International Journal for Modern Trends in Science and Technology
Volume 11, Issue 05, pages 1145-1150
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol11issue05.html
DOI: https://doi.org/10.5281/zenodo.15477817

Pioneering Lung Health Diagnostics: Leveraging Transfer Learning with Biopsy Imaging for Accurate Lung Carcinoma Detection

Rama Krishna Raju Chekuri[1], Dr.R.V.V.S.V.Prasad[2], Simhadri Sai Renu Chandrika[3], Vegesna Ahladita Naga Vaishnavi[4], Kotipalli Manvitha Lakshmi Sai[5]

Department of Information Technology Swarnandhra College of Engineering and Technology (A), Seetharampuram, Narsapur, AP 534280
rajuchekuri29@gmail.com, ramayanam.prasad@gmail.com, renuchandrika700@gmail.com, vegesnavaishnavi1@gmail.com, manvithakotipalli914@gmail.com

To Cite this Article
Rama Krishna Raju Chekuri, Dr.R.V.V.S.V.Prasad, Simhadri Sai Renu Chandrika, Vegesna Ahladita Naga Vaishnavi & Kotipalli Manvitha Lakshmi Sai (2025). Pioneering Lung Health Diagnostics: Leveraging Transfer Learning with Biopsy Imaging for Accurate Lung Carcinoma Detection. International Journal for Modern Trends in Science and Technology, 11(05), 1145-1150. https://doi.org/10.5281/zenodo.15477817

Article Info
Received: 17 April 2025; Accepted: 18 May 2025.; Published: 21 May 2025.

KEYWORDS

Lung cancer, EfficientNet, Deep learning, Convolutional Neural Networks, Cancer detection, Automated diagnosis

ABSTRACT

Lung cancer, one of the most fatal diseases worldwide, necessitates early and accurate diagnosis to improve survival rates. This study explores the application of Convolutional Neural Networks (CNNs) for automating the detection of lung cancer, specifically leveraging histopathological biopsy images. To address the limitations of manual diagnosis, including its time-consuming and error- prone nature, a pre-trained EfficientNet-based model was employed for precise image classification. The proposed approach classifies lung cancer into benign and malignant categories, demonstrating remarkable performance with robust accuracy and interpretability. The model was fine- tuned to enhance subtype differentiation and adapt to real- time intraoperative analysis, showcasing its scalability to other cancer types as well. Experimental results reveal that the EfficientNet architecture not only surpasses conventional CNN models in terms of accuracy and efficiency but also minimizes computational requirements, making it a viable tool for large-scale clinical implementation. This project underscores the transformative potential of machine learning in radiology and oncology, paving the way for improved diagnostic precision, personalized treatments, and better patient outcomes.

1. INTRODUCTION

Lung carcinoma is one of the most prevalent and deadly forms of cancer, accounting for a significant percentage of cancer-related fatalities worldwide. Early and accurate detection is critical for improving patient survival rates, as timely intervention can greatly enhance treatment efficacy. Traditional diagnostic methods, particularly histopathological analysis of biopsy samples, are considered the gold standard. However, these methods often involve labor-intensive processes, requiring skilled pathologists to manually examine tissue samples under a microscope, which can lead to potential diagnostic variability and delays. [1]

In recent years, deep learning has emerged as a powerful tool for medical image analysis, offering automated, precise, and scalable solutions for disease detection. Among the various deep learning architectures, convolutional neural networks (CNNs) have demonstrated remarkable success in image- based classification tasks, including medical diagnostics. By learning hierarchical features from imaging data, CNNs can distinguish between subtle patterns in biopsy samples that may not be easily detectable by the human eye. This has opened new avenues for improving the accuracy and efficiency of lung carcinoma detection. One of the key challenges in training deep learning models for medical imaging is the need for large, well-annotated datasets. [2] Since collecting and labelling medical images is time- consuming and requires domain expertise, transfer learning has become a widely adopted strategy. Transfer learning allows models pre-trained on large-scale datasets, such as ImageNet, to be fine-tuned on specific medical imaging tasks, significantly reducing the need for extensive labelled data while enhancing performance. By leveraging pre-trained feature representations, transfer learning accelerates model convergence and improves generalization to new datasets. [3]

