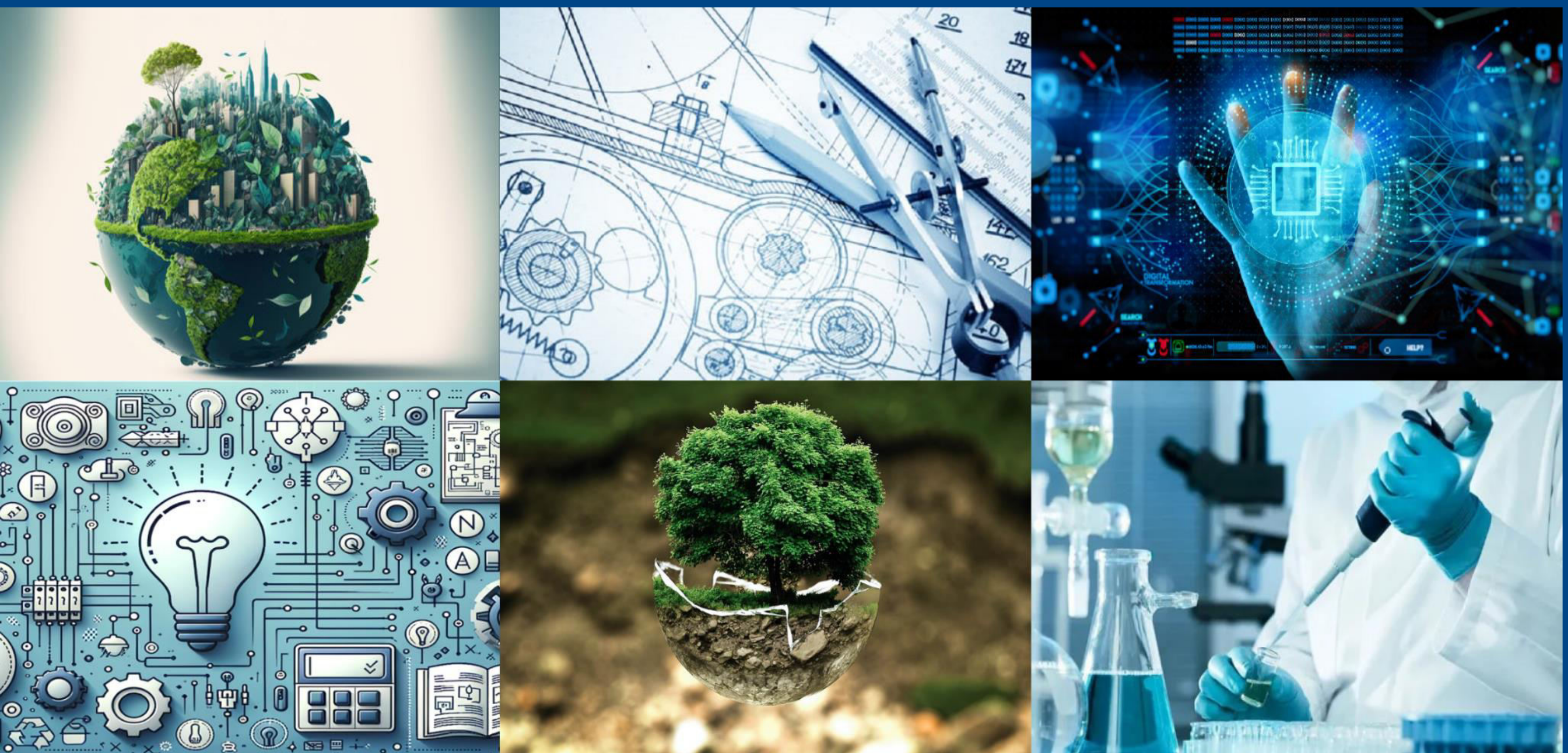
ISSN: 2582-7219

# International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206

Volume 9, Issue 3, March 2026

© 2026 IJMRSET | Volume 9, Issue 3, March 2026 DOI:10.15680/IJMRSET.2026.0903206

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

# Lung Cancer Classification using Deep Learning

**V.Bhavana, S.Janaki Ram, S.Lokesh, P.Venu Gopi**

UG Student, Department of ECE, R.V.R & J.C.C.E, Chowdavaram, India

UG Student, Department of ECE, R.V.R & J.C.C.E, Chowdavaram, India

UG Student, Department of ECE, R.V.R & J.C.C.E, Chowdavaram, India

UG Student, Department of ECE, R.V.R & J.C.C.E, Chowdavaram, India

**ABSTRACT**: Lung cancer is a serious health problem that affects a large number of people around the world. Detecting it at an early stage is very important, but analyzing CT scan images manually can take a lot of time and may sometimes lead to mistakes. To make this process easier and more accurate, this project focuses on using deep learning techniques for automatic lung cancer classification.

