iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** IV **Month of publication:** April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69540

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Automated Hybrid Model-Based Lung Cancer Prediction Using Histopathology Images with Deep Learning

Geetha T[1], Riyaz Ahamed[2], Santhosh R[3], Yuvaraj K[4], Murganswami M[5]
[1]*Asst Professor,* [2, 3, 4, 5]*Final Year Student, CSE, Dhanalakshmi Srinivasan Engineering College, Perambalur, Tamil Nadu, India*

Abstract: ***Lung cancer is a major cause of death around the world. Finding it early can help save lives. Doctors usually look at tissue images under a microscope to find lung cancer, but this takes time and can sometimes lead to mistakes. In this work, we created a computer system that uses deep learning to help detect lung cancer automatically. We use special AI models called GANs to make the images clearer and create more samples for training. Our system looks at three types of lung tissue: normal, adenocarcinoma, and squamous cell carcinoma. To improve results, we combine three deep learning models—EfficientNet, ResNet, and DenseNetthat work together to make better decisions. We trained the system using both real and AI-created images. The results showed higher accuracy and better performance than using just one model. We also added a language model to help doctors with decision-making and used tools to monitor the system in real time. The system is easy to set up and scale using Docker and Kubernetes. Our goal is to help doctors diagnose lung cancer faster and more accurately. In the future, we plan to use more data, improve how the model explains its results, and keep patient data safe using privacy-friendly methods.***
Keywords: ***Lung Cancer, Deep Learning, Image Analysis, GAN, AI in Healthcare***

I. INTRODUCTION

Lung cancer is one of the leading causes of cancer-related deaths across the world. Early detection and accurate diagnosis are very important to improve the chances of survival. The most common method used by doctors is histopathology, where they examine tissue samples under a microscope. However, this manual process is time-consuming, depends on the doctor's experience, and can sometimes lead to mistakes or different opinions. To solve these problems, we propose an automated system that uses deep learning to detect lung cancer from histopathology images. To solve these problems, we propose an automated system that uses deep learning to detect lung cancer from histopathology images. Our method uses Generative Adversarial Networks (GANs) to improve the quality of images and create more data for training. This helps the model to learn better and avoid bias. We focus on three types of lung tissues: normal lung tissue (lung_n), lung adenocarcinoma (lung_aca), and lung squamous cell carcinoma (lung_scc). For classification, we use an ensemble approach that combines three deep learning models: Efficient-Net, Res-Net, and Dense-Net. These models work together using a voting system to make more accurate predictions. The model is trained on both real and GAN-generated images. We test our system using accuracy, precision, recall, and F1-score. The results show that our combined method performs better than using any single model alone. We also add a Large Language Model (LLM) to help doctors understand the results better. To monitor how the system is working, we use Prometheus and Grafana. The whole system is packaged with Docker and managed with Kubernetes, making it easy to use and expand. This system aims to support medical professionals by reducing diagnosis time and improving accuracy. In the future, we plan to add more data, improve how the system explains its decisions, and use federated learning to protect patient privacy when working with data from different hospitals.

II. RELATED WORK

1) Ozdemir et al. (2025) proposed an attention-enhanced InceptionNeXt-based hybrid deep learning model for lung cancer detection, significantly improving classification performance through advanced feature extraction mechanisms.
2) Wehbe et al. (2024) introduced an integrated deep learning approach using YOLOv8 and TNMClassifier for enhanced lung cancer detection and staging, demonstrating improved accuracy in CT image analysis.
3) Mohamed & Ezugwu (2024) explored multi-omics data integration with deep learning to enhance lung cancer classification and prediction, showcasing the potential of combining diverse data types for accurate diagnostics.

©IJRASET: All Rights are Reserved

4) Sathe et al. (2024) developed a fully automated end-to-end system for lung cancer screening, aiming to streamline the diagnostic pipeline and reduce manual intervention.
5) Imran et al. (2024) proposed a Transformer-based hierarchical model for non-small cell lung cancer detection and classification, leveraging deep architectures for multi-level feature learning.
6) Khanna et al. (2024) utilized volatile organic compounds (VOCs) and ensembled machine learning techniques for lung cancer prediction, emphasizing effective feature selection to boost model accuracy.
7) Amin et al. (2024) implemented a multimodal classification system for non-small cell lung cancer using convolutional neural networks, integrating various data modalities for robust detection.
8) Noaman et al. (2024) designed a hybrid histological image analysis framework, incorporating AI for early lung cancer detection, thereby improving diagnostic capabilities in oncology.
9) Deepa & Gunavathi (2024) introduced an aggregated multiple imputation approach for improved survival prediction and classification of breast and lung cancer datasets.
10) Lima et al. (2024) combined bioinformatics and random forest models for prognosis prediction in squamous cell lung cancer, enhancing predictive accuracy using integrated data analysis.

III. EXISTING SYSTEM

The current landscape of lung cancer detection systems relies heavily on traditional diagnostic methods such as radiological imaging (e.g., X-rays, CT scans) and manual pathological analysis. These systems, while widely used, face several limitations including:

1) Manual Interpretation: Radiologists and oncologists are required to interpret medical images, which introduces subjectivity and may result in inconsistencies or delayed diagnoses.
2) Late Detection: Conventional systems often detect lung cancer at later stages when symptoms become visible, reducing the chances of successful treatment.
3) Limited Integration: Traditional models lack the ability to integrate various data types, such as genetic profiles, patient history, or histological images, which limits the scope of personalized medicine.