In this study, we utilize EfficientNetB3, a highly efficient CNN architecture, to detect lung carcinoma from biopsy images. EfficientNetB3 is part of the EfficientNet family, which optimizes model scaling using a compound coefficient that balances network depth, width, and resolution. This results in superior performance with fewer parameters and reduced computational costs compared to conventional CNN models. By fine-tuning EfficientNetB3 on biopsy imaging datasets, we aim to achieve high diagnostic accuracy while maintaining efficiency suitable for real-world clinical applications. The advantage of using EfficientNetB3 lies in its ability to extract complex histopathological features with minimal computational overhead. Its optimized architecture enhances feature representation, allowing the model to differentiate between malignant and benign tissue samples more effectively. Additionally, the model's capability to generalize across diverse biopsy images ensures robustness, reducing the likelihood of misdiagnosis due to variations in tissue morphology, staining techniques, or image quality. This makes it a promising candidate for improving lung carcinoma diagnostics. [4] As deep learning continues to evolve, its integration into medical diagnostics holds immense promise for improving patient outcomes and revolutionizing cancer detection methodologies. The adoption of AI-driven biopsy imaging analysis has far-reaching implications in clinical oncology, enabling early detection, improved patient outcomes, and optimized treatment planning. This research bridges the gap between computational pathology and AI, demonstrating how transfer learning with EfficientNetB3 can revolutionize lung carcinoma detection. By providing a scalable, cost-effective, and highly accurate diagnostic tool, this deep learning-based system has the potential to be integrated into real-world medical workflows, empowering healthcare professionals with cutting-edge technology to combat lung cancer more effectively.[5]

2. LITURETURE REVIEW

2.1 Hybrid Deep Learningfor Histopathological Classification

A hybrid CNN model integrating InceptionV3, HOG, and DAISY feature extractors achieved exceptional performance (99.97% accuracy) on the LC25000 dataset. This model emphasizes the strength of ensemble feature extraction to handle complex tissue structures effectively, enabling accurate classification across multiple cancer types [1].

2.2 Deep Learning with RNN-CNN and PSO

A novel integration of Recurrent Neural Networks (RNN) and CNN with Particle Swarm Optimization (PSO) was used for detecting nodules from chest radiographs. This hybrid model focused on improving training speed and diagnostic accuracy, particularly in

early-stage detection using CT scan images [2].

2.3 Importance of Preprocessing and Visualization

Another study focused on early-stage lung cancer identification using CNNs and stressed the significance of preprocessing techniques and data visualization in managing histopathological image datasets. Artificial Neural Networks (ANNs) were effectively employed for categorization, promoting interpretability and accuracy[3].

2.4 DL with DIP for Lung and Colon Cancer

Using Digital Image Processing (DIP) with CNNs, researchers showed AI's capability to classify cancer types—achieving 96% accuracy for colon polyps. This reaffirms the adaptability of CNN architectures across various cancers beyond lung carcinoma [4].

2.5 Deep Learning with Feature Fusion

Combining InceptionV3 with feature descriptors like HOG and DAISY, another model reported 99.96% accuracy. Implemented using PyTorch, the framework also involved weighted average precision and confusion matrix analysis for robust performance evaluation.

2.6 ML Algorithms with CatBoost

A study applied multiple machine learning algorithms (CatBoost, LDA, LR, CART) and achieved 99.80% accuracy using CatBoost on 15,000 histopathology images. The research highlighted the effectiveness of non-CNN models in certain scenarios, especially with high-dimensional data.

2.7 EfficientNet Variants and AI-Assisted Grading

The role of EfficientNetB2 was explored in grading cancer stages through clustering and classification. The model aimed to reduce pathologist workload and support regions with limited access to medical facilities. Reported accuracies ranged from 86% to 99% depending on the dataset and model tuning.

2.8 Lightweight End-to-End CNN Model

Recent works also proposed a lightweight, end-to-end CNN approach using EfficientNetB1 for real-time classification. These models exhibited excellent scalability and minimized overfitting, achieving over 99.5% accuracy in some cases and demonstrating suitability for intraoperative use.

