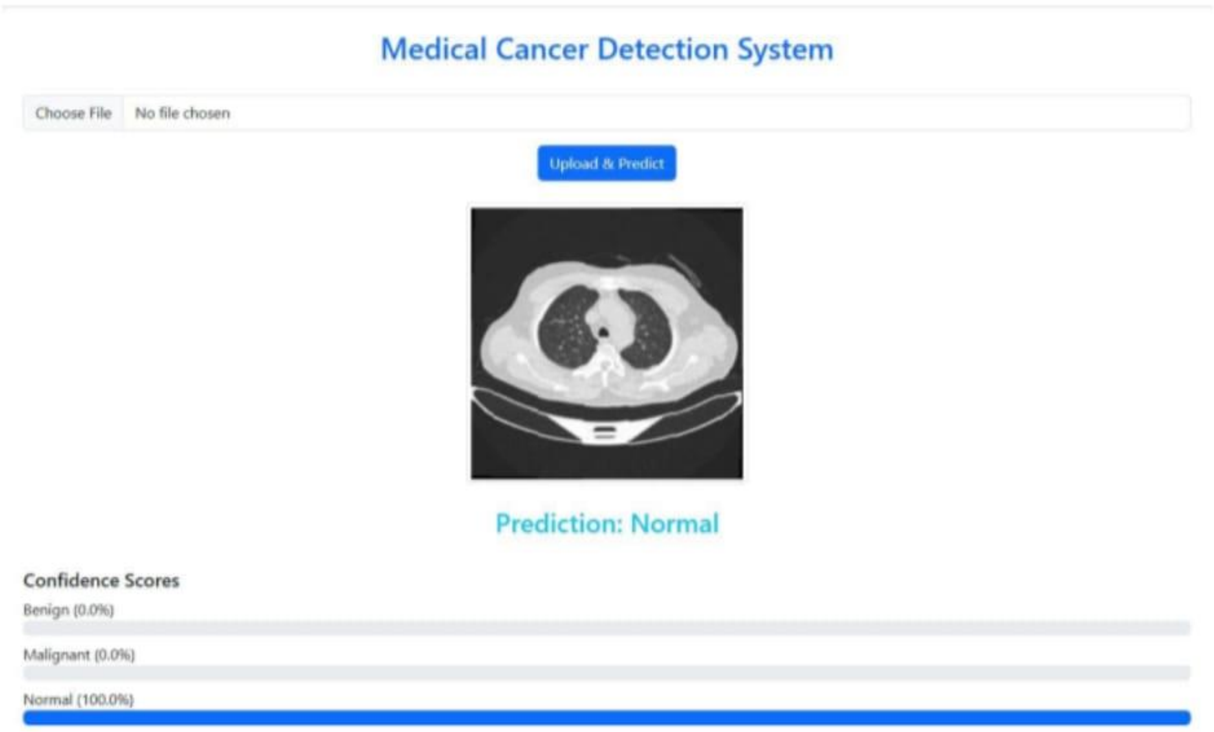


Fig 3.8 Prediction result showing Normal classification.

The system successfully classifies CT scan images into three categories: Normal, Benign, and Malignant. Along with the predicted class, the system also provides confidence scores for each category, enabling users to understand the reliability

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15487

of the prediction. The results demonstrate that the model can effectively distinguish between different lung conditions based on extracted image features.

