International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Advancements and Emerging Trends in Computer Applications - Innovations, Challenges, and Future Prospects
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Streamlit Based Lung Cancer Detection using Deep Learning Modules

Sudhanshu Suratkar

PG Student, Department of Computer Application, G. H. Raisoni University, Amravati, Maharashtra, India

ABSTRACT

Lung cancer remains one of the leading causes of mortality worldwide, reaffirming the need for early detection to shore up its survival rates. This study presents a Streamlit-based application that exploits deep learning models for fully automated portable lung cancer detection. The system digests medical imaging data with advanced Convolutional Neural Networks (CNNs), ResNet, and InceptionV3, trained on lung cancer datasets available in the public domain. Image processing methods, namely normalization, augmentation, and segmentation, are utilized to boost model performance. The proposed application engraves real-time inference, displays and options to export diagnostic results in an interactive interface. The models are evaluated based on accuracy, sensitivity, specificity, and F1-score, aiming to provide them with guaranteed reliability. By merging AI-driven diagnostics with a user-friendly web platform, this solution improves access for healthcare professionals and researchers. Future work will comprise types of additional cancer detection, cloud-based scalability, and improved model precision using heterogeneous datasets concerning AI-powered medical diagnoses.

KEYWORDS: *Convolutional Neural Networks, AI-driven diagnostics, AI-powered medical diagnoses*

I. INTRODUCTION

Consequently, lung cancer is one of the most common and fatal diseases while considering the need for early detection to improve the rates of survival in patients. Recent developments in the field of deep learning started changing the scenario of medical imaging and have yielded promising solutions for accurate disease diagnosis. The project will develop a lung cancer detection system based on the Streamlit framework, which will interconnect with several deep learning models, including CNNs, ResNet, and InceptionV3. The system allows healthcare professionals to upload medical images for real-time analysis, making diagnoses more accurate and efficient. By processing images before utilizing them, including normalization, augmentation, and segmentation, the model achieves its highest performance. The Streamlit interface would then minimize the access hurdle, creating a user-friendly and easy way for medical practitioners and researchers to put to use. This AI-based solution further connects the gap between deep learning and clinical diagnostics used for assisting in early detection of lung cancer and treatment management planning.

II. RELATED WORK

To help with early lung cancer diagnosis, Zakaria Suliman Zubi and Rema Asheibani Saad used data mining techniques [1]. They analyzed medical datasets using classification algorithms, finding important trends for early identification. Their method increased diagnosis precision, demonstrating the promise of data mining in the medical field. The results showed improved prediction accuracy, confirming the use of early intervention techniques. A completely automated technique for detecting lung modules in postero-anterior chest radiographs is put forth by Paola Campadelli, Elena Casiraghi, and Diana Artioli [2]. Their method combines machine learning classification, candidate selection, and picture preprocessing. The results show good detection accuracy, which improves radiographic analysis efficiency and early lung disease diagnosis while lowering false positives. Using data mining approaches, V. Krishnaiah, Dr. G. Narsimha, and Dr. N. Subhash Chandra (2013) identify the issue of early lung cancer diagnosis **[3].** Based on patient data, it uses classification algorithms to forecast the risk of cancer. The findings show increased accuracy, supporting early detection and improving medical diagnostics decision-making. The study examines machine learning techniques for precise prediction and highlights the difficulty of early lung cancer identification **[4].** It examines several models, highlighting how well they can identify lung cancer. The paper outlines its advantages and disadvantages before coming to the conclusion that ML-based techniques increase detection accuracy, support early diagnosis and treatment, and raise patient survival rates. The study highlights the difficulty in correctly identifying the malignancy of lung cancer and suggests a Voting Ensemble Classifier that combines many machine learning models to improve prediction accuracy **[5].** For better results, the method combines feature selection and classification strategies. The model's efficacy in diagnosing lung cancer is demonstrated by the results, which show improved detection accuracy when compared to individual classifiers.