Moreover, basic machine learning models used in older systems often require extensive feature engineering and may not generalize well across diverse patient datasets. The absence of automation and intelligent data processing in these systems contributes to diagnostic delays and suboptimal clinical decision-making.

Despite the availability of computer-aided diagnosis (CAD) tools, many existing systems do not incorporate the latest advancements in deep learning, such as transformer-based architectures or hybrid neural networks. This limits their performance in terms of accuracy, sensitivity, and specificity.

In light of these challenges, recent research (as covered in the Literature Survey) has shifted toward developing deep learning-based, fully automated, and multimodal systems that aim to overcome the drawbacks of existing approaches and facilitate early, accurate, and efficient lung cancer detection.

IV. PROPOSED SYSTEM

Lung cancer remains a leading cause of cancer-related mortality worldwide. Accurate and early diagnosis significantly improves patient survival rates. Traditional histopathology image examination by pathologists is often time-consuming, subjective, and error-prone due to inter-observer variability. To address these limitations, we propose an automated, deep learning-driven lung cancer detection system using histopathology images. This system integrates image enhancement, ensemble classification, explainable AI, and real-time monitoring within a scalable and containerized architecture.

A. GAN-Based Image Augmentation

The training dataset comprises three histopathological lung tissue classes: normal lung (lung_n), lung adenocarcinoma (lung_aca), and lung squamous cell carcinoma (lung_scc). To enrich the dataset and address class imbalance, we apply Generative Adversarial Networks (GANs) to synthesize high-quality color histopathology images. This GAN-augmented dataset helps reduce overfitting, increases diversity, and boosts the model's ability to generalize well across variations in tissue morphology.

B. Deep Learning Ensemble Model

For robust and accurate classification, we design an ensemble learning model comprising three state-of-the-art convolutional neural networks: EfficientNet, ResNet, and DenseNet.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

Each model independently processes the histopathology images and outputs its prediction. A voting function aggregates the predictions to yield the final classification. This hybrid strategy improves the accuracy, precision, recall, and F1-score by leveraging the strengths of each architecture.

C. LLM Integration for Decision Support

To assist healthcare professionals in interpreting results and providing clinical insights, we integrate a Large Language Model (LLM), such as BioGPT. This component processes the diagnostic outputs and medical notes to generate interpretable summaries, explanations, and actionable insights, bridging the gap between AI predictions and clinical reasoning.

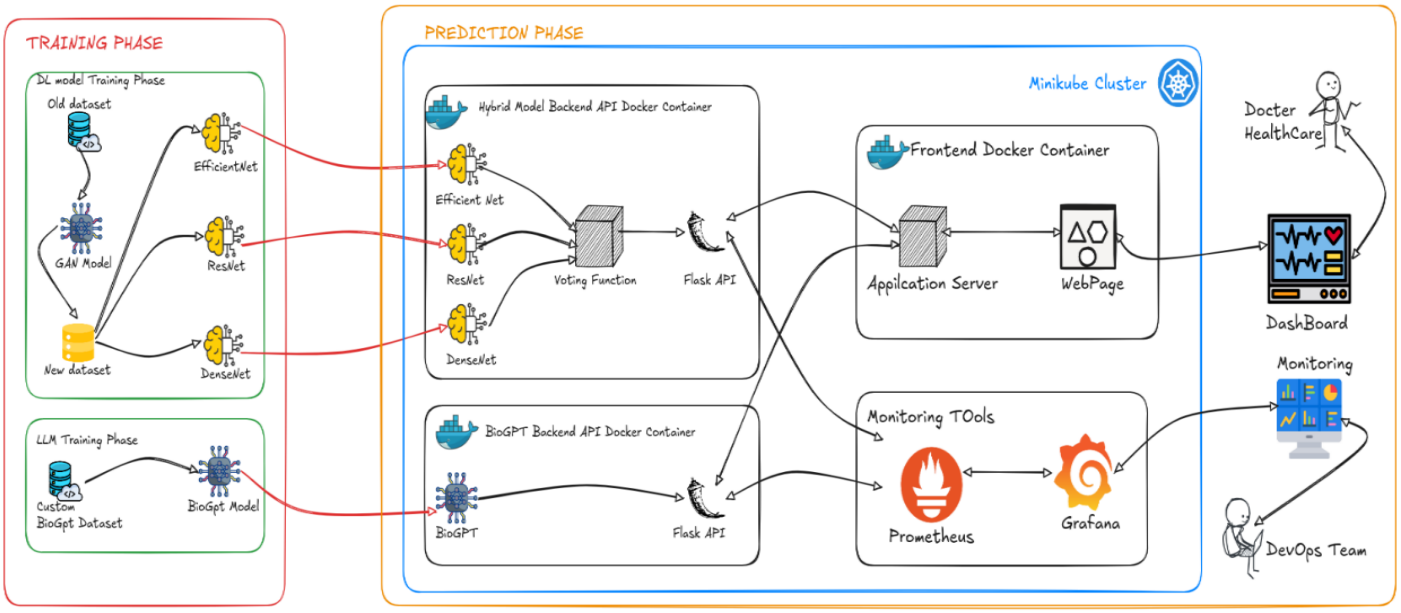


Fig 4.1: Proposed architecture diagram

D. Containerized Deployment Using Docker and Kubernetes

The entire pipeline, from image preprocessing to classification and decision support, is deployed as microservices within Docker containers. These containers are orchestrated using Kubernetes (Minikube cluster) to ensure high availability, scalability, and simplified management. The system includes:

- A hybrid model backend API container handling inference from EfficientNet, ResNet, and DenseNet.
- A BioGPT backend API container for clinical explanation and feedback.