In this work, models like CNN, ResNet50, and EfficientNet-B4 are used to analyze CT scan images and classify them as normal or cancerous. The process includes preparing the data, improving it using augmentation, training the models, and checking their performance using different evaluation metrics. Among all models, ResNet50 gives better results. A simple web interface is also developed so that users can upload images and get predictions easily. This system helps reduce manual effort and supports faster decision-making.

**KEYWORDS**: Lung Cancer Detection ,Deep Learning, Transfer Learning, CNN, ResNet50, EfficientNet-B4

## I. INTRODUCTION

Lung cancer is one of the leading causes of death worldwide, and the number of cases is increasing every year. It usually starts when abnormal cells in the lungs grow uncontrollably and may spread to other parts of the body. Detecting the disease early can make a big difference in treatment and survival.

In most hospitals, doctors examine CT scan images to identify lung cancer. However, this process can be difficult because there are many images to analyze, and it requires a lot of experience. Sometimes, it can also lead to delays or errors.

With the development of artificial intelligence, especially deep learning, it has become possible to analyze medical images more efficiently. In this project, deep learning models are used to automatically detect lung cancer from CT scan images. The aim is to make the diagnosis process faster, easier, and more reliable.

## II. LITERATURE REVIEW

Lung cancer detection using medical imaging has gained significant attention with the advancement of deep learning techniques. Earlier approaches mainly relied on traditional machine learning methods, where features were manually extracted from CT images. However, these methods were limited in their ability to capture complex patterns, leading to lower accuracy. With the emergence of Convolutional Neural Networks (CNNs), researchers began to achieve better performance in medical image classification.

| Author& Year | Method Used | Dataset | Performance |
|---|---|---|---|
| Shen et al. (2017) | Multi-scale CNN | LIDC-IDRI | ~86% Accuracy |

© 2026 IJMRSET | Volume 9, Issue 3, March 2026 DOI:10.15680/IJMRSET.2026.0903206

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

| | | | |
|---|---|---|---|
| Kumar et al. (2018) | Transfer Learning (AlexNet, VGG) | Private CT Dataset | Accuracy ~89% |
| Coudray et al. (2018) | InceptionV3 CNN | TCGA Dataset | ~95% Accuracy |
| Ardila et al. (2019) | 3D Deep Learning Model | NLST Dataset | AUC ~0.94 |
| Hosny et al. (2018) | Radiomics + Deep Learning | NLST, TCIA | AUC ~0.90 |

## III. PROBLEM STATEMENT

Even though medical technology has improved a lot, detecting lung cancer accurately is still not easy. Doctors need to go through many CT scan images, which increases their workload and can sometimes lead to mistakes. In some places, there are not enough specialists, which causes delays in diagnosis. Existing systems also have some drawbacks, such as limited data and lower accuracy in real-world conditions. Because of these issues, there is a need for a better system that can provide accurate results and help in early detection.

## IV. OBJECTIVES

The main goal of this project is to create a system that can automatically detect lung cancer using deep learning.
The objectives are:
To prepare CT scan images for model training
To improve data quality using augmentation techniques
To use models like CNN, ResNet50, and EfficientNet-B4
To apply transfer learning for better results
To classify images into normal and cancer categories
To evaluate the model performance
To build a simple interface for user interaction