3. PROPOSED SYSTEM

The proposed system leverages deep learning and transfer learning techniques to develop an accurate lung cancer detection model using CT scan images. It processes the dataset by resizing images to 256 × 256 pixels and applies batch-wise training for efficient learning. The system utilizes convolutional neural networks (CNNs) to extract meaningful features from lung images and classify them into three categories: normal lungs, adenocarcinoma, and squamous cell carcinoma. By implementing data augmentation and optimization techniques, the model enhances its generalization ability. The system is designed to provide high-accuracy predictions, assisting medical professionals in early and precise lung cancer diagnosis.

3.1 System architecture

The system architecture for lung cancer detection using deep learning follows a structured pipeline, starting with data acquisition and preprocessing, where CT scan images are collected, resized, normalized, and augmented to enhance model robustness. The dataset is then split into training and validation sets for model development. A Convolutional Neural Network (CNN) extracts features from the images through convolutional and pooling layers, followed by fully connected layers for classification into normal lungs, adenocarcinoma, or squamous cell carcinoma. The model undergoes training and fine-tuning, optimizing parameters to improve accuracy using loss functions and optimization techniques. After evaluation using metrics like accuracy, precision, and recall, the best-performing model is deployed via APIs or web applications for real-world use, assisting medical professionals in accurate lung cancer diagnosis.

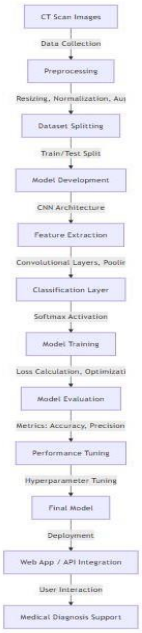

Fig1: System Architecture

3.2 Data describe

The dataset used in this project consists of CT scan

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

images of lungs, categorized into three distinct classes:

1. **Normal (lung_n):** Represents healthy lung images with no signs of cancer.
2. **Adenocarcinoma (lung_aca):** A type of lung cancer that originates in the glandular cells of the lung.
3. **Squamous Cell Carcinoma (lung_scc):** Another type of lung cancer that develops in the squamous cells lining the airways.

Each image in the dataset has been resized to 256 × 256 pixels for consistency and is normalized to enhance model training efficiency. The dataset is split into training and validation sets, ensuring a balanced distribution of classes for effective learning. Data augmentation techniques such as rotation, flipping, and contrast adjustments may be applied to improve model generalization. This structured dataset enables the deep learning model to effectively learn patterns and accurately classify lung cancer conditions.

3.3 Evaluation Metrics

To assess the performance of the deep learning model, the following evaluation metrics are used:

1. Accuracy

Accuracy measures the proportion of correctly classified images over the total number of images.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- **TP (True Positives):** Correctly predicted lung cancer cases
- **TN (True Negatives):** Correctly predicted normal lung cases
- **FP (False Positives):** Normal lungs incorrectly classified as cancer
- **FN (False Negatives):** Cancerous lungs incorrectly classified as normal

2. Precision (Positive Predictive Value)

Precision evaluates the fraction of correctly identified lung cancer cases among all cases predicted as cancer. A high precision score indicates fewer false positives

1. Recall (Sensitivity or True Positive Rate)

Recall measures the model's ability to correctly detect actual lung cancer cases.

A high recall score ensures that most lung cancer cases are correctly identified.

1. **F1-Score**

F1-Score is the harmonic mean of precision and recall, balancing both metrics.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

It provides a single score to evaluate the model's performance, especially in imbalanced datasets.

Specificity (True Negative Rate)

Specificity measures how well the model identifies non-cancerous cases.

$$Specificity = \frac{TN}{TN + FP}$$

A high specificity score indicates that the model correctly classifies normal lungs without unnecessary false alarms.

6 ROC-AUC Score (Receiver Operating Characteristic - Area Under Curve)

The ROC-AUC score evaluates the model's ability to differentiate between cancerous and non-cancerous cases across different threshold values. A score closer to **1.0** indicates better classification performance.

4.RESULTS AND DISCUSSION

This image shows the results of a cancer classification project, displaying histopathological images of tissue samples labeled with their corresponding diagnoses. The images are stained with hematoxylin and eosin (H&E), which color the nuclei of cells dark purple and other tissue components pink, highlighting structural details.The top row contains images labeled as "Squamous Cell Carcinoma" and "No Cancer". Squamous cell carcinoma is a type of skin or mucosal cancer that shows densely packed, irregularly shaped cells with prominent nuclei, while the "No Cancer" images display healthy tissue with normal cellular structures and scattered red blood cells.