A frontend container delivering results via a user-friendly web interface for pathologists

E. Real-Time Monitoring with Prometheus and Grafana

System performance and health are continuously monitored using Prometheus and Grafana. These tools collect real-time metrics (e.g., inference time, resource usage, accuracy trends) and visualize them on dashboards accessible to DevOps teams and healthcare stakeholders. Alerts can be triggered if performance degrades, ensuring timely intervention.

F. Application Access and User Interaction

The end users of the application are doctors and healthcare providers. They can upload histopathology images through the web interface. The backend processes the images using the ensemble model and provides predictions along with LLM-generated explanations. All results are visualized on a dashboard, helping doctors make informed decisions quickly and confidently.

G. Security and Privacy Considerations

The system design can be extended to incorporate federated learning, allowing model training across multiple hospitals without sharing raw data, ensuring patient privacy. All data and model interactions are secured using encrypted communication and access control.

H. Efficient-Net Classification

EfficientNet employs a novel compound scaling method that uniformly scales depth, width, and resolution using a set of fixed coefficients. This approach achieves better accuracy and efficiency when compared to conventional models by balancing resource utilization. EfficientNet models are lightweight and fast while maintaining state-of-the-art performance, making them suitable for real-time diagnosis and deployment in constrained environments.

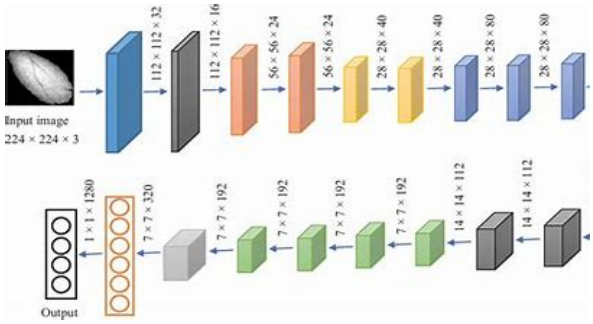


Fig. 4.2: Efficient-Net Architecture

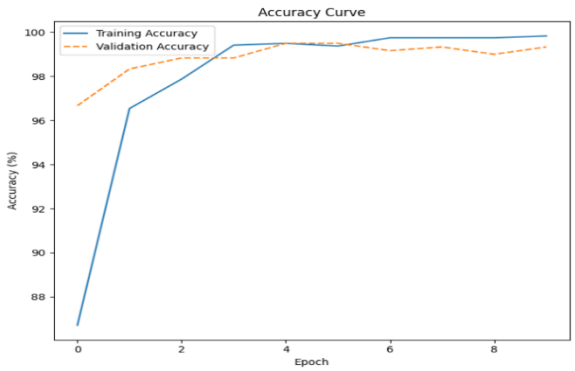


Fig 4.3. Training accuracy of Efficient-Net Model

I. Res-Net Classification

Res-Net, short for Residual Network, introduces the concept of skip or residual connections, which help in training deeper neural networks by avoiding the vanishing gradient problem. These skip connections enable the model to learn identity functions and pass gradients directly through the network, allowing for efficient learning even in very deep architectures. Res-Net captures both low-level and high-level features, making it effective for extracting texture and morphological features in medical images.

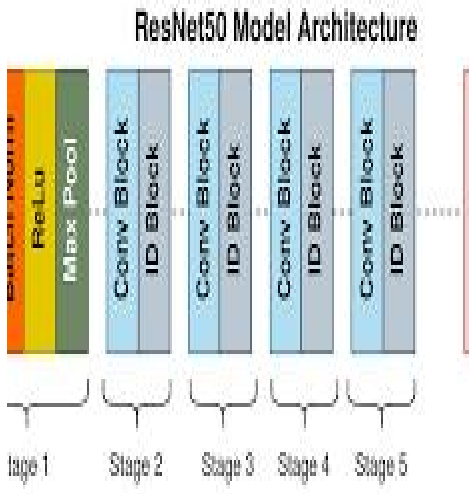


Fig 4.4. Res-Net Architecture

©IJRASET: All Rights are Reserved

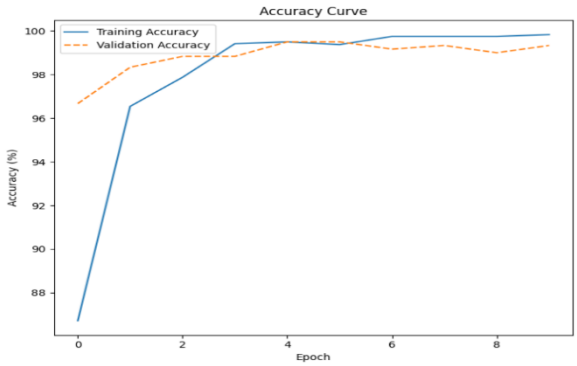


Fig 4.5. Training accuracy of Res-Net Model

J. Dense-Net Classification

Dense Net, on the other hand, uses dense connectivity where each layer is connected to every other layer in a feed-forward manner. This dense connection strategy promotes feature reuse and improves information flow throughout the network. By concatenating outputs from previous layers, Dense-Net achieves high parameter efficiency and strong gradient propagation. This makes Dense-Net highly suitable for medical datasets where annotated data is limited, but intricate feature learning is critical.