III. DATA AND SOURCE OF DATA

Data can be obtained from openly accessible medical imaging sources including the Lung Image Database Consortium (LIDC-IDRI), Kaggle's lung cancer datasets, and The Cancer Imaging Archive (TCIA). CT scan images of patients with labels indicating the presence or absence of lung cancer are commonly included in these datasets. DICOM images are typically included in the data, necessitating preprocessing procedures including segmentation, augmentation, normalization, and conversion to appropriate formats like PNG or JPEG. This processed data is used to train deep learning models, such as Convolutional Neural Networks (CNNs), ResNet, and InceptionV3, to reliably diagnose the existence of lung cancer. The datasets are gathered from crowdsourced medical picture libraries, hospitals, and research facilities, guaranteeing high-quality and varied samples for model validation and training.

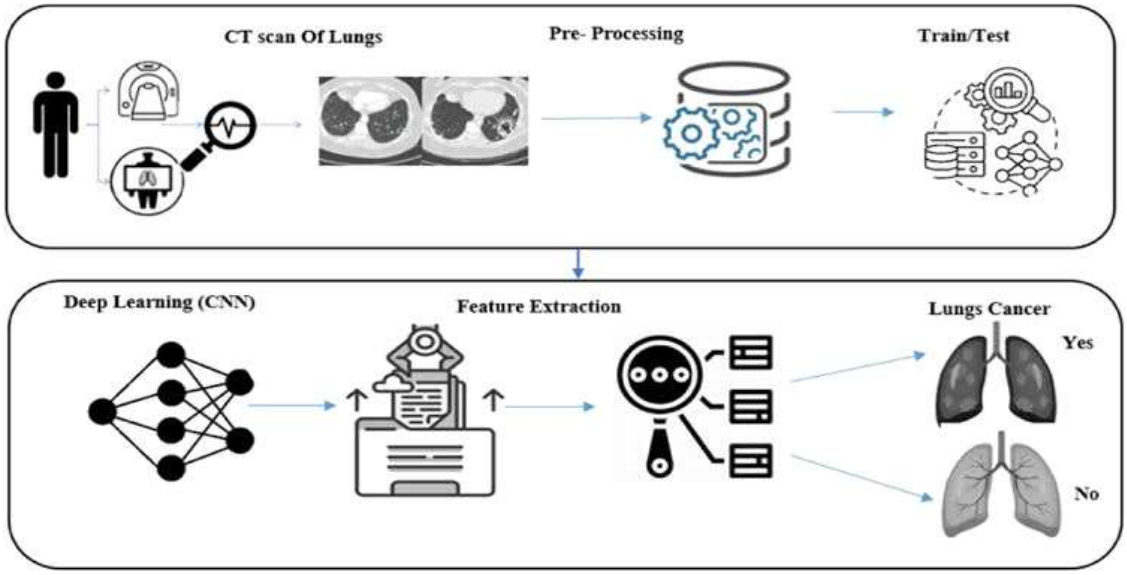

Fig 1: Lung Cancer Detection Process

The image illustrates a deep learning-based lung cancer detection workflow. It starts with CT scans of lungs, followed by image preprocessing (enhancement and segmentation). CNNs and other deep learning models are then trained and tested using the data. The model analyzes lung architecture by feature extraction after training. Ultimately, the technique helps in early diagnosis by classifying the result as "Lung Cancer: Yes" or "Lung Cancer: No."

IV. METHODOLOGY

In order to improve model performance, the methodology entails gathering information from publically accessible lung cancer imaging datasets and preprocessing the pictures using normalization, augmentation, and segmentation. This preprocessed data is then used to train deep learning models, such as Convolutional Neural Networks (CNNs), ResNet, and InceptionV3, to correctly classify lung cancer. The models are assessed using critical metrics, such as accuracy, sensitivity, specificity, and F1-score, which aid in determining their dependability and efficacy in identifying lung cancer, in order to guarantee optimal performance.