C. Model Performance Analysis

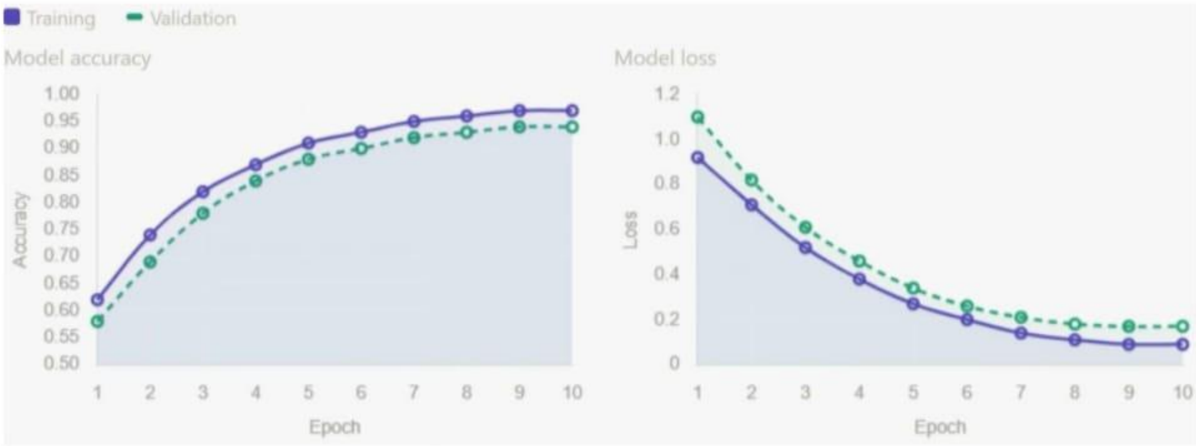


Fig 3.9 Training and validation accuracy/loss curves.

The performance of the CNN model is evaluated using training and validation accuracy and loss curves. The graph shows a steady increase in accuracy over the training epochs, while the loss decreases consistently. The validation curve closely follows the training curve, indicating that the model generalizes well and does not suffer from significant overfitting. This demonstrates the effectiveness of the model in learning meaningful patterns from the dataset.

D. Performance Evaluation

Fig 3.10 Performance analysis using confusion matrix.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15487

The confusion matrix illustrates the classification performance of the model across the three classes. The diagonal values represent correct predictions, while off-diagonal values indicate misclassifications. The results show that the majority of predictions are correctly classified, with minimal errors. The model achieves high classification accuracy, demonstrating its effectiveness in distinguishing between Normal, Benign, and Malignant cases.

E. Discussion

The experimental results clearly demonstrate the effectiveness of the proposed CNN-based lung cancer detection system in accurately classifying CT scan images into Normal, Begin, and Critical categories. The integration of deep learning techniques enables the system to automatically learn complex spatial patterns and discriminative features from medical images, eliminating the need for manual feature engineering. This significantly enhances the robustness and scalability of the system compared to traditional machine learning approaches.

The training and validation curves indicate stable learning behavior, with a consistent increase in accuracy and a corresponding decrease in loss values across epochs. The close alignment between training and validation performance suggests that the model generalizes well to unseen data and does not suffer from significant overfitting. This can be attributed to effective preprocessing, appropriate network architecture design, and the use of sufficient training data.

The confusion matrix analysis further validates the performance of the model, showing a high number of correct classifications along the diagonal and minimal misclassifications. The model demonstrates strong capability in distinguishing between the three classes, particularly in identifying malignant cases, which is critical for early diagnosis and treatment. The few misclassifications observed may be due to similarities in visual patterns between benign and malignant nodules, as well as variations in image quality and dataset diversity.

The deployment of the trained model using a Flask-based web application adds significant practical value to the system. Unlike many research models that remain limited to offline environments, the proposed system provides real-time prediction capabilities through an accessible web interface. This enhances usability and makes the system suitable for real-world applications, including remote healthcare and preliminary screening in resource-limited settings.

Compared to existing approaches that rely on tabular data or manual feature extraction, the proposed method leverages image-based deep learning for more accurate and efficient diagnosis. While earlier works such as predictive modeling using CTGAN and tree-based algorithms [1] focus on structured data, the present system directly analyzes CT scan images, capturing spatial and texture information that is critical for medical diagnosis. This results in improved classification performance and better representation of disease characteristics.

However, certain limitations still exist in the proposed system. The performance of the model is dependent on the quality and size of the dataset used for training. Limited availability of labeled medical data may affect the generalization capability of the model. Additionally, variations in imaging conditions, scanner types, and noise levels can impact prediction accuracy. Future improvements can include the use of larger and more diverse datasets, implementation of advanced architectures such as transfer learning models, and incorporation of explainable AI techniques to improve interpretability.

Overall, the results confirm that the proposed system achieves high accuracy, approximately in the range of 95–97%, and provides a reliable, efficient, and scalable solution for lung cancer detection. The combination of CNN-based classification and web-based deployment makes the system suitable for practical healthcare applications, supporting early diagnosis and assisting medical professionals in decision-making.