## V. FLOWCHART

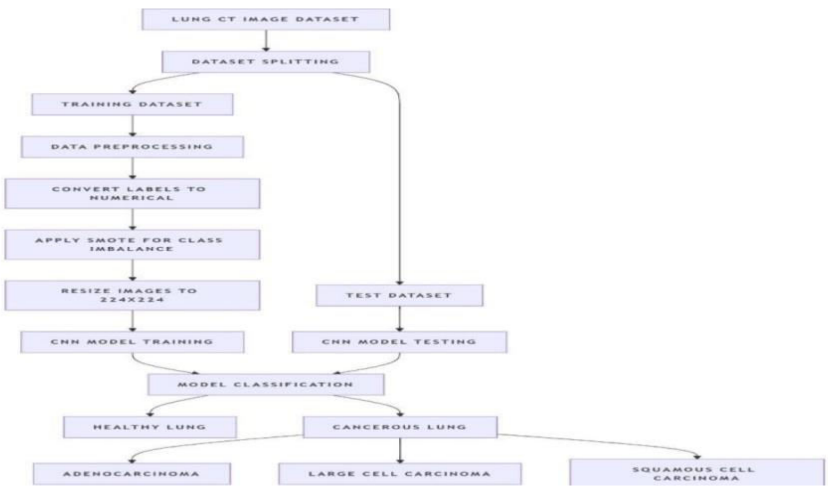



Fig 1: Flow Chart

The proposed system begins with collecting a lung CT image dataset, which forms the foundation for training and evaluation. The dataset is divided into training and testing sets to ensure proper model validation. In the training phase, the images undergo preprocessing steps such as noise removal, normalization, and enhancement to improve quality. The class labels are then converted into numerical form so that they can be processed by the model. To handle class imbalance, the SMOTE technique is applied, which generates synthetic samples for underrepresented classes. All

© 2026 IJMRSET | Volume 9, Issue 3, March 2026 DOI:10.15680/IJMRSET.2026.0903206

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

images are resized to 224×224 pixels to maintain uniform input dimensions for the CNN model. The processed data is then used to train the convolutional neural network, allowing it to learn important features from the images.

In parallel, the test dataset is used to evaluate the trained model. After training and testing, the model performs classification to determine whether the lung is healthy or cancerous. If classified as cancerous, the system further identifies the specific type of lung cancer, such as adenocarcinoma, large cell carcinoma, or squamous cell carcinoma. This multi-level classification helps in providing more precise diagnostic insights.

## VI. METHODOLOGY

**Overview**
The proposed system follows a structured approach to detect and classify lung cancer from CT scan images using deep learning techniques. The process includes data collection, preprocessing, model training, evaluation, and deployment through a web interface.

**Step 1: Data Collection**
The first step is to collect CT scan images of lungs from available datasets. The dataset contains both normal and cancerous images, including different types of lung cancer. This data is used to train and test the models.

**Step 2: Data Preprocessing**
Before giving the images to the model, preprocessing is performed to improve data quality. This includes:
Resizing images to a fixed size
Normalizing pixel values
Labeling images into categories
These steps help the model learn more effectively.

**Step 3: Data Augmentation**
To increase the amount of training data and avoid overfitting, augmentation techniques are applied. Some common methods used:
Rotation
Flipping
Zooming
This makes the model more robust and improves performance.

**Step 4: Model Development**
In this project, three deep learning models are used:
CNN (basic model)
ResNet50 (deep model with better learning ability)
EfficientNet-B4 (optimized model for performance and efficiency)
Transfer learning is applied to ResNet50 and EfficientNet-B4 using pre-trained weights, which helps in achieving better accuracy in less time.

**Step 5: Model Training**
The models are trained using the prepared dataset.
The training process uses an optimizer (like Adam)
Loss function is used to measure error
The model learns patterns from CT scan images
Training is done for multiple epochs until the model performs well.

**Step 6: Model Evaluation**
After training, the models are evaluated using standard metrics:
Accuracy
Precision
Recall

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

F1-score
These metrics help in comparing different models and selecting the best one.

**Step 7: Model Selection**
Based on the evaluation results, the best-performing model is selected.
In this project, ResNet50 provides better accuracy compared to CNN and EfficientNet-B4, so it is chosen for deployment

**Step 8: Deployment with Web Interface**
The final model is integrated into a web application.
Backend is developed using Flask
Frontend is created using HTML, CSS, and JavaScript
Users can upload CT scan images through the interface, and the system will display the prediction result (normal or cancer)

## VII. IMPLEMENTATION OF THE SYSTEM REQUIREMENTS FOR HARDWARE AND SOFTWARE

**Hardware Requirements**
The system requires a standard computing device with sufficient processing capability for training and testing deep learning models. A system with at least an Intel i5 processor (or equivalent), 8 GB RAM, and optional GPU support is recommended for faster computation. Adequate storage is also required to handle medical image datasets.

