Comprehensive Lung Disease Detection Using Deep Learning Models and Hybrid Chest X-ray Data with Explainable AI

Shuvashis Sarker*, Shamim Rahim Refat†, Faika Fairuj Preotee‡, Tanvir Rouf Shawon§ Raihan Tanvir¶,

*†‡§¶Department of Computer Science and Engineering,

*†‡§¶Ahsanullah University of Science and Technology,Dhaka,Bangladesh

Email:{*shuvashisofficial, †n.a.refat2000, ‡faikafairuj2001, §shawontanvir95}@gmail.com

¶raihantanvir.cse@aust.edu.

Abstract—**Advanced diagnostic instruments are crucial for the accurate detection and treatment of lung diseases, which affect millions of individuals globally. This study examines the effectiveness of deep learning and transfer learning models using a hybrid dataset, created by merging four individual datasets from Bangladesh and global sources. The hybrid dataset significantly enhances model accuracy and generalizability, particularly in detecting COVID-19, pneumonia, lung opacity and normal lung conditions from chest X-ray images. A range of models, including CNN, VGG16, VGG19, InceptionV3, Xception, ResNet50V2, InceptionResNetV2, MobileNetV2 and DenseNet121 are applied to both individual and hybrid dataset. The results show superior performance on the hybrid dataset, with VGG16, Xception, ResNet50V2 and DenseNet121 each achieving an accuracy, precision, recall and f1-score of 99%. This consistent performance across the hybrid dataset highlights the robustness of these models in handling diverse data while maintaining high accuracy. To understand the models implicit behavior, explainable AI techniques are employed to illuminate their black-box nature. Specifically, LIME is used to enhance the interpretability of model predictions, especially in cases of misclassification, contributing to the development of reliable and interpretable AI-driven solutions for medical imaging.**

Index Terms—**COVID-19, Lung Opacity, Normal, Pneumonia, Hybrid Dataset, Convolutional Neural Network (CNN), Explainable AI (XAI), Local Interpretable Model-agnostic Explanations (LIME)**

I. INTRODUCTION

Palmo aka lungs often referred to as the *silent powerhouses* of the body, play a vital role in sustaining life by ensuring the delivery of essential oxygen to every cell and the removal of carbon dioxide. Lung diseases, encompassing a broad spectrum of conditions, can range from mild afflictions to severe, life-threatening disorders. The advent of the COVID-19 pandemic, caused by the SARS-CoV-2 virus, has brought unprecedented attention to the significance of respiratory health, as the virus primarily targets the respiratory system. The rapid global spread of COVID-19 transformed it into a pandemic, with the World Health Organization (WHO) reporting an overwhelming 775,867,547 cases and 7,057,145 deaths worldwide as of August 4, 2024[1]. The DGHS found 2,051,190 cases and 29,499 deaths in Bangladesh. The age group 15–34 had the most cases, and the age group 60 and older had the most deaths[2].

The complexities of pulmonary diseases across diverse populations and the necessity of a more comprehensive approach to appreciating their impact are recognized in our investigation. The ability to apply of findings across broader demographics may be restricted by the fact that existing research has frequently concentrated on region-specific datasets. In order to resolve this matter, we have combined datasets from various regions, including two from Bangladesh and two from other regions, into a singular hybrid dataset. This method allows for a comprehensive analysis of the impact of increased data diversity on the efficacy of deep learning models in the context of lung diseases. The objectives of our study are threefold:

i. To evaluate the performance of Deep learning and Transfer learning models on individual Regional Datasets.

ii. To assess how these models perform when applied to a Hybrid Dataset created from multiple regional datasets, including those from Bangladesh and global sources.

iii. To utilize Explainable AI techniques to identify and understand the reasons behind misclassifications, thereby enhancing the interpretability of the results.