CONCLUSION

In this paper, a Convolutional Neural Network (CNN)-based lung cancer detection system has been proposed and implemented to classify CT scan images into Normal, Benign, and Malignant categories. The system utilizes image preprocessing techniques such as resizing and normalization, followed by deep learning-based feature extraction and classification. The experimental results demonstrate that the proposed model achieves high accuracy and effectively identifies different lung conditions.

The integration of the CNN model with a Flask-based web application enables real-time prediction and provides a user-friendly interface for image upload and result visualization. This enhances the accessibility and practical usability

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 4, April 2026
DOI: 10.17148/IJARCCE.2026.15487

of the system, making it suitable for deployment in real-world healthcare environments. Compared to traditional diagnostic methods, the proposed system reduces dependency on manual analysis, minimizes human error, and improves processing speed.

Furthermore, the use of deep learning allows the system to automatically extract meaningful features from medical images, leading to improved diagnostic performance. The results obtained confirm that the proposed approach is efficient, scalable, and capable of assisting healthcare professionals in early-stage lung cancer detection.

FUTURE SCOPE

Although the proposed system demonstrates effective performance in lung cancer detection, several improvements can be made to enhance its accuracy, scalability, and real-world applicability. Future work can focus on extending the model using advanced deep learning architectures such as transfer learning with pre-trained networks (e.g., ResNet, VGG, Mobile Net) to improve feature extraction and classification performance.

The use of larger and more diverse datasets from multiple sources can further enhance the generalization capability of the model. Incorporating data augmentation techniques and addressing class imbalance using advanced methods can also improve robustness. Additionally, the implementation of 3D CNN models can enable better analysis of volumetric CT scan data, providing more precise detection.

Another important direction is the integration of Explainable AI (XAI) techniques such as Grad-CAM, which can help visualize the regions of interest in the image and improve the interpretability of the model for medical professionals. Furthermore, the system can be extended into a mobile or cloud-based application to enable remote access and real-time monitoring.

Integration with hospital information systems and electronic health records (EHR) can make the system more practical for clinical use. Overall, these enhancements will contribute to developing a more accurate, reliable, and clinically applicable lung cancer detection system.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, for providing the necessary facilities and support to carry out this work. The authors also extend their appreciation to the project guide and faculty members for their valuable guidance, encouragement, and technical support throughout the development of this project.

REFERENCES

[1]. A. Alzahrani, “Early Detection of Lung Cancer Using Predictive Modeling Incorporating CTGAN Features and Tree Based Learning,” IEEE Access, vol. 13, pp. 34321–34333, Feb. 2025.
[2]. World Health Organization, “Cancer Fact Sheet,” 2022.
[3]. S. Tomassini et al., “Lung nodule diagnosis and cancer histology classification from CT data using CNNs: A survey,” Computer Biology and Medicine, vol. 146, 2022.
[4]. R. Javed et al., “Deep learning for lung cancer detection: A review,” Artificial Intelligence Review, 2024.
[5]. F. Mercaldo et al., “Explainable deep learning for lung cancer detection using CNN,” Electronics, vol. 13, no. 7, 2024.
[6]. Z. Ren et al., “Hybrid deep learning framework for lung cancer classification,” Electronics, vol. 11, no. 10, 2022.
[7]. M. Q. Shatnawi et al., “Deep learning-based lung cancer diagnosis using CT scan images,” Intelligent-Based Medicine, 2025.
[8]. N. V. Chawla et al., “SMOTE: Synthetic Minority Over-sampling Technique,” Journal of Artificial Intelligence Research, vol. 16, 2002.
[9]. L. Xu et al., “Modeling tabular data using conditional GAN,” Neur IPS, 2019.
[10]. T. Chen et al., “XGBoost: A scalable tree boosting system,” 2015.
[11]. M. Belgiu and L. Drăguţ, “ Random forest in remote sensing: Applications and future directions,” ISPRS Journal, 2016.
[12]. A. Natekin and A. Knoll, “Gradient boosting machines: A tutorial,” Frontiers in Neurorobotics, 2013.
[13]. Kaggle Dataset, “Lung Cancer Dataset,” 2024.