**Software Requirements**
The implementation is carried out using Python programming language along with deep learning and data processing libraries such as TensorFlow, Keras, NumPy, and OpenCV. Development is performed using environments like Jupyter Notebook or Google Colab. A web framework such as Flask is used for building the user interface, and standard tools like HTML, CSS, and JavaScript are used for frontend development.

## VIII. DEEP LEARNING MODELS

**1.Convolutional Neural Network (CNN)**
CNN is a basic and widely used deep learning model for image classification tasks. It automatically extracts features such as edges, textures, and shapes from input images.The model consists of convolution layers, pooling layers, and fully connected layers. These layers work together to learn patterns from CT scan images and classify them into different categories. CNN does not require manual feature extraction, making it efficient for medical imaging tasks.However, since it is a relatively simple model, its performance may be lower compared to advanced architectures when dealing with complex datasets.

**2. ResNet50**
ResNet50 is a deep neural network with 50 layers that uses a concept called residual learning. It introduces shortcut connections (skip connections) that allow the network to bypass certain layers, solving the problem of vanishing gradients in deep networks.This enables the model to learn deeper and more complex features from CT images. In this project, ResNet50 is used with transfer learning, where a pre-trained model is fine-tuned for lung cancer classification.ResNet50 provides very high accuracy and is highly effective for multi-class classification problems in medical imaging.

**3. EfficientNet-B4**
EfficientNet-B4 is an advanced deep learning model designed to achieve high accuracy with optimal efficiency. It uses a technique called compound scaling, which balances the network's depth, width, and resolution. This model captures fine details in CT images and performs well even with limited data. It also requires fewer parameters compared to other deep networks while maintaining high performance.EfficientNet-B4 is known for providing a good balance between accuracy and computational cost, making it suitable for real-world applications.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

## IX. IMPLEMENTATION OF MODEL

The proposed system integrates a deep learning model with a web-based frontend for lung cancer classification. Initially, CT scan images are pre processed by resizing, normalization, and data augmentation to improve model performance.

The classification model is developed using CNN, ResNet50, and EfficientNet-B4 architectures. Transfer learning is applied using pre-trained weights, and the models are trained using the Adam optimizer with categorical cross-entropy loss. Among the models, ResNet50 achieves the best performance and is selected for deployment.

For user interaction, a web-based frontend is developed using HTML, CSS, and JavaScript. The backend is implemented using Flask, where the trained model is loaded for prediction. Users can upload CT scan images through the interface, and the system processes the image and provides instant classification results, indicating whether the lung is normal or cancerous along with the specific cancer type.

This integration ensures a simple, efficient, and user-friendly system that can assist medical professionals in diagnosis.

## X. MODELS RESULTS

**1. CNN**

Confusion matrix for CNN

The confusion matrix shows most predictions on the diagonal, indicating correct classification, with minor confusion between cancer classes.

The model achieves 89% accuracy, performing best on normal images. Performance can be improved using data augmentation and transfer learning models like ResNet50.

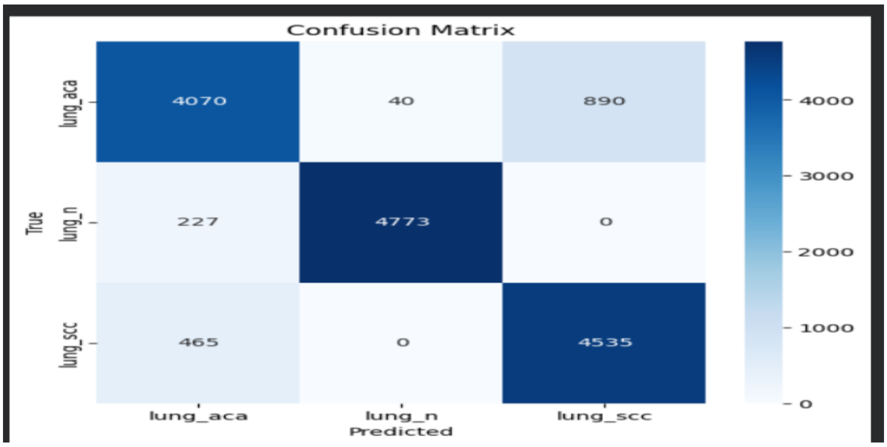



Fig 2: Confusion matrix for CNN

Classification Report for CNN

This classification report shows that the ResNet50 model achieved 100% accuracy with perfect precision, recall, and F1-score for all classes.