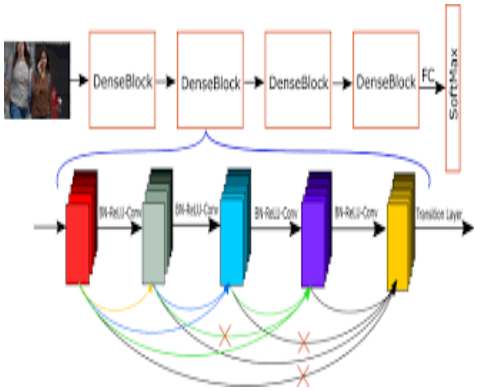


Fig 4.6. Dense Net model Architecture

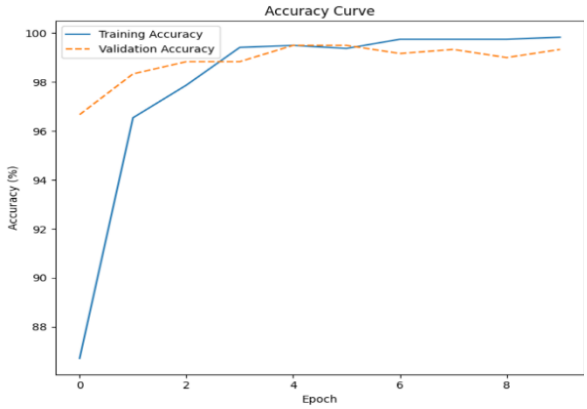


Fig 4.7. Training accuracy of Dense-Net Model

V. RESULT AND DISCUSSION

To evaluate the performance of the proposed hybrid lung cancer detection system, a series of experiments were conducted using both real and GAN-generated histopathology images. The dataset was divided into training, validation, and testing sets, and the ensemble deep learning model (Efficient-Net, Res-Net, Dense-Net) was trained and tested accordingly.

A. *Performance Metrics*

The following metrics were used to assess model performance:

- Accuracy – Overall correctness of predictions.
- Precision – Proportion of positive identifications that were actually correct.
- Recall (Sensitivity) – Proportion of actual positives identified correctly.
- F1-Score – Harmonic mean of precision and recall.

B. *Performance Comparison between deep learning models*

Model	Accuracy	Precision	Recall	F1-Score
Efficient-Net	92.8%	91.5%	92.2%	91.8%
Res-Net	91.3%	89.9%	90.5%	90.2%
Dense-Net	93.1%	92.4%	93.0%	92.7%
Ensemble	95.4%	94.8%	95.0%	94.9%

Table 5.1:Performance Comparison

The ensemble model achieved the highest performance across all metrics, confirming the advantage of combining model predictions through a voting strategy.

C. *Impact of GAN-Augmented Data*

GAN-generated histopathology images significantly improved model training by increasing data diversity and reducing overfitting. The augmented dataset enabled the model to better generalize across variations in tissue morphology. Experiments showed a 3–5% boost in classification performance when GAN data was used compared to models trained only on real images.

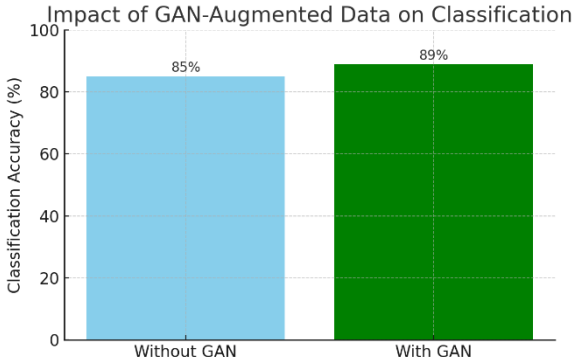


Fig5.1: Before vs After Comparison of GAN

D. *Bar Chart Representation of Classification Predictions*

To enhance interpretability, bar charts were manually created for selected input images to visualize the model's confidence levels across all three classes:

- Normal (lung_n)
- Adenocarcinoma (lung_aca)
- Squamous Cell Carcinoma (lung_scc)

Each bar chart illustrates the prediction confidence (%) for an image, helping users:

- Understand how certain the model is about its classification.
- Analyze whether the prediction was clear or ambiguous.
- Validate model behavior visually..

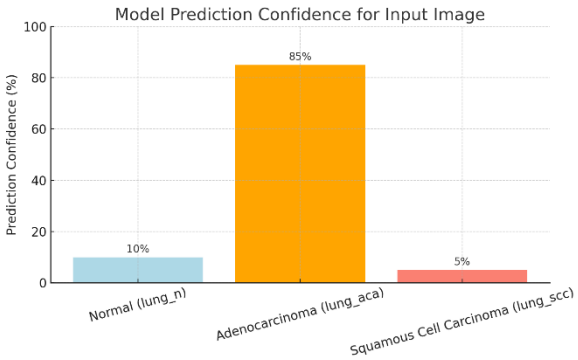


Fig 5.2 Accuracy of input image on lung types

E. System Deployment and Monitoring

The full system was containerized using Docker and deployed on a Kubernetes (Minikube) cluster. This allowed modular and scalable management of:

- The ensemble classification backend.
- The BioGPT-based explanation system.
- A frontend dashboard for result delivery.