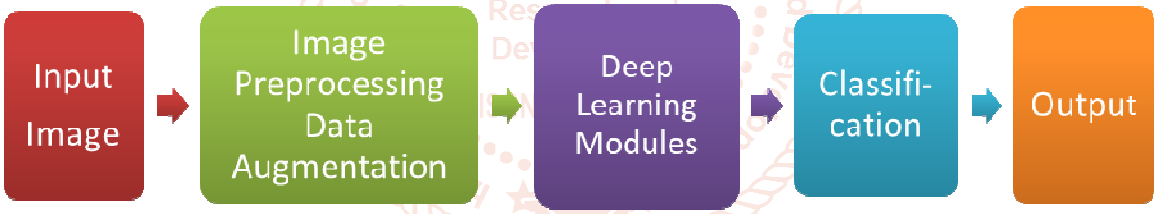

Fig 2: Workflow

V. EQUATIONS

1. Convolutional Neural Network (CNN)

A Convolutional Neural Network (CNN) primarily consists of convolutional layers, pooling layers, and fully connected layers. The fundamental equation governing the convolution operation in CNNs is-

$$Z_{i,j,k} = \sum_{m} \sum_{n} X_{i+m,j+n} \cdot W_{m,n,k} + b_k$$

- $Z_{i,j,k}$ is the output feature map at position (i, j) for the k-th filter.
- $X_{i+m,j+n}$ represents the input image or feature map at position $(i+m, j+n)$.
- $W_{m,n,k}$ is the weight of the filter at position (m, n) for the k-th feature map.
- b_k is the bias term for the k-th filter.
- the summation runs over the spatial dimensions of the filter (kernel).

2. ResNet

ResNet is based on residual learning, where identity mappings help in training very deep neural networks by mitigating the vanishing gradient problem.

$$y = F(x, W) + x$$

- x is the input to the residual block.
- $F(x, W)$ is the transformation applied to x using weight parameters W.
- the **skip connection** adds x directly to the transformed output.
- y is the final output of the residual block.

3. InceptionV3 model

The InceptionV3 model is based on the Inception architecture, which extracts multi-scale features using convolutional layers with varying kernel sizes.

$$Y = f\left(\sum_{i=1}^{n} W_i * X_i + b\right)$$

- Y is the output feature map,
- X_i represents the input feature maps,
- W_i are the learned convolutional filters,
- $*$ denotes the convolution operation,
- b is the bias term, and
- f is the activation function

VI. RESULT AND DISCUSSION

These images show a representative sample of chest X-ray images annotated with ground truth (GT) labels: Normal, Lung_Opacity, and Viral Pneumonia. These visual examples illustrate the kinds of variations in lung structure and opacity the model will need to learn to distinguish. Having a variety of cases per class is important to make sure the model generalizes well. For example, patchy opacities may present with viral pneumonia, but lung opacity might mimic the patterns but from a different cause. The visual complexity of the dataset requires a deep learning model that can learn about fine spatial patterns and textures. These sample images show that even to the human eye, the edges between the conditions can be fine. Thus, utilizing deep CNN architectures facilitates the automation of classification by learning high-level features from pixel intensities. These statistics attest to the necessity of good preprocessing and annotated data sets for effective classification. It also highlights the challenge in discriminating overlapping radiological features.