The opportunities and challenges of utilizing a variety of datasets to analyze pulmonary disease are the focus of our research. This method improves our understanding of the efficacy of models in various contexts. The remainder of this paper is structured as follows: Section II discusses the related work, while Section III outlines the proposed methodology, and Section IV presents the analysis of the results. Techniques related to Explainable AI are detailed in Section V, Limitations and Future Works are described in Section VI and concluding remarks are provided in Section VII

II. RELATED WORK

A. CNN and Pre-trained Models

Talukder et al. (2024) [1] utilized two datasets, the COVID-19 X-ray dataset and the Chest X-ray Image Dataset [2], to

---

[1]https://data.who.int/dashboards/covid19/cases?n=

[2]https://dashboard.dghs.gov.bd/pages/covid19.ph

fine-tune several deep learning models for COVID-19 and lung disease detection. They tested models like Xception, InceptionResNetV2, ResNet50, ResNet50V2, EfficientNetB0, and EfficientNetB4, achieving up to 100% accuracy in COVID-19 detection with EfficientNetB4. Additionally, EfficientNetB4 also demonstrated high accuracy (99.17%) in lung disease detection, highlighting the effectiveness of fine-tuned transfer learning models for medical imaging tasks. Similarly,Hariri et al (2023) [3] created a simple and efficient CNN model to differentiate between COVID-19, viral pneumonia, and bacterial pneumonia using chest X-rays. Their model achieved an accuracy of 89.89%, performing better than other popular models like EfficientNet B2, which had an accuracy of 85.7%. At the same time, Hasan et al. (2023) [4] applied deep transfer learning using VGG16 and ResNet50 models to classify COVID-19 and pneumonia in chest X-ray images. They fine-tuned these pre-trained models on a dataset of 9218 images, achieving accuracies of 89.23% and 88.80%, respectively. In 2023 study, Saleh et al. [5] also analyzed AI models for predicting COVID-19 using a meticulously curated hybrid dataset [6] of 36,170 chest X-ray images. The dataset addressed disparities in histogram characteristics, patient demographics, and scanning environments, ensuring consistency through normalization and bi-cubic interpolation. This diverse dataset enabled models like DenseNet121, InceptionResNetV2, ResNet152V2, VGG16, and Xception—particularly ResNet152V2—to achieve high accuracy (96.17%) and strong generalization across different populations, highlighting the value of a well-prepared hybrid dataset in medical image analysis. Additionally, Roy et al. (2022) [7] evaluated several CNN models, including ResNet-50, InceptionV3, Xception, DenseNet-121, VGG-16, and VGG-19, for detecting COVID-19 from chest X-ray images. They addressed class imbalance in the highly imbalanced CXR dataset by using SVD-CLAHE Boosting and a BWCCE loss function. The ResNet-50 model, with these enhancements, outperformed the others, achieving 94% accuracy and a 96% AUC. Sunyoto et al. [8] evaluated the performance of the VGG16 model using nine publicly available chest X-ray datasets in 2022, comprising 38,181 images across four categories: normal, pneumonia, viral pneumonia, and COVID-19. The model achieved a maximum accuracy of 97.99%, demonstrating robust performance. Srivastava et al.introduced CoviXNet in 2022, a lightweight CNN for COVID-19 detection using chest X-rays [9]. Comparing it with models like InceptionV3, Modified EfficientNet B0 & B1, and ResNetV2 on 6207 X-rays, CoviXNet achieved 99.47% accuracy in binary classification and 96.61% in 3-class classification, offering strong performance with fewer parameters. InceptionV3 and Modified EfficientNet slightly outperformed in binary classification (99.78% accuracy) but required more computational resources.

B. Explainable Artificial Intelligence

Ifty et al. [10] did a study in 2024 that looked at how well different deep learning models could grade lung diseases from chest X-rays. Using stratified 5-fold cross-validation, the Xception model got the best results, with a 96.21% success rate. It did better than DenseNet121 and InceptionV3, which both got to about 92%. The study also looked at combination models, and the InceptionV3+Xception hybrid did very well. XAI methods like Grad-CAM and LIME were used to make the explanations more visual. Mahamud et al. (2024) [11] developed a Dense CNN with Modified Dense Blocks and Transition Layers to classify lung diseases like COVID-19, pneumonia, and tuberculosis. Using transfer learning and data augmentation, their model achieved 99.20% accuracy with 99% precision and recall, outperforming EfficientNetB0, InceptionV3, and LeNet. The study also employed Explainable AI techniques, including SHAP, LIME, Grad-CAM, and Grad-CAM++, to enhance model interpretability. Gasca Cervantes et al. [12] used 10,686 chest X-rays to test four CNN models that had already been trained: VGG16, DenseNet201, ResNet50, and EfficientNetB3. The most accurate network was DenseNet201, with 96.47% of the time. It was followed by VGG16 (95.71%), ResNet50 (93.15%), and EfficientNetB3 (92.39%). DenseNet201 did the best, but VGG16 didn't focus on useful lung regions when LIME-based XAI analysis was done. Ong et al. (2021) [13] further utilized the SqueezeNet to classify chest X-rays into the categories of COVID-19, normal, and pneumonia with an 84.3% accuracy on the COVIDx dataset. They used XAI techniques, LIME and SHAP, for interpretability, where SHAP gives more consistent and accurate explanations in model understanding through the importance of the highlighted critical regions of lungs for an improved interpretation in medical imaging.