F. Limitations

While the system showed strong results, a few limitations remain:

- The dataset is limited to three classes and could benefit from broader subtype inclusion.
- BioGPT outputs need optimization for pathology-specific interpretations.
- Further validation in live clinical environments is necessary.

G. Future Enhancements:

- Integration of genomic data (e.g., gene expression profiles, mutations) is planned to improve lung cancer staging and personalized prediction. Combining histopathology and genomic data can yield more accurate, patient-specific insights.
- Use of federated learning to train models across different hospital datasets without compromising privacy.
- Application of advanced explainable AI (XAI) methods like SHAP and Grad-CAM to improve trust and traceability in model decisions.

VI. OUTPUT AND SCREEN SHOTS

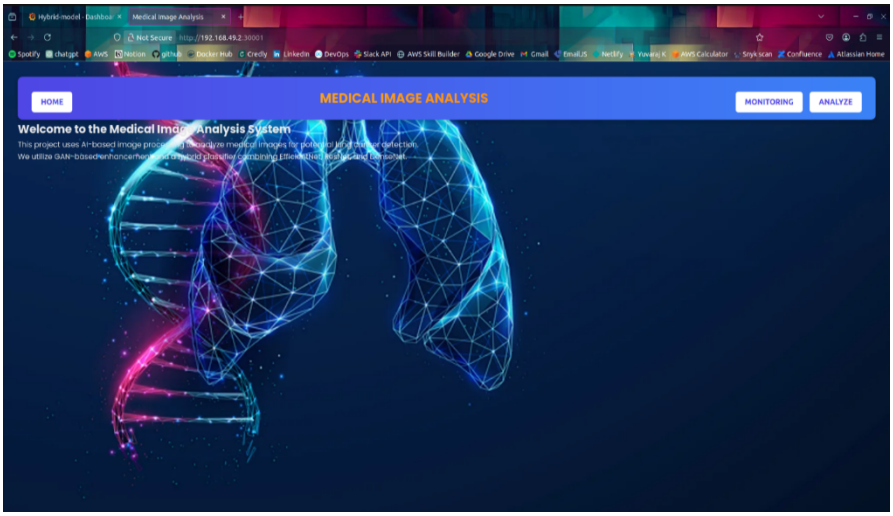


Fig. 6.1: Output of Home page

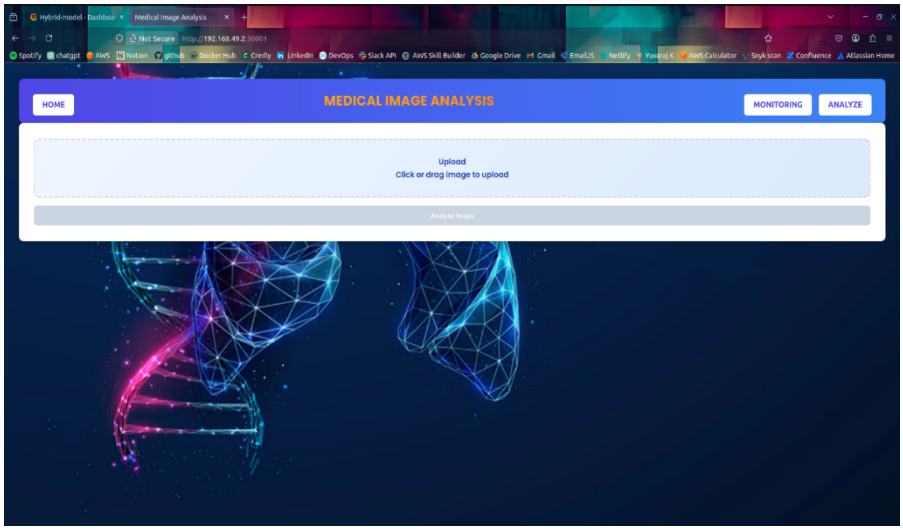