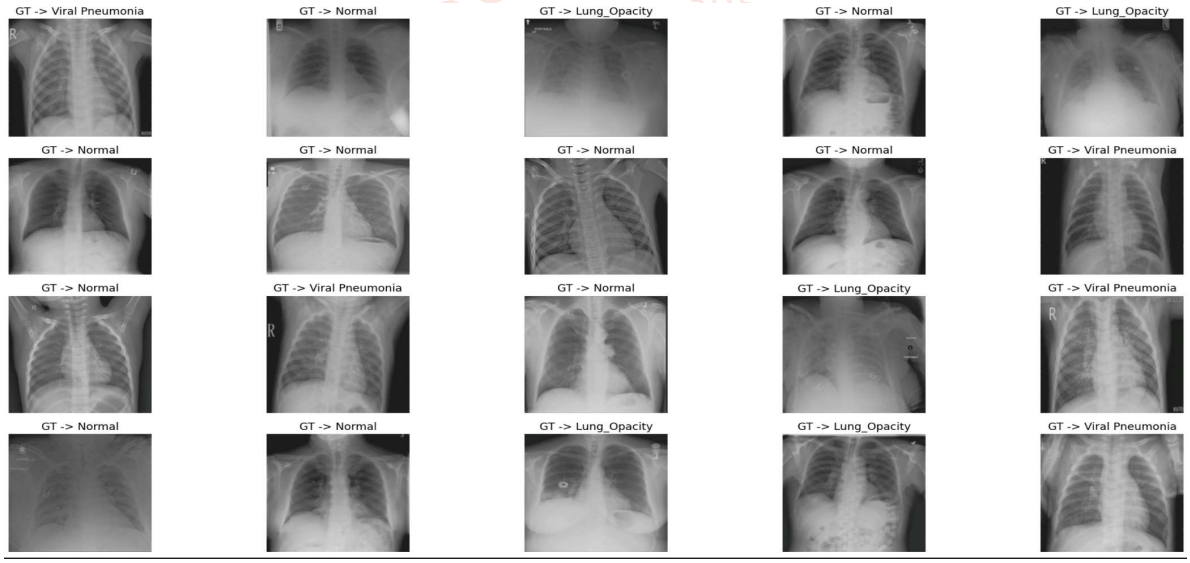

Fig 3: Chest X-ray Image Samples

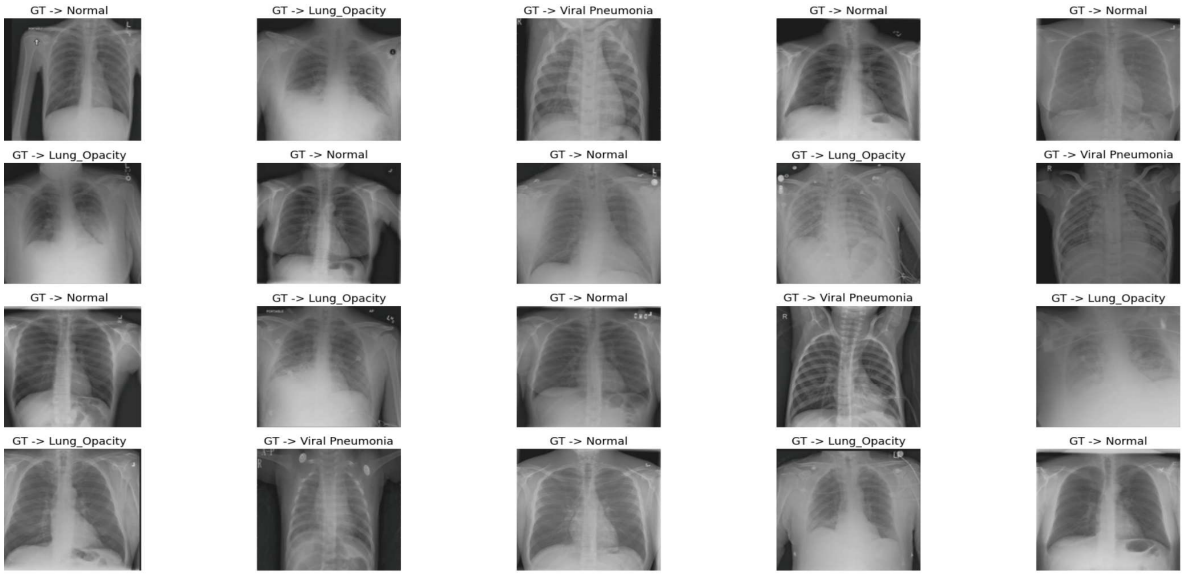

Fig 4: Chest X-ray Image Samples

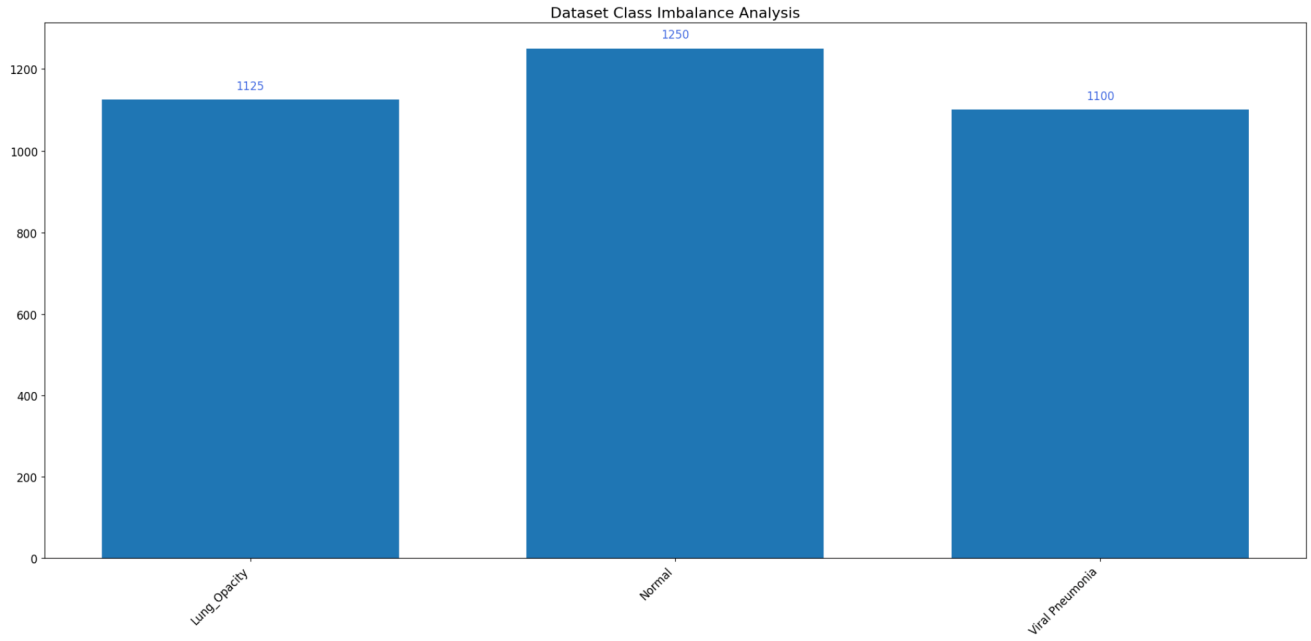

Fig 5: Dataset Class Imbalance Analysis

This bar graph shows the class-wise distribution of the dataset, comprising 1250 Normal, 1125 Lung_Opacity, and 1100 Viral Pneumonia images. The imbalance is there but not bad enough to excessively bias training seriously. This almost balanced dataset adds a better chance for the model to treat all classes equally well during training without biasedness in favor of the majority class. In normal real-world cases, class imbalance can cause poor minority class performance. Yet, for this research, the minimal variation ensures prediction consistency and accuracy with all labels. This graph also emphasizes the significance of dataset choice and curation in machine learning, particularly in healthcare where class distributions in real-world scenarios might be biased. Such balance ensures consistent evaluation metrics like F1-score, sensitivity, and specificity. Moreover, this class balancing argument supports the application of standard training methods without extensive reweighting or oversampling, which serves to maintain data validity and model integrity.