Each class has equal support, so macro and weighted averages are also 1.00.

Although the model shows perfect performance, further validation on independent test data is necessary to confirm generalization

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Classification Report:

| | precision | recall | f1-score |
|---|---|---|---|
| lung_aca | 85% | 81% | 83% |
| lung_n | 99% | 95% | 97% |
| lung_scc | 84% | 91% | 87% |
| accuracy | | | 89% |
| macro avg | 89% | 89% | 89% |
| weighted avg | 89% | 89% | 89% |

Fig 3 . Classification Report for CNN

Graphs for CNN

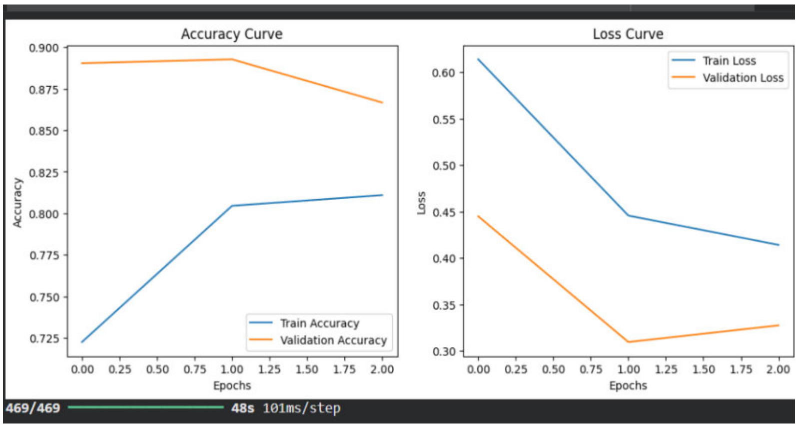



Fig 4: Graphs for CNN

**2. EFFICIENT NET B4**

Confusion Matrix for EfficientNet B4

The confusion matrix shows that EfficientNetB4 performs well with high correct predictions on the diagonal. However, there is significant misclassification between lung_scc and lung_aca, and some normal cases are incorrectly classified as cancer. The overall accuracy is around 94%, which is better than CNN but slightly lower than ResNet50.

Misclassification between cancer subtypes occurs due to similar visual patterns in CT images.

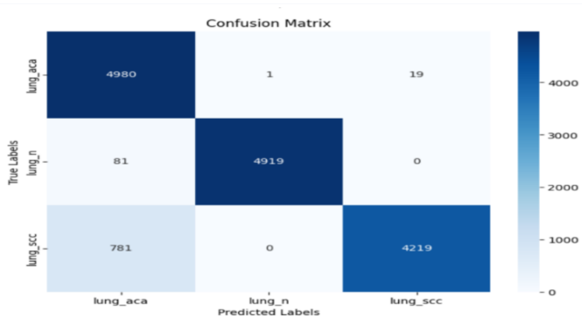



Fig 5: Confusion Matrix for EfficientNet B4

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

**Classification Report for EfficientNetB4**

This classification report shows the performance of the EfficientNetB4 model.The overall accuracy of the model is 94.12%, indicating good performance.