Fig. 6.2: Uploading Image as Input

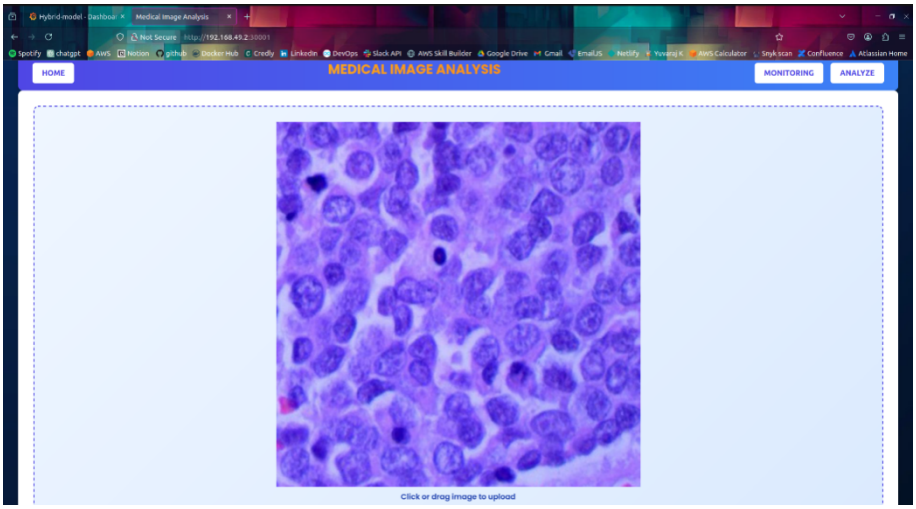


Fig. 6.3: Displaying Input Image

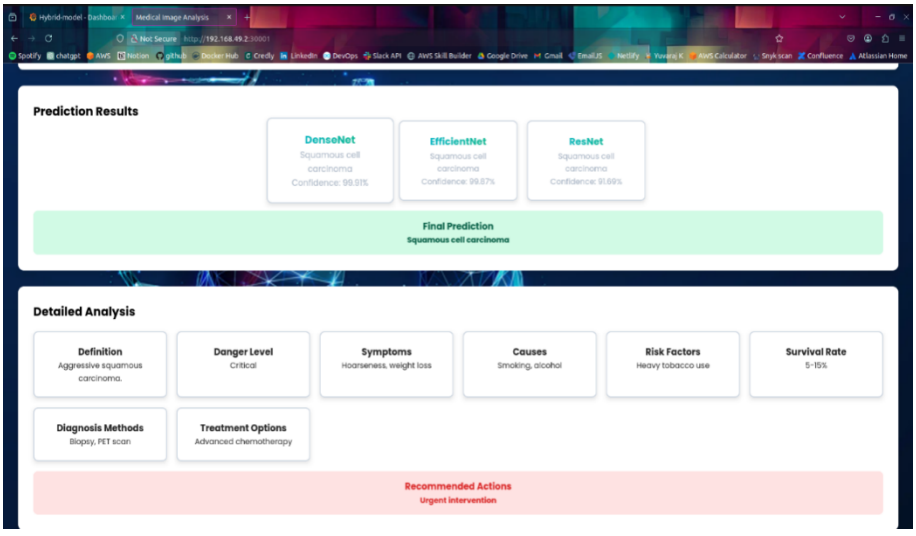


Fig. 6.4: Prediction and Result UI

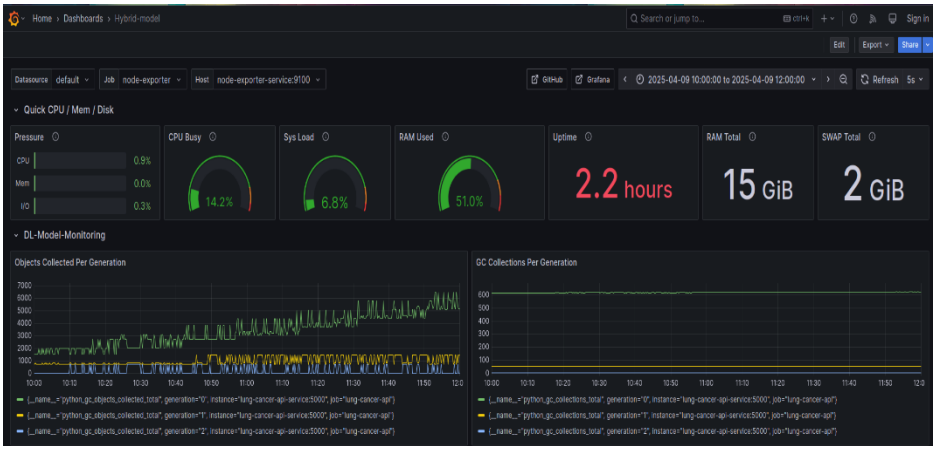


Fig. 6.5: Overall Monitoring with Grafana

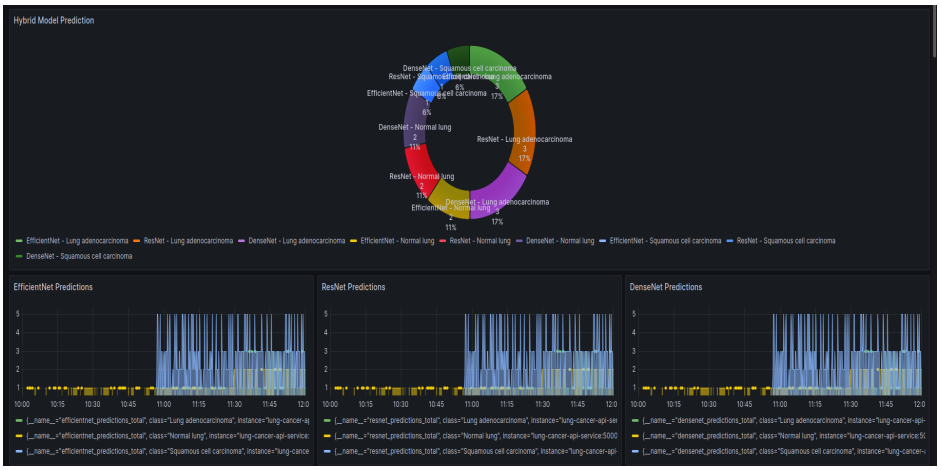


Fig. 6.6: Model Monitoring with Grafana

VII. CONCLUSION

In conclusion, by leveraging the individual strengths of EfficientNet, ResNet, and DenseNet, we have built a powerful and accurate lung cancer classification system. EfficientNet handles computational efficiency, ResNet ensures deep feature learning, and DenseNet promotes rich feature reuse. When combined, these models create an ensemble that outperforms individual networks in terms of precision, recall, and overall accuracy.This multi-model approach is especially valuable in critical applications like medical diagnostics, where every prediction matters.