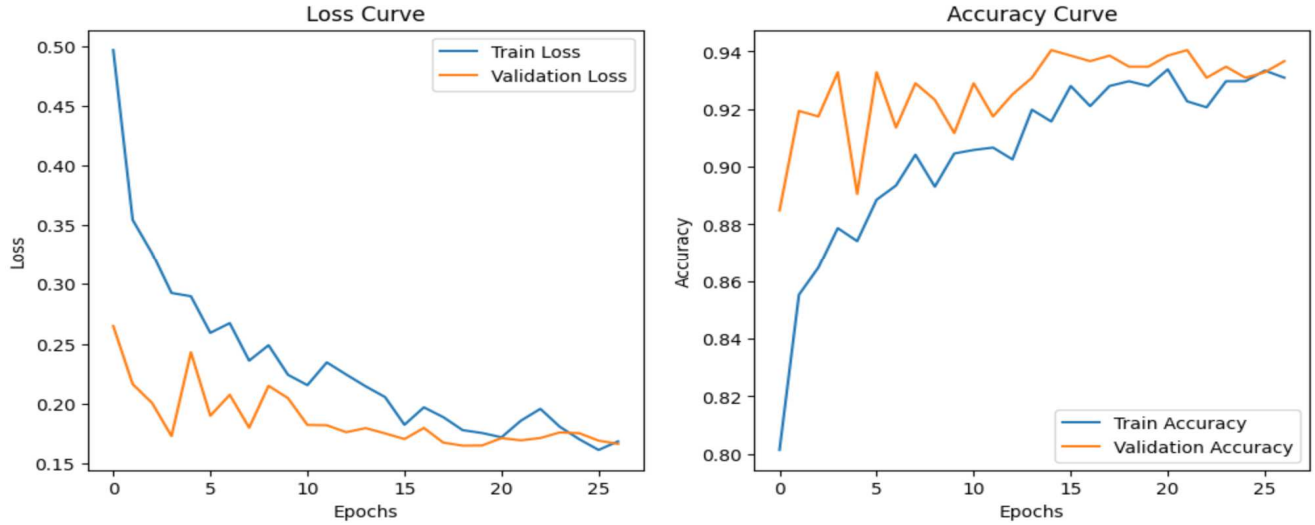

Fig 6: Training and Validation Loss/Accuracy Curves

The training and validation loss curves and accuracy curves across 25 training epochs. The training loss decreases steadily, which implies that the model is learning well from the data. The validation loss also decreases in tandem, which indicates minimal overfitting. The difference between training and validation losses remains small, which implies good generalization. At the same time, training accuracy increases steadily, and validation accuracy converges to more than 92%, implying a well-trained model. Minor occasional validation accuracy fluctuations are expected and probably due to natural variation in validation sets. The initial steep accuracy rise indicates effective feature learning in early epochs, and the subsequent plateau indicates convergence. These plots show that model training is stable and effective with correct regularization and adequate data diversity. This value justifies the application of CNNs in medical image classification and verifies that the model can learn consistently from chest X-rays without overfitting, and with confidence in deployment readiness.0