III. METHODOLOGY

This research aims to develop a system to identify COVID-19, pneumonia, lung opacity, and normal lung conditions using chest X-ray images. Figure 2 illustrates the methodology of our study which involves collecting a comprehensive dataset, preprocessing iteratively, fine-tuning hyperparameters, creating a hybrid dataset for data analysis, and implementing a LIME explainer on the best-performing model to enhance transparency and trustworthiness.

A. Dataset

In our study, we have utilized four publicly available datasets to enhance the detection capabilities of lung conditions through Chest X-ray images. **Dataset 1**, *Balanced Augmented Covid CXR Dataset*, developed by Mrinal Tyagi [14] and published on Kaggle, is an improved version of the widely recognized COVID-19 Radiography Database. This dataset, containing a total of 29,998 images, was balanced using various techniques to address significant class imbalance issues. **Dataset 2**, *Chest X-Ray Image Dataset*, compiled by Md Alamin Talukdar [2] and available on Mendeley Data, includes 4,350 images were collected from hospitals in Bangladesh, meticulously categorized into four classes to facilitate research and diagnostic efforts related to lung conditions. **Dataset 3**, *Lung X-Ray Image Dataset*, also assembled by Md Alamin Talukdar [15] and available on Mendeley Data,

comprised 3,475 X-ray images classified into three distinct categories, gathered from various healthcare sources to support the detection and diagnosis of lung conditions. Lastly, **Dataset 4**, *Curated Dataset for COVID-19 Posterior-Anterior Chest Radiography Images (X-Rays)*, was prepared by Unais Sait et al [6]. and publicly available on Mendeley, is a comprehensive collection obtained by aggregating 15 publicly available datasets, providing a diverse range of COVID-19 related X-ray images with a total of 4,350 images. In this dataset, the two classes *Viral Pneumonia* and *Bacterial Pneumonia* are merged into a single general class named *Pneumonia* to maintain consistency across the datasets. To ensure uniformity, the class name *Viral Pneumonia* in other datasets has also changed to *Pneumonia.*

A proposed **hybrid dataset** which have been used in our study is a combined collection of these multiple datasets, integrated to create a more comprehensive and diverse data source for analysis and training purposes. Figure 1 summarizes the distribution of images across different categories in the four datasets, highlighting the variety and volume of data available for each condition.
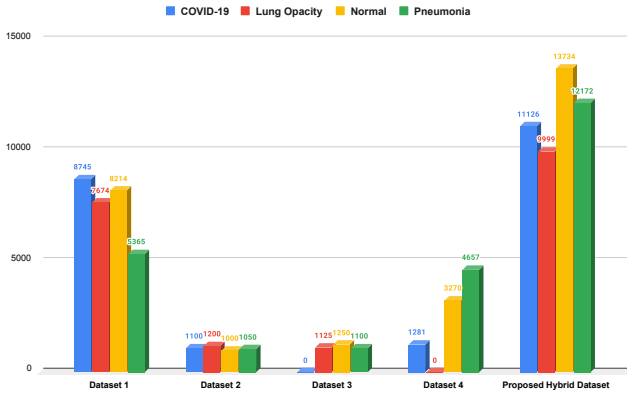

Fig. 1: Comparison of Different Datasets by Categories

B. Data Preprocessing

The dataset we use containing X-ray images with varying shapes and resolutions. To standardize the input data and ensure consistency across the entire dataset,initially, all images from Dataset 2 and Dataset 3 were converted to grayscale for uniform analysis.After that, the images from Dataset 1 and the Hybrid Dataset were resized to 128x128 pixels, while the images from Datasets 2, 3, and 4 were standardized to 256x256 pixels. To prepare the data for model training, each dataset, including the Hybrid Dataset, was split into an 80:10:10 ratio using the Keras Utils library. In this split, 80% of the images were allocated for training, 10% for validation, and 10% for testing.