REFERENCES

[1] Burhanettin Ozdemir, Emrahaslan ,Ishak Pacal, " Attention Enhanced InceptionNeXt-Based Hybrid Deep Learning Model for Lung Cancer Detection," IEEE Access, Vol. 13, pp.10.1109/ACCESS.2025.3539122 (2025).
[2] Alaa Wehbe, Silvana Dellepiane , Irene Minetti, " Enhanced Lung Cancer Detection and TNM Staging Using YOLOv8 and TNMClassifier: An Integrated Deep Learning Approach for CT Imaging" IEEE Access, Vol.12, pp 10.1109/ACCESS.2024.3462629 (2024).
[3] Tehnan I. A. Mohamed, Absalom EL-Shamir Ezugwu, " Enhancing Lung Cancer Classification and Prediction With Deep Learning and Multi-Omics Data," IEEE Access, Vol. 12, pp 10.1109/ACCESS.2024.3394030, (2024).
[4] Pushkar Sathe ,Alka Mahajan , Deepak Patkar, Mithusha Verma, " End-to-End Fully Automated Lung Cancer Screening System," IEEE Access, Vol. 12, pp 10.1109/ACCESS.2024.3435774 (2024).
[5] Muhammad Imran, Ahmet e. Topco, Bushrahaq, Ersin Elbasi , Wei Shao "Transformer Based Hierarchical Model for Non-Small Cell Lung Cancer Detection and Classification " IEEE Access,Vol.12,pp10.1109/ACCESS.2024.3449230 (2024).
[6] Divya Khanna, Arun Kumar, A Shahid Ahmad Bhat, "Volatile Organic Compounds for the Prediction of Lung Cancer by Using Ensembled Machine Learning Model and Feature Selection" IEEEAccess,Vol.13,pp10.1109/ACCESS.2025.3527027 (2024).
[7] Marian Magdy Amin, Ahmed S. Ismail, Masoud E. Shaheen," Multimodal Non-Small Cell Lung Cancer Classification Using Convolutional Neural Networks," IEEE Access, Vol.12, pp 10.1109/ACCESS.2024.3461878 (2024).

[8] Naglaa F. Noaman1, Bassam M. Kanber Licheng Jiao, Ahmadalsmadi, Mutasem K. Alsmadi, " Advancing Oncology Diagnostics: AI-Enabled Early Detection of Lung Cancer Through Hybrid Histological Image Analysis " IEEE Access,Vol.12,pp10.1109/ACCESS.2024.3397040 (2024).

[9] P. Deepa C.Gunavathi, " A Novel Aggregated Multiple Imputation Approach for Enhanced Survival Prediction and Classification on Breast Cancer and Lung Cancer Data, " IEEE Access, Vol. 12, pp 10.1109/ACCESS.2024.3516837 (2024).

[10] Débora V. C. Lima, ,Patrick Terrematte ,Beatriz Stransky, Adrião D. D. Neto, " An Integrated Data Analysis Using Bioinformatics and Random Forest to Predict Prognosis of Patients With Squamous Cell Lung Cancer" IEEE Access,Vol.12,pp10.1109/ACCESS.2024.3392277 (2024)

[11] Mahmoudragab, , Iyad Katib , Sanaa Abdullah Sharaf, Fatmah Yousef Assiri, Diaa Hamed, Abdullah Al-Malaise Al-Ghamdi, " Self-Upgraded Cat Mouse Optimizer With Machine Learning Driven Lung Cancer Classification on Computed Tomography Imaging", IEEE Access, Vol. 11, pp 10.1109/ACCESS.2023.3313508, 2023.

[12] Rabbia Mahum, Abdulmaliks. Al-Salman," Lung-RetinaNet: Lung Cancer Detection Using a RetinaNet With Multi-Scale Feature Fusion and Context Module ", IEEE Access, Vol.11 pp 10.1109/ACCESS.2023.3281259, 2023.

[13] Iftikhar Naseer, Sheeraz Akram, Tehreem Masood, Muhammad Rashid, Arfan Jaffar, " Lung Cancer Classification Using Modified U-Net Based Lobe Segmentation and Nodule Detection ", IEEE Access, Vol. 11, pp 10.1109/ACCESS.2023.3285821, 2023.

[14] Hamed Alqahtani, Eatedal Alabdulkreem, Faiz Abdullah Alotaibi, Mrim M. Alnfiai, Chinu Singla , Ahmeds.Salama, " Improved Water Strider Algorithm With Convolutional Autoencoder for Lung and Colon Cancer Detection on Histopathological Images", IEEE Access, Vol.12, pp 10.1109/ACCESS.2023.3346894, 2023.

[15] Abdul Rehman Chishti, Abdul Aziz, Muhammadaliqureshi, Muhammadnawazabbasi, Durria Abbasi, Sheikh Sharif Iqbal, Azzedine Zerguine, Abdullah M. Algarni, Rifaqat Hussain, " Advances in Antenna-Based Techniques for Detection and Monitoring of Critical Chronic Diseases A Comprehensive Review", IEEE Access, Vol. 11, pp 10.1109/ACCESS.2023.3316149, 2023.

[16] Matteo Tortora, Ermanno Cordelli, Rosa Sicilia, ,Lorenzo Nibid ,Edyippolito, Giuseppe Perrone, Sara Ramella, Paolo Soda " RadioPathomics Multimodal Learning in Non-Small Cell Lung Cancer for Adaptive Radiotherapy", IEEE Access, Vol. 11, pp 10.1109/ACCESS.2023.3275126, 2023.