For some cases it shows imbalance between precision and recall, meaning some cases are misclassified.Overall, the model performs well but can be improved for better cancer subtype detection.

| | precision | recall | f1-score |
|---|---|---|---|
| lung_aca | 85.24 % | 99.6 % | 91.86 % |
| lung_n | 99.98 % | 98.38 % | 99.17 % |
| lung_scc | 99.55 % | 84.38 % | 91.34 % |
| accuracy | 94.12 % | 94.12 % | 94.12 % |
| macro avg | 94.93 % | 94.12 % | 94.13 % |
| weighted avg | 94.93 % | 94.12 % | 94.13 % |

Fig 6: Classification Report for EfficientNetB4

Graphs for Efficient Net B4

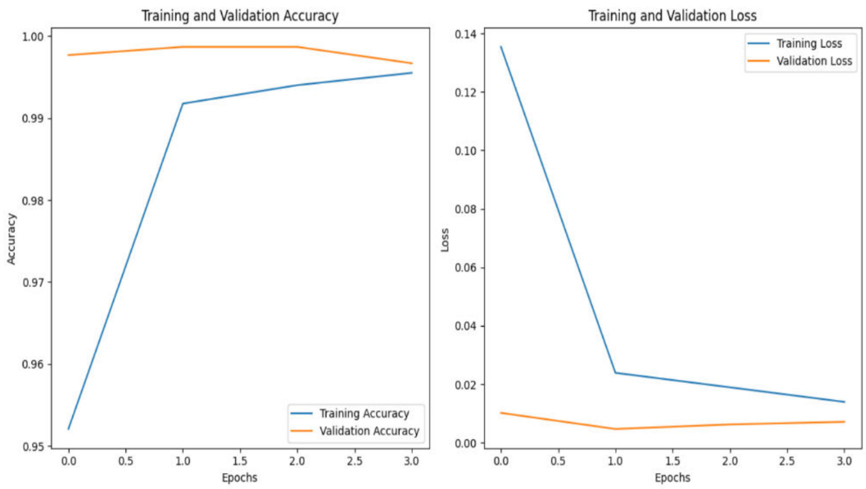



Fig 7: Graphs for Efficient Net B4

**3. Resnet 50**

Confusion Matrix for ResNet 50This confusion matrix shows that the ResNet50 model performs almost perfectly, with most predictions on the diagonal indicating correct classification.

The normal class is classified with 100% accuracy, while very few misclassifications occur between adenocarcinoma and squamous cell carcinoma. Overall, the model achieves near-perfect accuracy.

© 2026 IJMRSET | Volume 9, Issue 3, March 2026 DOI:10.15680/IJMRSET.2026.0903206

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

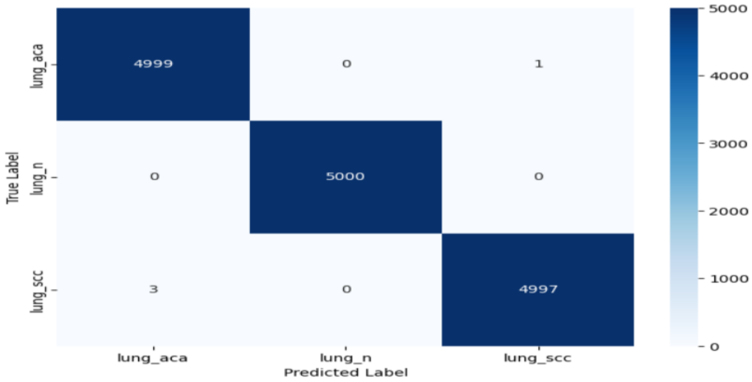



Fig 8: Confusion Matrix for ResNet 50

Classification Report for Resnet 50This classification report shows that the ResNet50 model achieved 100% accuracy with perfect precision, recall, and F1-score for all classes.

Each class has equal support, so macro and weighted averages are also 1.00.

Although the model shows perfect performance, further validation on independent test data is necessary to confirm generalization.

| | precision | recall | f1-score |
|---|---|---|---|
| lung_aca | 99.80% | 99.94% | 99.87% |
| lung_n | 100.00% | 100.00% | 100.00% |
| lung_scc | 99.94% | 99.80% | 99.87% |
| accuracy | 99.91% | 99.91% | 99.91% |
| macro avg | 99.91% | 99.91% | 99.91% |
| weighted avg | 99.91% | 99.91% | 99.91% |