C. Proposed Method

The proposed model architecture integrates a range of deep convolutional neural networks (CNNs), including VGG16,

VGG19, ResNet50V2, InceptionV3, Xception, MobileNetV2, InceptionResNetV2, and DenseNet121, which are renowned for their image classification capabilities. VGG16 and VGG19, based on 3x3 convolutions, are widely used for feature extraction, with VGG19 having additional layers to capture more complex patterns. ResNet50V2 leverages residual connections to avoid vanishing gradients and improve training efficiency. InceptionV3 employs multi-scale feature extraction via various filter sizes (1x1, 3x3, 5x5) and reduction modules to minimize complexity. Xception, a variant of Inception, applies depthwise separable convolutions to improve efficiency. InceptionResNetV2 merges Inception modules with residual connections for enhanced feature learning and reduced vanishing gradients. MobileNetV2 is optimized for resource-constrained devices, utilizing inverted residuals and bottlenecks to balance performance and computational efficiency. Lastly, DenseNet121 ensures strong feature propagation by densely connecting each layer to all subsequent layers, thus mitigating the vanishing-gradient problem and reducing parameters [16].

In our study, we employed both CNNs and pre-trained models on four distinct datasets, as well as a hybrid dataset created by merging the four. Our CNN architecture consists of three Conv2D layers, increasing from 32 to 128 filters, and MaxPooling2D layers for dimensionality reduction. After flattening the data, two dense layers with 1024 and 512 neurons are applied, followed by dropout layers to reduce overfitting. The final softmax layer outputs four neurons corresponding to the dataset classes. For pre-trained models, we retained their base architecture and added custom fully connected layers (1024 and 512 neurons), along with dropout layers (0.5 and 0.25 dropout rates). Fine-tuning these pre-trained networks with our dataset allowed us to harness their feature extraction capabilities, improving performance while reducing training time.

IV. RESULT ANALYSIS

The methodologies are executed on a computer system, with the best results selected for analysis. The system is equipped with an Intel Core i5 13400F processor, an NVIDIA GeForce RTX 3060 12GB GPU, and 16GB of DDR4 RAM. The models are implemented using Keras [3], a high-level neural networks API, and TensorFlow [4] as the backend. To ensure robustness, each model runs five times, and the best outcomes are recorded.

The Table I presents the performance metrics for various models across different datasets. For the individual datasets, Dataset 1 [14] shows exceptionally high accuracy across all models, with scores consistently reaching **99%** for advanced models such as **VGG16**, **VGG19**, **InceptionV3**, **Xception**, **ResNet50V2**, **InceptionResNetV2**, **MobileNetV2**, and **DenseNet121**. This indicates that these models are highly effective when applied to Dataset 1 [14]. In contrast, Dataset 2 [2] shows more variability, with **CNN** achieving an accuracy
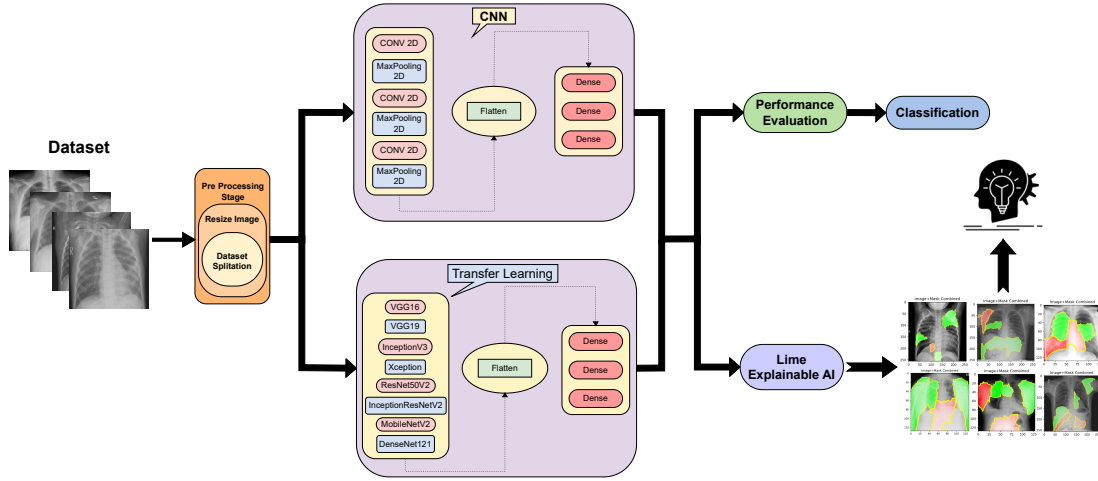
Fig. 2: A Flow Diagram of Our Proposed Model