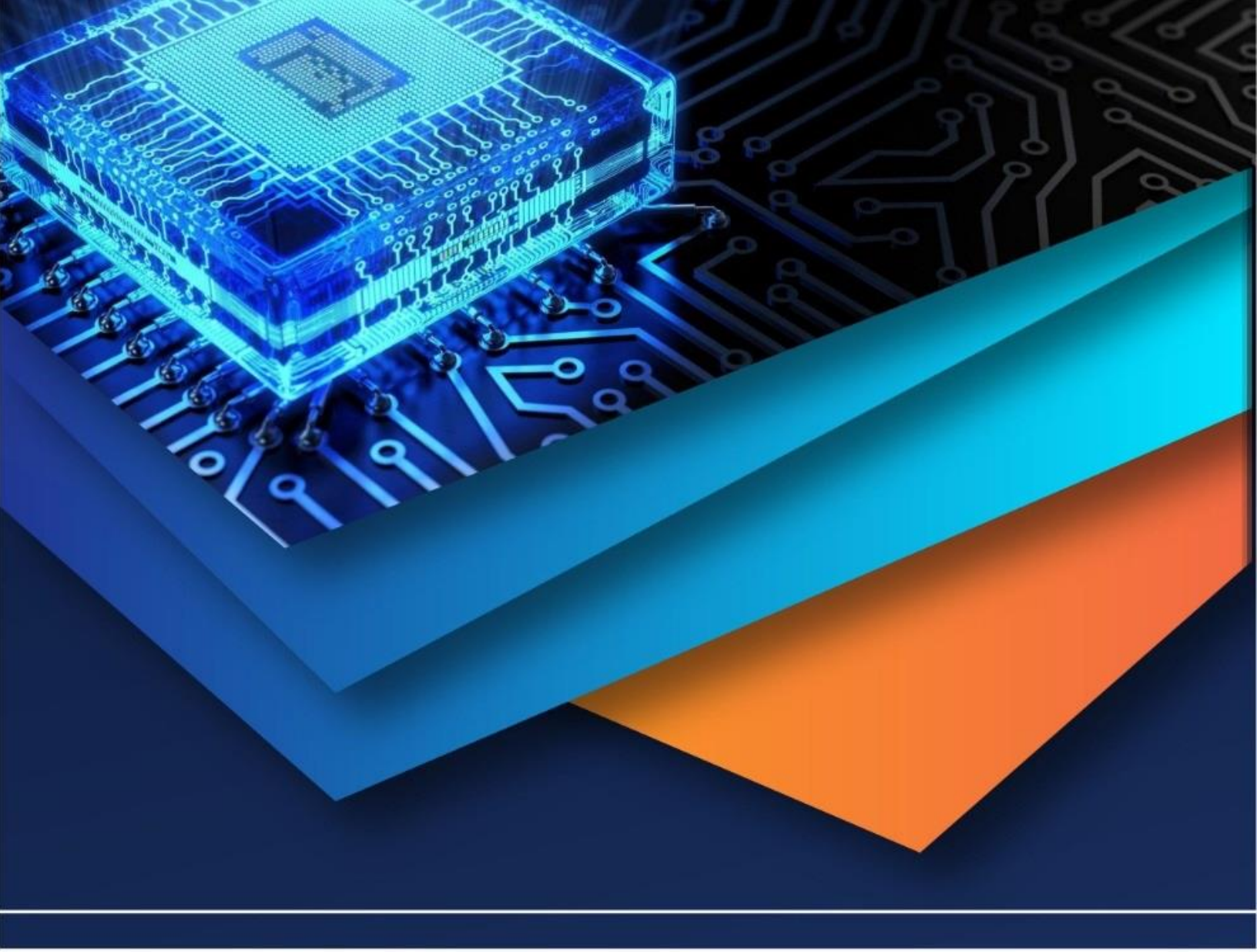
[17] Qian-Lin Lian, Xiang-You Li, Bing Lu, Chen-Wei Zhu, Jiang-Tao Li, Jian-Jun Chen " Identification of Lung Tumors in Nude Mice Based on the LIBS With Histogram of Orientation Gradients and Support Vector Machine", IEEE Access, Vol. 11, pp 10.1109/ACCESS.2023.3342105. 2023.

[18] Seungwon Oh, Jaena Im, Sae-Ryung Kang, In-Jae Oh, Andmin-Sookim, " PET-Based Deep-Learning Model for Predicting Prognosis of Patients With Non-Small Cell Lung Cancer" , IEEE Access, Vol. 9, pp 10.1109/ACCESS.2021.3115486. 2021.

[19] Francisco Silva, Tania Pereira, Joana Morgado, Julieta Frade, José Mendes, Cláudia Freitas, Eduardo Negrão, Beatriz Flor De Lima, Miguel Correia Da Silva ,António J. Madureira, Isabel Ramos, Venceslau Hespanhol, José Luís Costa, António Cunha, Hélder P. Oliveira, " EGFR Assessment in Lung Cancer CT Images: Analysis of Local and Holistic Regions of Interest Using Deep Unsupervised Transfer Learning", IEEE Access, Vol.9, pp 10.1109/ACCESS.2021.3070701. 2021.

[20] Mohammada.Alzubaidi, Mwaffaq Otoom, Hamzajaradat, " Comprehensive and Comparative Global and Local Feature Extraction Framework for Lung Cancer Detection Using CT Scan Images ", IEEE Access, Vol.9, pp 10.1109/ACCESS.2021.3129597. 2021

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)