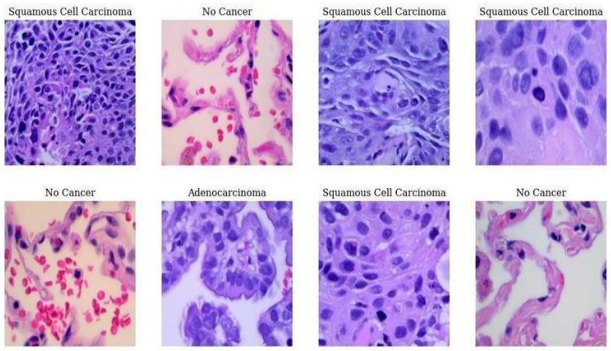

Fig 2: Histopathological Image Classification of Cancer vs. No Cancer

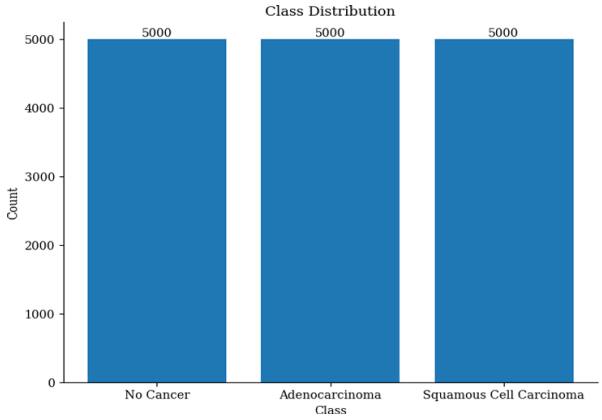

Fig 3: Class Distribution of Histopathological Images for Cancer Classification.

This image is a bar chart showing the Class Distribution of a dataset used for cancer classification. The x-axis represents the three classes: No Cancer, Adenocarcinoma, and Squamous Cell Carcinoma, while the y-axis shows the Count of samples in each class. Each class has exactly 5,000 samples, indicating a balanced dataset. A balanced class distribution is essential for training machine learning models, as it helps prevent bias towards the majority class and promotes fair learning across all categories.

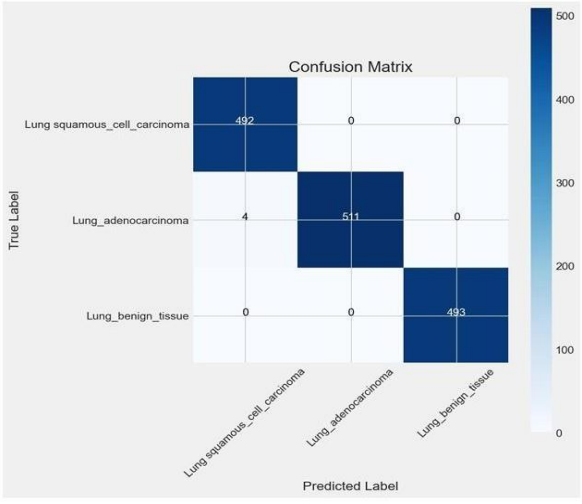

Fig 4: Confusion Matrix for Cancer Classification Model Performance

The confusion matrix presented in the image evaluates the performance of a classification model designed to distinguish between three lung conditions: lung squamous cell carcinoma, lung adenocarcinoma, and lung benign tissue. The matrix indicates that the model performs exceptionally well, as it correctly classifies the vast majority of samples with minimal misclassification`. Specifically, 492 cases of lung squamous cell carcinoma were correctly identified, with no instances of misclassification. Similarly, 493 cases of lung benign tissue were accurately classified without any errors. The model also successfully identified 511 cases of lung adenocarcinoma, though it misclassified four instances as lung squamous cell carcinoma.