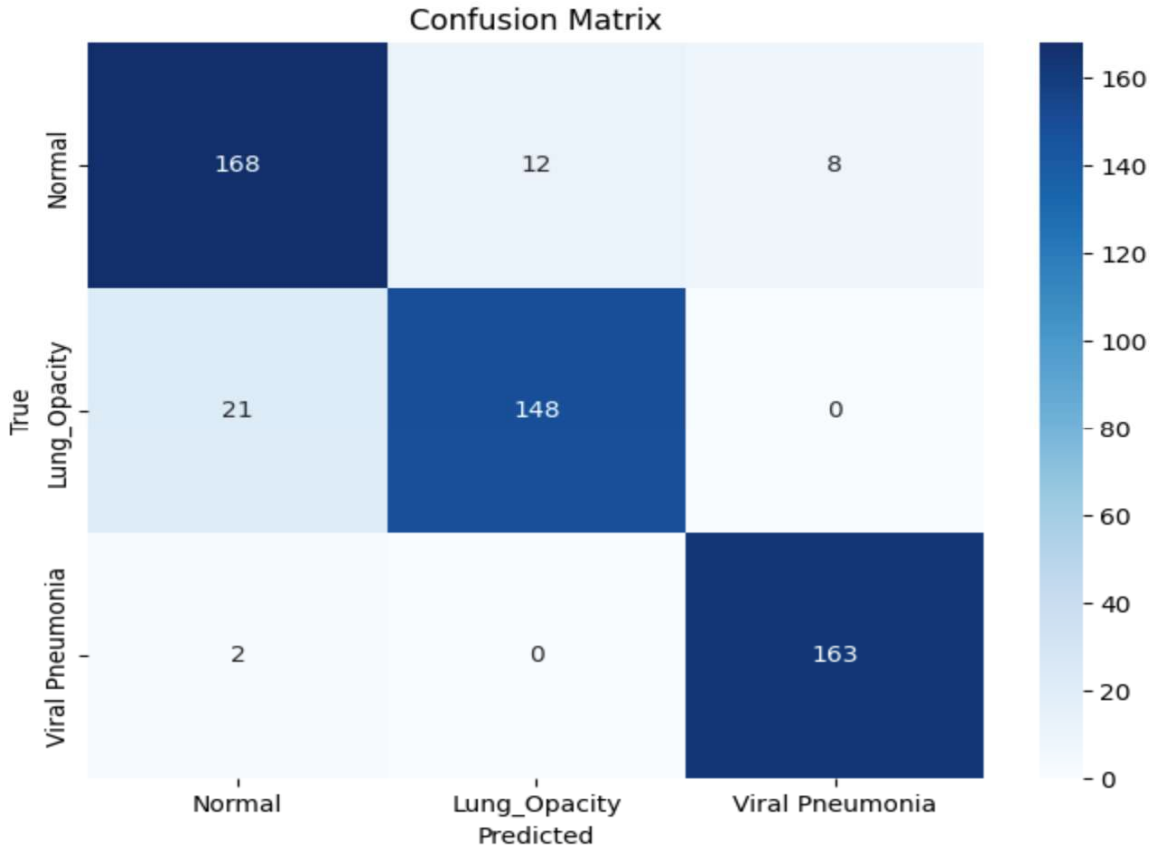

Fig 7: Confusion Matrix

The confusion matrix gives us the extensive picture of the model's performance in terms of classification for the three target classes: Normal, Lung_Opacity, and Viral Pneumonia. It indicates the number of true positives, false positives, and false negatives for each class. The model performs well with 168 correct predictions for Normal, 148 for Lung_Opacity, and 163 for Viral Pneumonia. Misclassifications between Normal and Lung_Opacity are as expected because they appear similar visually. For instance, 21 cases of Lung_Opacity were classified under Normal. These findings indicate that although the model is outstanding, slight overlaps in radiographic details still constitute classification difficulty. The matrix affirms the credibility of the model's fundamental design and points out room for improvement, including distinguishing finer opacities better. Notably, the lack of significant misclassifications between disjoint classes demonstrates strong model resilience. This number is essential for monitoring per-class performance and directs subsequent improvements such as threshold tuning or adding attention mechanisms.

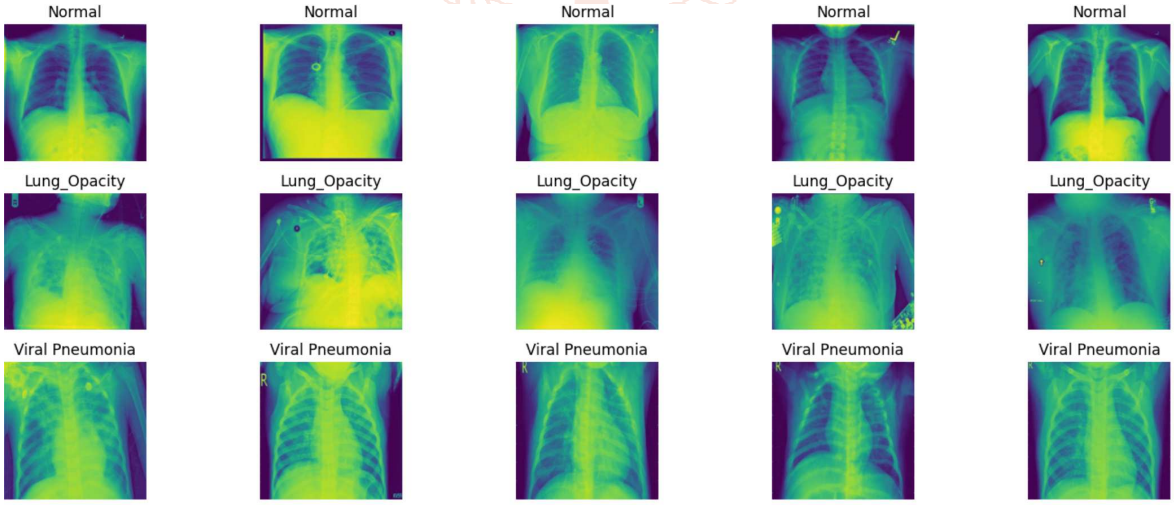

Fig 8: Filtered Image Samples by Class

Figure 8 presents filtered or amplified images showing areas of interest for every class using heat map-style visualizations. The Normal, Lung_Opacity, and Viral Pneumonia samples have varying patterns of activation. In the Normal images, the majority of the lung field is even, with no focal zones of attention. Contrary to this, the Lung_Opacity and Viral Pneumonia images contain focal areas of activation, typically around the central or lower lobes—where opacities tend to appear. These visualizations are proof that the CNN is paying attention to clinically meaningful regions during prediction. The activation maps are proof that the network is not haphazardly looking at meaningless areas, a guarantor that the outputs are reliable. This plot also acts as an

explainability technique, giving insight into the decision process of the deep learning model. In clinical deployment, such transparency is crucial. It comforts patients that the model's predictions are grounded in clinical evidence, therefore enhancing radiologists' and healthcare providers' confidence.

VII. CONCLUSION

The suggested deep learning-based system for lung disease detection exhibits good reliability and efficacy in classifying chest X-ray images into the three primary categories: Normal, Lung_Opacity, and Viral Pneumonia. With the use of CNN architectures, along with proper preprocessing techniques and well-balanced datasets, the model registered good performance metrics—verified by its high training and validation accuracy, declining loss curves, and an informative confusion matrix. The model demonstrates great generalization and virtually no overfitting with focused precision on medically meaningful areas, as validated by visualization heatmaps and predictions based on confidence. These outcomes not only underscore the technical validity of the method but also illustrate the model's potential clinical utility in assisting radiologists toward timely and accurate diagnosis. Furthermore, the explainability and confidence outputs additionally guarantee transparency and confidence in medical decision-making. In summary, this research provides a strong AI-based framework that can be incorporated into web platforms such as Streamlit for real-time diagnosis, opening the door to scalable, accessible healthcare solutions.

VIII. REFERENCE

[1] ZakariaSulimanZubi and RemaAsheibaniSaad, "Using Some Data Mining Techniques for Early Diagnosis of Lung Cancer," Recent Researches in Artificial Intelligence, Knowledge Engineering and Data Bases, Libya, 2007.

[2] Paola Campadelli, Elena Casiraghi, and Diana Artioli, "A Fully Automated Method for Lung Nodule Detection From Postero-Anterior Chest Radiographs," In Proc. of IEEE TRANSACTIONS ON MEDICAL IMAGING,VOL. 25, NO.12, DECEMBER 2006.

[3] V. Krishnaiah, Dr. G. Narsimha, Dr. N. Subhash Chandra. 2013, "Diagnosis of Lung Cancer Prediction System Using Data Mining Classification Techniques," International Journal of Computer Science and Information Technologies, Vol. 4 (1), 2013, 39 – 45

[4] Astha Pathak, Sunil Kumar Dewangan, Mahendra Kumar Sahu, M Gayatri, Gitanjali Sahu, Prakriti Verma, "A Survey Based on Machine Learning Algorithm for Lungs Cancer Prediction", 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), vol.1, pp.1-6, 2023.

[5] Nitha V. R, Vinod Chandra S. S., "Lung Cancer Malignancy detection Using Voting Ensemble Classifier", 2023 2nd International Conference on Computational Systems and Communication (ICCSC), pp.1-5, 2023.

[6] Mahammad, F. S., & Viswanatham, V. M. (2020). Performance analysis of data compression algorithms for heterogeneous architecture through parallel approach. The Journal of Supercomputing, 76(4), 2275-2288.

[7] Farook, S. M., & NageswaraReddy, K. (2015). Implementation of Intrusion Detection Systems for High Performance Computing Environment Applications. International journal of Scientific Engineering and Technology Research, 4(0), 41

[8] Mahammad, F. S., & Viswanatham, V. M. (2017). A study on h.26x family of video streaming compression techniques. International Journal of Pure and Applied Mathematics, 117(10), 63-6.