TABLE I: Performance metrics for different models across datasets

Dataset	Model	Accuracy	Precision	Recall	F1 Score
Dataset 1 [14]	CNN	0.97	0.98	0.98	0.98
	VGG16	**0.99**	**0.99**	**0.99**	**0.99**
	VGG19	**0.99**	**0.99**	**0.99**	**0.99**
	InceptionV3	**0.99**	**0.99**	**0.99**	**0.99**
	Xception	**0.99**	**0.99**	**0.99**	**0.99**
	ResNet50V2	**0.99**	**0.99**	**0.99**	**0.99**
	InceptionResNetV2	**0.99**	**0.99**	**0.99**	**0.99**
	MobileNetV2	**0.99**	**0.99**	**0.99**	**0.99**
	DenseNet121	**0.99**	**0.99**	**0.99**	**0.99**
Dataset 2 [2]	CNN	0.84	0.85	0.84	0.84
	VGG16	0.93	0.93	0.93	0.93
	VGG19	0.91	0.91	0.91	0.91
	InceptionV3	0.89	0.89	0.89	0.89
	Xception	0.91	0.91	0.91	0.91
	ResNet50V2	0.90	0.90	0.90	0.90
	InceptionResNetV2	0.87	0.87	0.87	0.87
	MobileNetV2	0.90	0.90	0.90	0.90
	DenseNet121	**0.94**	**0.94**	**0.94**	**0.94**
Dataset 3 [15]	CNN	0.91	0.93	0.91	0.92
	VGG16	0.93	0.93	0.93	0.93
	VGG19	0.90	0.90	0.90	0.90
	InceptionV3	0.92	0.92	0.92	0.92
	Xception	0.91	0.91	0.91	0.91
	ResNet50V2	0.92	0.92	0.92	0.92
	InceptionResNetV2	0.90	0.90	0.90	0.90
	MobileNetV2	0.91	0.92	0.91	0.91
	DenseNet121	**0.94**	**0.94**	**0.94**	**0.94**
Dataset 4 [6]	CNN	0.97	0.98	0.96	0.97
	VGG16	**0.98**	**0.99**	**0.98**	**0.99**
	VGG19	**0.98**	**0.99**	**0.99**	**0.99**
	InceptionV3	**0.98**	**0.98**	**0.98**	**0.98**
	Xception	**0.98**	**0.99**	**0.98**	**0.99**
	ResNet50V2	**0.98**	**0.98**	**0.98**	**0.98**
	InceptionResNetV2	0.97	0.98	0.97	0.97
	MobileNetV2	**0.98**	**0.98**	**0.98**	**0.98**
	DenseNet121	**0.98**	**0.98**	**0.98**	**0.98**
Proposed Hybrid Dataset	CNN	0.96	0.96	0.95	0.96
	VGG16	**0.99**	**0.99**	**0.99**	**0.99**
	VGG19	0.98	**0.99**	0.98	**0.99**
	InceptionV3	0.98	0.98	0.98	0.98
	Xception	**0.99**	**0.99**	**0.99**	**0.99**
	ResNet50V2	**0.99**	**0.99**	**0.99**	**0.99**
	InceptionResNetV2	0.98	0.98	0.97	0.98
	MobileNetV2	0.98	0.98	0.98	0.98
	DenseNet121	**0.99**	**0.99**	**0.99**	**0.99**

of **84%**, while **DenseNet121** achieves the highest accuracy of **94%**. Similarly, Dataset 3 [15] results range from **91%** accuracy with **CNN** to **94%** with **DenseNet121**. Dataset 4 [6] also performs strongly, with **VGG16**, **VGG19**, and several other models consistently achieving accuracy around **98%** to **99%**.

For the hybrid dataset, the proposed combination of datasets maintains high performance, with most models achieving near-perfect accuracy. The **VGG16**, **Xception**, **ResNet50V2**, and **DenseNet121** models stand out, each reaching an accuracy of **99%**, showcasing their robustness and ability to generalize well when faced with a more diverse dataset. This high consistency across the hybrid dataset demonstrates the effectiveness of these models in handling varied data while maintaining top-tier accuracy.

Overall, while all models have performed well across both individual and hybrid datasets, the advanced models such as **VGG16**, **Xception**, **ResNet50V2**, and **DenseNet121** have consistently showed superior accuracy, especially with the hybrid dataset. Table II shows that our proposed model, utilizing a hybrid dataset with multiple CNN architectures and XAI techniques, achieves competitive accuracy compared to existing methods, many of which do not incorporate XAI. This consistency demonstrates that these models are highly effective at generalizing across diverse datasets, achieving the near-perfect accuracy we aimed for. The results confirm that our approach of combining multiple datasets into a hybrid dataset to test model performance was successful, validating that these models can handle varied data sources while maintaining high accuracy. This indicates that these models can work effectively for data from any region, making them versatile tools for broader applications.

V. EXPLAINABLE AI

Explainable Artificial Intelligence (XAI) seeks to enhance the transparency of AI models by providing clear explanations of their decision-making processes, which aids in improving prediction accuracy and building trust between humans and AI.

TABLE II: Comparison of Our Proposed Approach with Existing Methods

Author Name	Approach	Accuracy	Using of XAI
Talukder et al. [1]	EfficientNetB4 + Fine-Tuning	99.17%	NO
Roy et al [7].	ResNet-50+ SVD-CLAHE Boosting+ BWCCE	94%	NO
Hariri et al [3].	CNN	89.89%	NO
Sunyoto et al. [8]	VGG16	97.99%	NO
Hasan et al [4].	VGG16	89.23%	NO
Srivastava et al. [9]	CoviXNet	99.47%	NO
Saleh et al [5].	ResNet50V2	96.17%	NO
Ong et al [13].	SqueezeNet	84.3%	YES (LIME & SHAP)
Gasca Cervantes et al. [12]	DenseNet201	96.47%	YES (LIME)
Ifty et al. [10]	Xception+ Fine-Tuning	96.21%	Yes (Grad-CAM & LIME)
Mahamud et al. [11]	Dense CNN	99.20%	YES (LIME,SHAPE, Grade-CAM, Grade-Cam++
Proposed Model	**Hybrid Dataset+ VGG16,Xception, ResNet50V2, DenseNet121**	**99%**	**YES (LIME)**

XAI helps users understand the strengths and weaknesses of AI models, leading to a better comprehension of their behavior.

A. LIME:

LIME (Local Interpretable Model-agnostic Explanations) is an Explainable AI technique developed by Ribeiro et al. [17] to clarify how complex classifiers and regressors make predictions. LIME operates by locally approximating any model to a simplified, interpretable counterpart, thereby emphasizing the relationship between input features and outputs. It is model-agnostic, which means that it can consider the original model as a black box, rendering it versatile across a variety of data types, including images, text, and tabular data [18]. LIME functions by executing four fundamental procedures: permuting input data to generate artificial data points, predicting classes for these points, calculating their similarity to the original data, and employing a linear classifier to identify the most significant features that influence the model's decisions. This method enhances the transparency and comprehensibility of AI models [19].

In Figure 3 shows cases where the model misclassified chest X-rays, with LIME explanations providing insights into the model's reasoning. In Figure 3a from Dataset 1 [14], the model predicts *Normal* when the true label is *Lung Opacity*, with green regions indicating areas the model interprets as *Normal*, leading to the misclassification. The red regions correspond to other classes, but the *Normal* class has a higher probability. In Figure 3b from Dataset 2 [2], the model incorrectly predicts *COVID* for a case of *Lung Opacity*, with LIME highlighting green regions indicative of *COVID*, which misleads the model. In Figure 3c from Dataset 3 [15], the model predicts
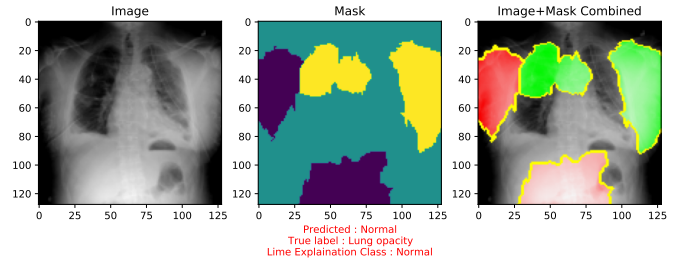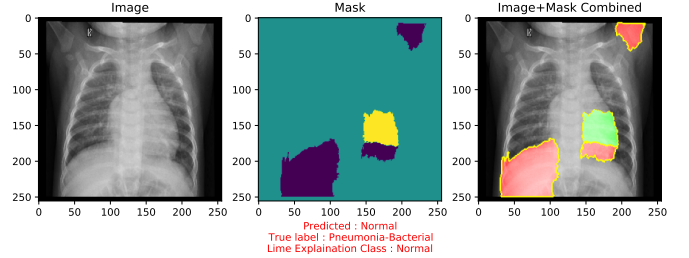

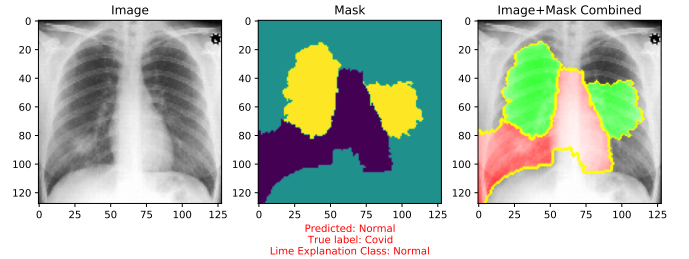(a) Misclassification Image of Dataset 1
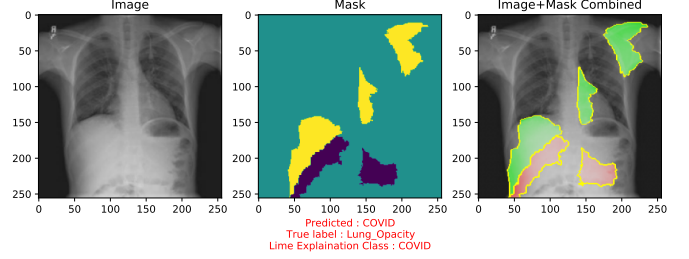
(b) Misclassification Image of Dataset 2
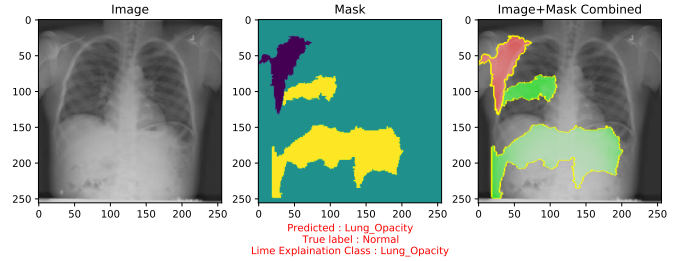
(c) Misclassification Image of Dataset 3

(d) Misclassification Image of Dataset 4

(e) Misclassification Image of Hybrid Dataset

Fig. 3: Some Sample Images of Misclassification by the Model

Lung Opacity instead of the correct *Normal* label, as the LIME explanation shows green areas the model associated with *Lung Opacity*, resulting in the error. For Figure 3d from Dataset 4 [6], the model predicts *Normal* when the

true label is *Pneumonia*, again with green regions supporting the *Normal* prediction, causing the misclassification. Finally, Figure 3e from the hybrid dataset, the model misclassifies a *Covid* case as *Normal*, with LIME explanations showing green areas that the model interprets as *Normal*, leading to the incorrect prediction. From these LIME explanations, we gain valuable insights into the model's black-box decision-making process, enhancing our understanding of its interpretability and identifying misclassification regions.

VI. Limitations & Future Works

The limitations of this study include regional constraints that may have impacted the generalizability of the findings, as the data primarily reflected specific populations. Additionally, the study employed only one explainable AI (XAI) technique, potentially restricting the depth of interpretability and analysis. Moreover, limitations in the use of transformers and ensemble techniques may have affected the overall performance and robustness of the models. Further research will involve examining larger datasets to enhance the model's robustness and reliability, employing various XAI techniques to explain its decision-making processes, and incorporating data from multiple geographies to mitigate restrictions. Enhancements will concentrate on employing picture segmentation for increased precision, optimizing model architectures, investigating novel techniques like as transformers and ensemble learning, and evaluating the system's applicability in practical medical environments.

VII. Conclusion

In conclusion, this study demonstrated that integrating regional datasets into a hybrid model significantly enhances the accuracy and generalizability of deep learning models in detecting lung diseases from chest X-rays. The hybrid approach proved effective across diverse datasets, and the use of LIME provided valuable insights into the models' decision-making processes, contributing to a better understanding of how these models classify different lung conditions. The successful application of these techniques underscores the potential for more accurate and interpretable AI-driven medical imaging solutions.

References

[1] M. A. Talukder, M. A. Layek, M. Kazi, M. A. Uddin, and S. Aryal, "Empowering covid-19 detection: Optimizing performance through fine-tuned efficientnet deep learning architecture," *Computers in Biology and Medicine*, vol. 168, p. 107789, 2024.

[2] M. A. Talukder, "Chest x-ray image." https://data.mendeley.com/datasets/m4s2jn3csb/1, 2023.

[3] M. Hariri and E. Avşar, "Covid-19 and pneumonia diagnosis from chest x-ray images using convolutional neural networks," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 12, no. 1, p. 17, 2023.

[4] M. R. Hasan, S. M. A. Ullah, and M. Hasan, "Deep learning in radiology: A transfer-learning based approach for the identification and classification of covid-19 and pneumonia in chest x-ray images," in *2023 Fourth International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, pp. 1–6, IEEE, 2023.

[5] R. A. Saleh, F. Al-Areqi, Z. Al-Huda, and M. A. Al-antari, "Comparative analysis of artificial intelligence for predicting covid-19 using diverse chest x-ray images," in *2023 3rd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, pp. 1–7, IEEE, 2023.

[6] UNAIS SAIT, "Curated dataset for covid-19 posterior-anterior chest radiography images (x-rays).." https://data.mendeley.com/datasets/9xhgts2s6/4, 2022.

[7] S. Roy, M. Tyagi, V. Bansal, and V. Jain, "Svd-clahe boosting and balanced loss function for covid-19 detection from an imbalanced chest x-ray dataset," *Computers in Biology and Medicine*, vol. 150, p. 106092, 2022.

[8] A. Sunyoto, Y. Pristyanto, A. Setyanto, F. Alarfaj, N. Almusallam, and M. Alreshoodi, "The performance evaluation of transfer learning vgg16 algorithm on various chest x-ray imaging datasets for covid-19 classification," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, 2022.

[9] G. Srivastava, A. Chauhan, M. Jangid, and S. Chaurasia, "Covixnet: A novel and efficient deep learning model for detection of covid-19 using chest x-ray images," *Biomedical Signal Processing and Control*, vol. 78, p. 103848, 2022.

[10] T. T. Ifty, S. A. Shafin, S. M. Shahriar, and T. Towhid, "Explainable lung disease classification from chest x-ray images utilizing deep learning and xai," *arXiv preprint arXiv:2404.11428*, 2024.

[11] E. Mahamud, N. Fahad, M. Assaduzzaman, S. Zain, K. O. M. Goh, and M. K. Morol, "An explainable artificial intelligence model for multiple lung diseases classification from chest x-ray images using fine-tuned transfer learning," *Decision Analytics Journal*, vol. 12, p. 100499, 2024.

[12] E. G. Cervantes and W.-Y. Chan, "Lime-enabled investigation of convolutional neural network performances in covid-19 chest x-ray detection," in *2021 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, pp. 1–6, IEEE, 2021.

[13] J. H. Ong, K. M. Goh, and L. L. Lim, "Comparative analysis of explainable artificial intelligence for covid-19 diagnosis on cxr image," in *2021 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 185–190, IEEE, 2021.

[14] M. Tyagi, "Balanced augmented covid cxr dataset." https://www.kaggle.com/datasets/tr1gg3rtrash/balanced-augmented-covid-cxr-dataset, September 2022. Accessed on July 05, 2024.

[15] M. A. Talukder, "Lung x-ray image." https://data.mendeley.com/datasets/9d55cttn5h/1, 2023.

[16] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, 2023.

[17] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[18] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: a systematic review on the application of lime and shap in alzheimer's disease detection," *Brain Informatics*, vol. 11, no. 1, p. 10, 2024.

[19] R. Maliyat, F. Sultana, S. Sultana, S. K. S. Joy, and M. T. R. Shawon, "Investigating the effectiveness of interpretable cost-sensitive neural network for pneumonia detection contemplating data imbalance," 2023.