5. CONCLUSION

We leveraged transfer learning with the EfficientNetB3 model to develop a highly accurate deep learning-based system for lung carcinoma detection using biopsy imaging. It demonstrate a highly effective deep learning model for lung cancer classification using histopathological images. The accuracy and loss graphs indicate rapid convergence and strong generalization, with training and validation accuracy nearing 100% and minimal loss, showcasing the model's ability to learn complex cellular features. The balanced class distribution ensures unbiased learning across all categories — Adenocarcinoma, Squamous Cell Carcinoma, and No Cancer. The model demonstrated exceptional performance, achieving near-perfect classification accuracy while maintaining low validation loss, indicating strong generalization capabilities. The confusion matrix confirmed high precision and recall, with minimal misclassifications among lung squamous cell carcinoma, lung adenocarcinoma, and benign lung tissue. The training and validation curves further validated the model's robustness, with both loss decreasing steadily and accuracy converging toward 100W. These results highlight the effectiveness of transfer learning in medical image analysis, reducing the need for extensive labelled datasets while significantly improving diagnostic accuracy. Our approach paves the way for more reliable, automated lung cancer detection, potentially aiding early diagnosis and enhancing clinical decision-making. Future work can focus on integrating additional imaging modalities, optimizing model

interpretability, and deploying the system in real-world clinical settings to further validate its effectiveness.

6 FUTURE SCOPE

The future scope of this project is vast, with multiple directions for further enhancement and real-world application. Integrating multi-modal data, such as combining biopsy images with CT scans, PET scans, and histopathology images, can improve classification accuracy and provide a more holistic understanding of lung carcinoma. Enhancing explainability and interpretability using techniques like Grad- CAM or SHAP values can help medical professionals understand the model's decision-making process, increasing trust in AI-assisted diagnostics.[1]

Deployment as a cloud-based or edge AI solution can enable real-time analysis of biopsy images in clinical settings, assisting pathologists in making faster and more precise diagnoses. Additionally, expanding the dataset to include diverse populations, different staining techniques, and various image resolutions can improve the model's generalizability, making it more robust across different clinical environments. Future research can also focus on self-supervised learning and federated learning approaches to improve privacy, scalability, and adaptability without requiring extensive labelled datasets.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Chen, M., Huang, S., Huang, Z. and Zhang, Z., 2021, September. Detection of lung cancer from pathological images using cnn model. In 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI) (pp. 352-358). IEEE

[2] .Sakr, A.S., 2022, October. Automatic Detection of Various Types of Lung Cancer Based on Histopathological Images Using a Lightweight End-to-End CNN Approach. In 2022 20th International Conference on Language Engineering (ESOLEC) (Vol. 20, pp. 141 146). IEEE.

[3] Nannapaneni, D., Saikam, V.R.S.V., Siddu, R., Challapalli, V.M. and Rachapudi, V., 2023, February. Enhanced Image-based Histopathology Lung Cancer Detection. In 2023 7th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 620-625). IEEE.

[4] Cañada, J., Cuello, E., Téllez, L., García, J.M., Velasco, F.J. and Cabrera, J., 2022, November. Assistance to lung cancer detection on histological images using Convolutional Neural Networks. In 2022 E-Health and Bioengineering Conference (EHB) (pp. 1-4). IEEE.

[5] Maheshwari, U., Kiranmayee, B.V. and Suresh, C., 2022, December. Diagnose Colon and Lung Cancer Histopathological Images Using Pre-Trained Machine Learning Model. In 2022 5th International Conference on Contemporary Computing and Informatics (IC3I) (pp. 1078- 1082). IEEE.

[6] Mohalder, R.D., Sarkar, J.P., Hossain, K.A., Paul, L. and Raihan, M., 2021, September. A deep learning based approach to predict lung cancer from histopathological images. In 2021 international conference on electronics, communications and information technology (ICECIT) (pp. 1-4). IEEE

[7] Radhakrishnan, J.K., Aravind, K.S., Nambiar, P.R. and Sampath, N., 2022, June. Detection of Non- small cell Lung Cancer using Histopathological Images by the approach of Deep Learning. In 2022 2nd International Conference on Intelligent Technologies (CONIT) (pp. 1-11). IEEE

[8] Mridha, K., Islam, M.I., Ashfaq, S., Priyok, M.A. and Barua, D., 2022, November. Deep Learning in Lung and Colon Cancer classifications. In 2022 International Conference on Advances in Computing, Communication and Materials (ICACCM) (pp. 1-6). IEEE

[9] Wild, C.P., Weiderpass, E. and Stewart, B.W., 2020. Cancer research for cancer prevention World Cancer Report. World Health Organization: Lyon, France. [10] PAHO and WHO, available from: Cancer - PAHO/WHO | Pan American Health Organization (Last accessed: 01/02/2024)