Fig 9 . Classification Report for Resnet 50

Graphs for Resnet50

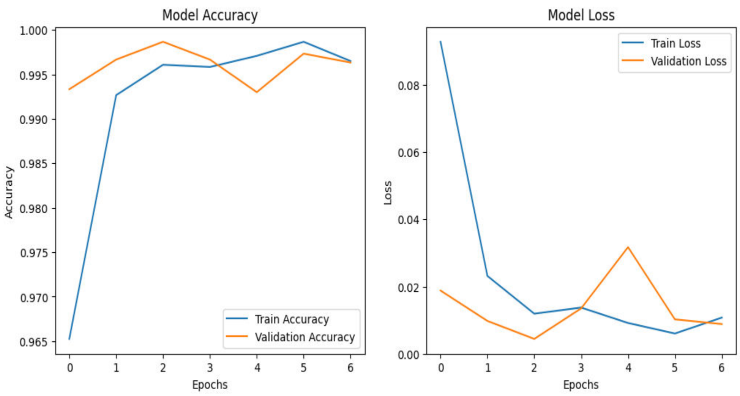



Fig 10: Graphs for Resnet50

© 2026 IJMRSET | Volume 9, Issue 3, March 2026 DOI:10.15680/IJMRSET.2026.0903206

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

## XI. WEB INTERFACE

The web application provides a simple interface for users to upload lung images and get predictions. After uploading, the system analyzes the image using the trained model and displays the detected cancer type along with a confidence score. The uploaded image is also shown for reference, making the process quick and easy to use.

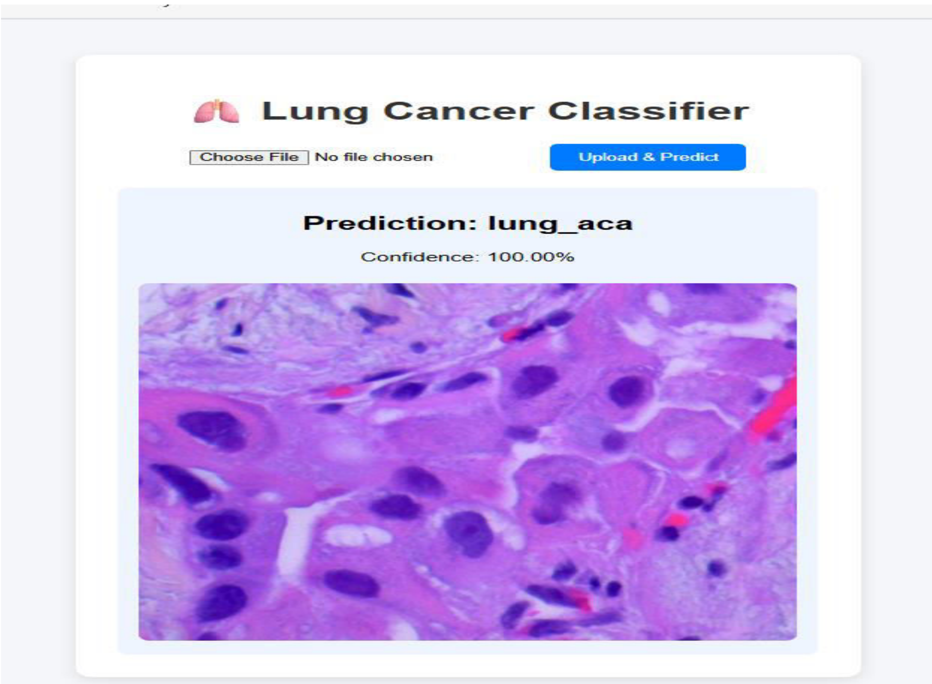



Fig 11: WEB INTERFACE

## REFERENCES

1. [W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," IEEE Transactions on Biomedical Engineering, vol. 64, no. 2, pp. 493–502, 2017. Setio et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge," Medical Image Analysis, vol. 42, pp. 1–13, 2017.
2. K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," OncoTargets and Therapy, vol. 8, pp. 2015–2022, 2015.
3. D. Ardila et al., "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," Nature Medicine, vol. 25, pp. 954–961, 2019.
4. N. Coudray et al., "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," Nature Medicine, vol. 24, no. 10, pp. 1559–1567, 2018.
5. M. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, "Artificial intelligence in radiology," Nature Reviews Cancer, vol. 18, no. 8, pp. 500–510, 2018.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.
7. M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proceedings of the International Conference on Machine Learning (ICML), 2019.

# INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY