Project Report on

Explainable AI for Pulmonary Disease Diagnosis

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

By

Nandakishore T J (U2103145)

Niranjana S Nair (U2103159)

Nithin Kurisingal Cimu (U2103161)

Shanker Menon (U2103193)

Under the guidance of

Ms. Dincy Paul

Department of Computer Science and Engineering
Rajagiri School of Engineering & Technology (Autonomous)
(Parent University: APJ Abdul Kalam Technological University)
Rajagiri Valley, Kakkanad, Kochi, 682039
April 2025

CERTIFICATE

This is to certify that the project report entitled ***"Explainable AI for Pulmonary Disease Diagnosis"*** *is a bonafide record of the work done by* ***Nandakishore T J (U2103145)****,* ***Niranjana S. Nair (U2103159)****,* ***Nithin Kurisingal Cimu (U2103161)****,* ***Shanker Menon (U2103193)****, submitted to the Rajagiri School of Engineering & Technology (RSET) (Autonomous) in partial fulfillment of the requirements for the award of the degree of Bachelor of Technology (B. Tech.) in "Computer Science and Engineering" during the academic year 2024-2025.*

Ms. Dincy Paul
Project Guide
Assistant Professor
Dept. of CSE
RSET

Ms. Sangeetha Jamal
Project Coordinator
Assistant Professor
Dept. of CSE
RSET

Dr. Preetha K G
Head of the Department
Professor
Dept. of CSE
RSET

ACKNOWLEDGMENT

We wish to express our sincere gratitude towards **Rev.Dr.Fr.Jaison Paul Mulerickal CMI**, Principal of RSET, and **Dr.Preetha K.G**, Head of the Department of ”Computer Science and Engineering” for providing us with the opportunity to undertake my project, ”Explainable AI for Pulmonary Disease Diagnosis”.

We are highly indebted to my project coordinator, **Ms. Sangeetha Jamal**, Assistant Professor, Department of Computer Science, for her valuable support.

It is indeed our pleasure and a moment of satisfaction for us to express our sincere gratitude to our initial project guide **Ms. Keerthimol P. P.**, Assistant Professor and our current project guide **Ms. Dincy Paul**, Assistant Professor, Department of Computer Science and Engineering, for their patience and all the priceless advice and wisdom they have shared with us.

Last but not the least, We would like to express our sincere gratitude towards all other teachers and friends for their continuous support and constructive ideas.

Nandakishore T J
Niranjana S Nair
Nithin Kurisingal Cimu
Shanker Menon

Abstract

Explainable Artificial Intelligence (XAI) becomes ever more significant as advanced intelligent systems develop, especially in high-stakes and sensitive areas such as medicine. For example, in the use of automated disease diagnosis, transparency is paramount because the results have direct impact on patients and need to be understandable and justifiable. Respiratory diseases such as covid and pneumonia are severe risks to human health. Detection and categorization of the diseases in early stages through chest radiographs can significantly enhance recovery and treatment. The system proposed here includes an innovative method that utilizes Convolutional Neural Networks (CNNs) to detect pulmonary disease automatically and provide explainable results of model predictions. To enhance the transparency of the decision-making process, the system incorporates Gradient-weighted Class Activation Mapping (Grad-CAM), which visually marks areas in the input images that have a strong impact on the model's outputs. This facilitates easier understanding by medical professionals of how and why specific diagnoses are being made.

The software is built with Python, with a custom-built, optimized interface developed using the PySide framework for seamless deployment on Windows. The interface is easy to use, enabling doctors to upload radiographs, obtain AI-driven diagnostic predictions, and see visual explanations in the form of heatmaps that highlight important areas of interest. Second, the program is extensively tested and validated through genuine datasets for its reliable and efficient diagnostic performance. Through emphasizing interpretability, such a system allows for higher reliance on AI-facilitated diagnostics as well as empowers healthcare experts in making better informed and assertive decisions towards tackling pulmonary illnesses.

Contents

List of Abbreviations

COPD - Chronic Obstructive Pulmonary Disease

AI - Artificial Intelligence

CNN - Convolutional Neural Network

Grad-CAM - Gradient-weighted Class Activation Mapping

SHAP - Shapley Additive Explanations

XAI - Explainable AI

CT - Computed Tomography

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Worldwide, respiratory diseases like pneumonia and in recent years diseases like COVID-19 top the list of morbidity and mortality. However, treatment becomes effective only with early and precise diagnosis. But the traditional methods of diagnosis using CT scans and X-rays require expert interpretation often, which can be cumbersome and varied in the interpretation between radiologists. The increasing workload for medical personnel is another aspect that makes it necessary to develop reliable,automated, , and diagnostic devices.

Breakthroughs in artificial intelligence (AI), particularly in deep learning, have revolutionized the analysis of medical images, making it more precise and effective than ever before. When it comes to analyzing complex medical images for disease identification, convolutional neural networks (CNNs) have proven to be highly efficient. In this research, we develop a pulmonary disease detection system that can accurately evaluate lung images to identify potential disease regions. Neural networks, particularly CNNs, have demonstrated remarkable success in medical imaging by learning intricate patterns and features, enabling automated and reliable diagnosis.

The requirement for explainability is a significant obstacle when implementing AI for clinical application. The model's conclusions must be clear and understandable for medical practitioners to have faith in and follow AI-based diagnosis. Grad-CAM (Gradient-weighted Class Activation Mapping), which our approach utilizes to address this, identifies particular regions of the image that the model concentrates on during prediction. This enhances interpretability and trust by giving clinicians a visual representation of the model's attention.

Moreover, for a more detailed interpretability approach, SHAP consistently contributes feature-level information to the model's output. This dual interpretability approach also provides all kinds of explanations, therefore enabling physicians to gain a profound understanding of why the model outputs a given set of diagnoses. This paper develops a system by combining XAI methods with an excellent detection model to develop not only an accurate but also trustworthy and appropriate one for clinical practice. In less resourced settings for healthcare, such tools This could significantly improve diagnostic skills, ease the workload for radiologists, and enable prompt and accurate therapies, all of which would improve patient outcomes.

1.2 Problem Definition

In many healthcare settings, diagnosing medical conditions is a challenging task. There is often inaccuracy and a lack of transparency within the systems. Doctors end up working harder on interpreting results that they cannot rely on. Such situations are harder in less developed places, where delays or mistakes can have a heavy price on patients.

Our project targets these challenges using the power of explainable AI (XAI) in highly effective detection. We will strive to build an accurate, easy-to-understand system that can aid doctors in this regard. This way, their work will be somewhat lightened; however, in the end, it will improve the care with which patients will be treated-thus helping communities needing it the most.

1.3 Scope and Motivation

1.3.1 Scope

The primary goals of this study are to explore the development and implementation of a system that combines AI and analytical capabilities to discern pulmonary conditions with the help of medical images or chest radiographs. The system employs advanced deep learning techniques, particularly convolutional neural networks (CNN), which are highly effective in automatically identifying and extracting complex features from lung scans. In addition to the core CNN architecture, methods such as data augmentation, transfer learning, and fine-tuning have been incorporated to enhance accuracy and robustness. These approaches allow the system to effectively diagnose conditions such as COVID-19 and pneumonia by recognizing subtle patterns indicative of lung disease.The model's decision is intelligible and explainable to medical practitioners, and it also deploys various advanced explainability methods such as Grad-CAM, and SHAP. Despite the fact that the framework initially focuses on pulmonary diseases, it can later be adapted to other medical imaging applications. In addition to providing meaningful visuals, the system should also be user-friendly as it sends information and predictive diagnoses to clinicians.

1.3.2 Motivation

With the increased incidence of lung diseases all over the world and a lack of qualified doctors in some places, there is a dire need for reliable, automated diagnostic methods.After the COVID-19 pandemic,we have seen the first-hand impact of needing accurate rapid diagnosis especially when most doctors are forced to fight on the frontlines. The conventional techniques of diagnosing through manual scanning of CT and X-ray images are cumbersome and susceptible to human error. The proposed work uses AI to hasten and make the diagnosis more accurate. Further, the introduction of explainability techniques, such as Grad-CAM, and SHAP, overcome the usual obstacle of AI ”black-box” models, increasing acceptance and trust from medical practitioners in the system. This initiative aims at improving patient outcomes by giving doctors advanced, easy-to-understand tools for the early identification of diseases.

1.4 Objectives

- Design a system along with the algorithm applications such as CNN in predicting the disease,i.e. pneumonia and COVID-19 in the medical image containing CXR.
- Develop explanation methods such as Grad-CAM, and SHAP integrated with AI-based models such that healthcare practitioners understand the outputs from the models better.
- Make the difference on a large number of available data for medical images. model, which you can fine-tune to its performance and derive high accuracy and reliability.
- Design a highly interactive user interface easy for doctors to upload and analyze images with visual explanation accompanying the predictive diagnosis.

- Evaluate the diagnostic performance and interpretability of the system using metrics such as accuracy, sensitivity, specificity, and explainability.

1.5 Challenges

Medical images often vary in quality and may contain subtle indicators of illness, making it a significant challenge to ensure the AI model remains accurate and generalizable across different datasets. Additionally, while methods like Grad-CAM, and SHAP are valuable for interpretability, they can come at the cost of reduced model performance and increased computational complexity. Striking the right balance between accurate predictions and clear explanations will require careful adjustments and fine-tuning.

1.6 Assumptions

1. The model was trained using a diverse set of high-quality, carefully labeled chest X-Ray images, capturing various stages of the lung diseases.
2. The deep learning model is designed to perform consistently and accurately over time when applied to X-rays, without any noticeable drop in performance.
3. The interpretable, meaningful insights that the explainability techniques (Grad-CAM, and SHAP) will offer will increase healthcare practitioners' confidence in and acceptance of the model's predictions.
4. The system will be set up in a medical setting where medical practitioners can access the computer power and user interfaces required for picture processing.
5. With very minor adjustments to the current architecture, The AI-driven detection system will be flexible enough to evolve with advancements in imaging technology and adapt to the detection of new illness categories.

1.7 Societal / Industrial Relevance

In both medical and industrial contexts, this lung illness detection technology is extremely pertinent. From a societal perspective, it tackles the increasing global burden of pulmonary illnesses by offering an early diagnosis tool that can enhance patient outcomes and lower death rates. Early identification is essential for conditions like COVID-19 because of its virality and lethality,especially among patients with pre-existing health conditions such as a compromised immune system.Thus prompt treatment can greatly increase the likelihood of a successful outcome. Healthcare workers can use AI-driven diagnostic help to make well-informed judgments in resource-constrained places where access to skilled radiologists may be limited. Industrially, the technology can be implemented by clinics, hospitals, and medical research organizations to decrease healthcare expenses, lessen radiologists' workloads, and expedite the diagnosis process.

1.8 Organization of the Report

The backdrop, precise definition of the problem, the scope of the research, and the rationale behind it are all covered in **Chapter 1**, which lays the groundwork for the study. Along with listing any assumptions made during the project, this chapter also describes the study's goals and points out any potential difficulties. Furthermore, it highlights the topic's wider significance by examining its sociological and industrial relevance. A thorough evaluation of related studies and research publications is provided by the literature survey in **Chapter 2**, which also summarizes important approaches and findings. Additionally, this chapter examines current solutions, pointing out weaknesses and difficulties in the literature, and talks about how these support and validate the methodology used in this investigation. **Chapter 3** is the system architecture's high-level view, where the system components and the various algorithms are explained as well, We can also find the dataset that has been identified, the tools and technologies used, the data flow diagram, and the key deliverables. The manner in which the modules are divided as well as each member's contributions to the project with its proper timeline. **Chapter 4** compares the performance of InceptionV3, ResNet-50, and a Custom CNN based on evaluation metrics. Although pre-trained models performed higher accuracy, the Custom CNN performed better in generalization and was less overfitted and thus more appropriate for real-world applications. **Chapter 5** showcases that the project applied explainable AI and deep learning for identifying pulmonary disease through chest X-rays. The future work would involve its expansion to other imaging tests such as CT or MRI and enhancing the capabilities to facilitate its quicker usage in clinics in real time.

1.9 Summary of the Chapter

Chapter 1, starts with the background to set the scene and emphasize the importance of the subject. The problem being addressed is then defined, along with the research's goals and scope, which highlights the significance of the study. While the problems section addresses potential roadblocks, the objectives clearly outline the goals of the investigation. Important presumptions are mentioned to shed light on any underlying circumstances. It is concluded by highlighting the topic's industrial and societal relevance, pointing out its wider implications and possible uses.

Chapter 2

Literature Survey

Deep learning models, specifically convolutional neural networks (CNNs), have transformed medical image analysis in recent years. They provide insightful solutions for automating the detection and classification of various diseases using techniques like computed tomography (CT) scans and chest X-rays (CXR). These advancements have made it possible to achieve faster More accurate diagnoses are vital for improving patient care and reducing the pressure on healthcare systems. To earn the trust of medical professionals and help them understand the reasoning behind the predictions, it's important to ensure that these AI-driven models are easy to use and interpret.In order to aid in highlighting regions of medical images contributing to a model's judgments, explainable AI (XAI) techniques like LIME or Local Interpretable Model-Agnostic Explanations have been used. The research proposed in this article checks new combinations of explainability techniques with deep learning models toward achieving their validity and reliability for application in a clinical setting. These models are designed to identify particular anomalies in medical imaging, not only for COVID-19 but also for a host of other diseases. These approaches aim to enhance the accuracy of diagnosis and help doctors make rational, informed decisions by providing interpretability on par with performance.

2.1 Related Works

Many studies address the application of deep learning models for improving detection and classification performance in COVID-19 from CT scans and chest radiographs (CXR). In one work, a pre-trained ResNet50-based deep learning model has been developed. The model had improved performance on two benchmark datasets - COVID-CT and COVIDNet. The algorithm successfully recognized CT scans and chest X-ray 6 images to COVID-19 positive or negative classes with an accuracy of 93% on COVID-CT and

97% on COVIDNet. Furthermore, the study improved the interpretability through LIME or Local Interpretable Model-Agnostic Explanations which was applied in highlighting the important regions of the photos which affect the result of classification.This method showcased the advantages of the integration of any explainable AI (XAI) method with deep learning models for medical diagnoses [1].

Another relevant paper employed specific models in Deep Learning named Faster R-CNN as well as Mask R-CNN, mainly specializing in detecting and segmenting objects, toward identifying any peculiar abnormalities related to COVID-19 from CT scan images. These 4000 images of annotated CT scans were accurately created by the authors. The fast R-CNN model performed better than the mask R-CNN model. In the test phase, it produced a mean average precision score of 97.72% against 93. 86% produced by the mask R-CNN model. Mask R-CNN enhanced the diagnostic sensitivity because it could delineate lesions and abnormalities on a microscopic level through the creation of masks for segmentation.[2].

In another paper, we find here some investigation to the XAI techniques on using them in an application on a COVID-19 detection case study. The developed heat maps presented in this contribution were made to highlight important local regions in chest radiographs to provide support to decision-making in classifying COVID-19. The areas which were highlighted contained bilateral subpleural ground-glass opacities which usually are considered the most indicative image features. This approach improved classification accuracy and clearly made model decision-making, hence helping medical practitioners understand the predictions of the model. [3].

To classify COVID-19 from chest X-ray pictures, additional research compared many deep learning architectures, including DenseNet169, MobileNet, and ResNet50. According to the study, ResNet50 was found to be better in terms of classification accuracy than the other models used. To further show ResNet50's promise as a valid diagnostic tool, the study also utilized LIME for higher interpretability and produced heat maps that closely matched the annotations by expert radiologists [4].

Another investigation was led by the goal of enhancing the segmentation of COVID-19 lesions in CT scans. To identify and classify anomalies linked to COVID-19, the scientists made use of both the Faster R-CNN and Mask R-CNN models. Further localization of COVID-19 lesions was made possible by the Mask R-CNN model's exceptional capacity to produce pixel-level segmentation masks. This study highlighted how crucial precise segmentation is to improving diagnosis accuracy [5].

In order to identify and classify COVID-19 abnormalities in CT images, the last relevant study in this context also used deep learning models, Mask R-CNN in particular. Comparison of Mask R-CNN with alternative techniques demonstrated that the accuracy of lesion detection was greatly increased by its micro-scaled segmentation capabilities. This study demonstrated how Mask R-CNN—can accurately identify and locate anomalies arising from COVID-19 [6].

All of these research highlight how important it is to combine explainable AI methods with deep learning models in order to increase the precision, reliability, and transparency of the diagnosis of COVID-19 diagnosis. The use of AI models in clinical settings greatly depends on their ability to classify and separate any and all irregularities resulting from COVID-19 while also giving us insight into the decision-making process.

2.2 Summary and Gaps Identified

Enhancing COVID-19 detection through the use of deep learning models on chest X-rays and CT scans has been the recent subject of several studies. A model based on ResNet50 was created in one study, attaining good accuracy (93%–97%) and employing LIME for interpretability. This model identified pertinent picture regions for diagnosis and was verified by skilled radiologists. With a high mean average precision (mAP) of 97.72%, another study used the 2 aforementioned R-CNNs for identification and segmentation of objects to improve diagnostic accuracy by accurate lesion localization.

Additionally, LIME was utilized to create heat maps that highlighted important COVID-19 characteristics, such as ground-glass opacities, which improved classification accuracy and matched expert annotations well. ResNet50 proved to be the most successful deep learning model in terms of accuracy and interpretability after additional study compared various models. According to studies, precise segmentation of CT scans is crucial, and Mask R-CNN offers pixel-level segmentation for improved lesion detection.

The accuracy, dependability, and transparency of COVID-19 diagnosis can be increased by integrating deep learning with explainable AI methodologies, as these research show. This boosts the credibility of AI-driven diagnostic tools for clinical application.

Paper	Techniques	Advantages	Disadvantages
Naz, Z. et al. An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs.	- A pretrained ResNet model was used with tools like Grad-CAM, SHAP, and LIME, evaluated using AUC-ROC.	93-97% accuracy with LIME explanations to assist radiologists.	Tailored to specific diseases and dependent on the data.
Singh et al. Transfer learning–based ensemble SVM for COVID-19 detection	-An ensemble combining VGG16 CNN and SVM.	Achieved 95.7% accuracy using efficient transfer learning and tested multiple classifiers.	High computational cost and limited generalization.
Patrik Szepesi et al. Detection of pneumonia using CNN and deep learning.	- An enhanced VGG-16 CNN.	97.2% accuracy with high recall (97.3%) and precision (97.4%), delivering competitive results..	Pediatric-focused, data-dependent, and reliant on the quality of Kaggle data.
M. R. Karim et al. "DeepCOVIDExplainer: Explainable COVID-19 Diagnosis from Chest X-ray Images"	- VGG-16/19, ResNet, DenseNet, with interpretability methods like Grad-CAM, Grad-CAM++, and LRP.	96.12% PPV, blending deep learning with XAI, and enhanced reliability with a large dataset.	Risk of overfitting and increased complexity.

Paper	**Techniques**	**Advantages**	**Disadvantages**
Teixeira et al. Impact of Lung Segmentation on the Diagnosis and Explanation of COVID-19 in Chest X-ray Images. Sensors 2021	-U-Net CNN for lung segmentation, combined with VGG, ResNet, Inception, and interpretability methods like Grad-CAM and LIME.	Effective segmentation with the inclusion of XAI techniques.	Cross-dataset challenges and potential data bias.
Sahin et al. Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images.	- Faster R-CNN with VGG-16 and Mask R-CNN with ResNet.	93.86% accuracy with Mask R-CNN achieving a mAP of 97.72%, demonstrating robustness to dataset variations.	Limited focus on interpretability.

Table 2.1: Summary of Related Works

COVID-19 detection can be improved with deep learning models on chest X-rays and CT images has been the subject of numerous investigations. One study created a ResNet50-based model that highlighted important picture regions verified by skilled radiologists, achieved 93%–97% accuracy, and used LIME for interpretability. Another improved diagnostic accuracy by using Mask and Faster R-CNN to detect and segment objects, achieving an impressive mean average precision (mAP) of 97.72%. Additionally, LIME assisted in producing heat maps that matched expert annotations by highlighting important COVID-19 characteristics including ground-glass opacities.

Subsequent analysis of other models revealed that ResNet50 was the most successful in terms of accuracy and interpretability. With Mask R-CNN offering comprehensive lesion detection, the significance of precise CT scan segmentation was again underlined. These research demonstrate the benefits of integrating explainable AI methods with deep learning to improve COVID-19 diagnosis, increasing the dependability and credibility of AI-driven tools for clinical use.

Chapter 3

System Design

This chapter will be focusing on the system architecture describing the whole step, algorithms, and components applied. It will then depict the data flow diagram of the technology tools used as well as the chosen dataset to train the detection model. The deliverables are expected to major, with the task division and module division being prominent. This would make everything clear and in line during the development process for every important feature to successfully conclude the project by having this extensive timeline of the project.

3.1 System Architecture

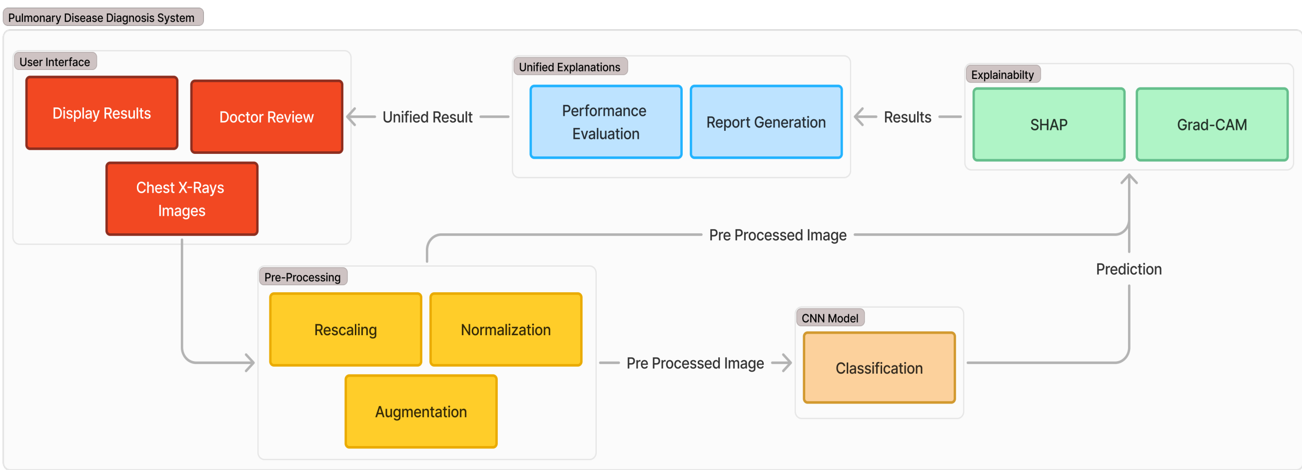


Figure 3.1: System Architecture

3.1.1 Front-End

An intuitive GUI has been designed using PySide, specifically crafted to streamline tasks for doctors. The interface allows doctors to log in safely with their username and password or create a new account if it doesn't already exist. Doctors can add new patients by entering details such as Patient ID, Name, Age, Gender, and Chest X-ray Images. Upon submission, the system easily redirects to the Pulmonary Disease Detection System, which generates a detailed diagnosis along with its explanations. All the patient information, including the diagnosis and its insights are securely stored in the patient database, ensuring efficient and comprehensive record management.

3.1.2 Back-End

After uploading the chest X-ray image from the dataset, it undergoes a pre-processing stage that includes operations like rescaling, normalization and augmentation to improve image quality and prepare it for analysis. The dataset is split for training, validation, and testing, and the pre-processed image is passed to a CNN model within a Jupyter Notebook environment. The CNN model identifies potential regions of interest and draws bounding boxes around important areas. These predictions and bounding boxes are then passed to the explainability module, where techniques such as Grad-CAM, and SHAP are applied to provide insights into the model's decision-making process. SHAP offers interpretability, while Grad-CAM produces a heatmap highlighting the most influential regions. The prediction, along with its explanation, is fused into one final result which is then evaluated by doctors. The diagnosis and explanations are stored in the database for future access. Performance evaluation modules ensures the ability to continuously discern the improvement. Finally, the diagnosis and explanation are presented to the user through the interface, allowing doctors to take any necessary follow-up actions. Jupyter Notebook is also used to monitor and visualize training progress using libraries like Keras for plotting and graphing.

3.1.3 Database

This system employs a database that can safely store and manage all the basic details of the patient such as Patient ID, Name, Age and Gender.It also stores the diagnosis obtained via the CNN model and explainability module. This will make it possible for the doctors to use historical data as reference in later treatment. The system ensures that all data is safely stored and makes it possible to update the model and improve future diagnosis processes.

3.2 Component Design

3.2.1 Data Collection and Pre-processing

The diagnostic process begins with the pre-processing of chest radiographs to ensure that they are prepared for analysis. This step involves:

- **Data Collection**: Images are gathered from datasets like COVID-Net and Covid-GAN, which include labeled radiographic data for the pulmonary conditions.
- **Preprocessing Steps**:
  - **Resizing**: Ensures all images conform to a standard size suitable for the model's input layer.
  - **Normalization**: Scales pixel values to a consistent range, improving model convergence during training.
  - **Data Augmentation**: Introduces variability (e.g., rotation, flipping, brightness adjustment) to improve the model's generalization capability.

3.2.2 Machine Learning Model

A Convolutional Neural Network is chosen for classification and diagnosis of the pulmonary diseases in chest X-rays.

Convolutional Neural Network (CNN)

CNN is one of the core models in deep learning, widely used for image processing tasks. It works by using layers like convolutional, pooling, and fully connected layers in sequence to capture and analyze spatial features from images.

- **Strengths:**

  - Efficiently identifies patterns such as edges, textures, and shapes in images.
  - Handles large-scale image data with reduced computational complexity compared to traditional methods.

- **Workflow:**
  1. **Input:** Preprocessed chest X-rays.
  2. **Convolution:** Extracts low-level features (e.g., edges, gradients).
  3. **Pooling:** Reduces spatial dimensions while retaining essential features.
  4. **Classification:** Assigns the image to one or more pulmonary disease categories.

- **Use Case:**
  CNN serves as the baseline model for understanding image features and disease classification but lacks the capability to focus on specific regions of interest in complex images.

3.2.3 Explainability Module

The explainability module bridges the gap between model predictions and human understanding by providing interpretable insights into the CNN's decision-making process. Key techniques include:

- **SHAP**
  Assigns Shapley values to different image regions, quantifying their contribution to the model's predictions.

- **Grad-CAM**
  Grad-CAM makes use of heatmaps that highlight the most influential and important regions in an image influencing a prediction. This allows clinicians to understand the reasoning behind a diagnosis, making it more transparent and relatable.

3.2.4 Post-Processing & Performance Evaluation

This module consolidates the diagnostic results and ensures they are presented in an interpretable and actionable format. Additionally, it evaluates the system's diagnostic performance to support ongoing optimization and reliability.

Post-Processing:

- Combines predictions and explanations (from the CNN and Explainability modules) into a cohesive format.
- Formats the results to highlight key findings, ensuring they are easily understandable by healthcare professionals.

Performance Evaluation:

- **Metrics Used:** Assesses the system's accuracy, specificity,precision, recall.
- **Comparison with Ground Truth:** Validates predictions against labeled data to ensure reliability and improve future iterations.
- **Continuous Improvement:** Provides quantifiable insights to benchmark the system's performance and guide enhancements.

3.2.5 User Integration Module

This module facilitates interaction between the system and healthcare professionals to streamline and enhance the decision making process.

Graphical User Interface (GUI):

- Developed using PySide for a user-friendly experience.
- Key functionalities:
  - Uploading chest X-ray images for diagnosis.
  - Viewing results with visual explanations generated by explainability techniques.
  - Downloading diagnosis reports in PDF format for clinical documentation.

Integration with Medical Systems:

Report Generation: Enables smooth generation of reports for each patient containing their diagnosis along with the visual explanations.

3.3 Algorithm Design

3.3.1 Preprocessing

During Training:

Step 1: Read the dataset from disk.

Step 2: Image to be converted to a consistent target size by resizing (e.g., 224x224).

Step 3: Normalize the pixel values by subtracting the channel-specific means [0.485, 0.456, 0.406] and then dividing by the corresponding standard deviations [0.229, 0.224, 0.225]. and standard deviation scaling.

Step 4: Apply random horizontal or vertical flips with a set probability.

Step 5: Convert the processed image into a tensor with the correct shape for the model.

Step 6: Group preprocessed images into batches for training or inference

During Prediction:

Step 1: Read the dataset from disk.

Step 2: Image to be converted to a consistent target size by resizing (e.g., 224x224).

Step 3: Normalize the pixel values by subtracting the channel-specific means [0.485, 0.456, 0.406] and then dividing by the corresponding standard deviations [0.229, 0.224, 0.225]. and standard deviation scaling.

Step 4: Convert the processed image into a tensor with the correct shape for the model.

Step 5: Group preprocessed images into batches for training or inference

3.3.2 CNN

InceptionV3

- **Step 1:** Resize the input image to 299×299 to ensure consistency.
- **Step 2:** Normalize pixel values by scaling to the range $[0, 1]$ and applying mean subtraction and standard deviation scaling.

- **Step 3:** Pass the preprocessed image through the InceptionV3 model, utilizing multiple convolutional layers with different kernel sizes to extract spatial features at various scales.
- **Step 4:** Apply global average pooling to reduce dimensionality while retaining essential feature representations.
- **Step 5:** Feed the extracted features into the fully connected layers to classify the image.
- **Step 6:** Use the softmax activation function to generate probability scores for each class.
- **Step 7:** Output the final classification result along with confidence scores.

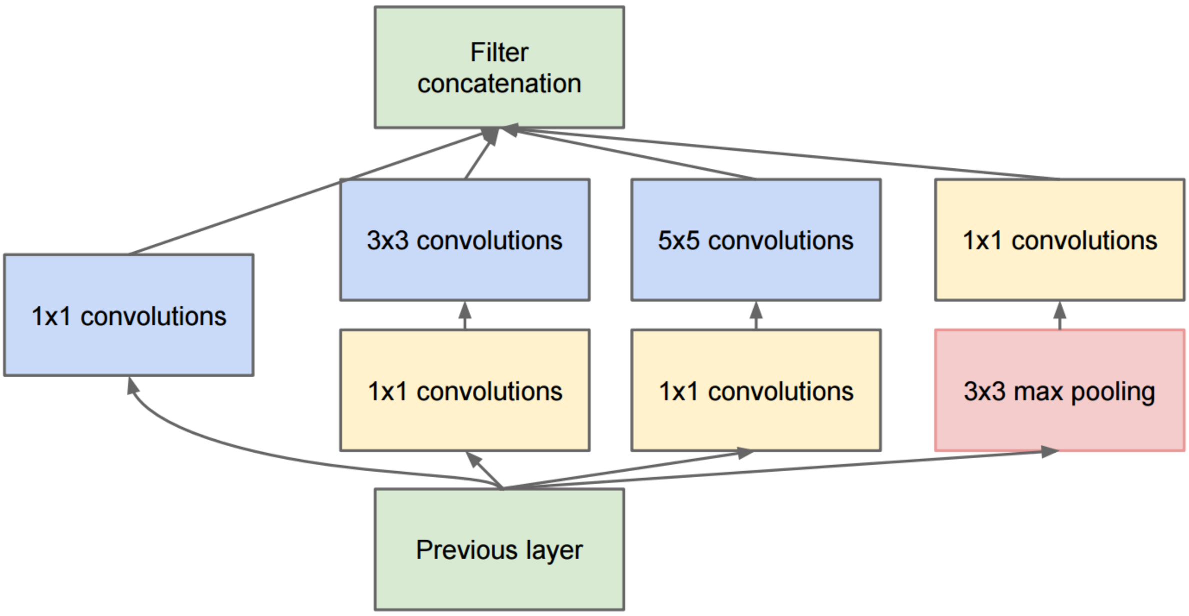


Figure 3.2: InceptionV3

ResNet-50

- **Step 1:** Resize the input image to 224×224 to ensure consistency.
- **Step 2:** Normalize pixel values by scaling to the range $[0, 1]$ and applying mean subtraction and standard deviation scaling.

- **Step 3:** Pass the preprocessed image through the ResNet-50 model, utilizing residual blocks to preserve feature integrity across layers.
- **Step 4:** Apply convolutional and pooling layers sequentially to extract hierarchical feature representations.
- **Step 5:** Use a global average pooling layer to reduce the feature map size while retaining essential information.
- **Step 6:** Feed the extracted features into the fully connected layers for classification.
- **Step 7:** Apply the softmax function to obtain probability scores for each class.
- **Step 8:** Output the final classification result along with confidence scores.

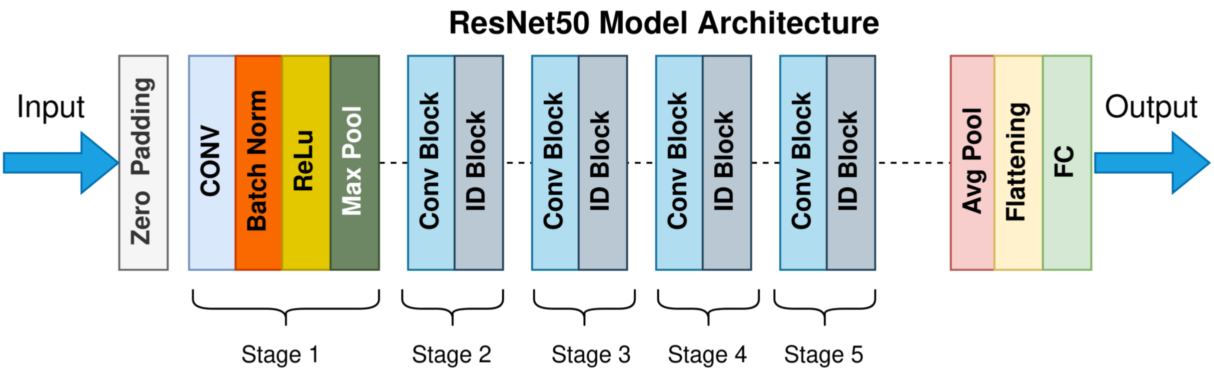


Figure 3.3: ResNET-50

3.3.3 XAI

SHAP

Step 1: Train or load a deep learning model for the desired task.

Step 2: Prepare a representative subset of the dataset for explanation purposes.

Step 3: Generate perturbed instances of the input sample and predict outputs using the model

Step 4: Use SHAP to calculate Shapley values for features across the dataset or specific samples.

Step 5: Aggregate SHAP values to understand global feature importance.

Step 6: Visualize SHAP's global explanations using bar charts or dependency plots.

Step 7: Compare SHAP results with domain knowledge to validate interpretability.

Step 8: Output both local and global explanations for comprehensive model insights.

Grad-CAM

Step 1: Pass the input image through the model and get predictions.

Step 2: Identify the target class for the explanation.

Step 3: Gradients of the target class with respect to the final convolutional layer can be obtained by performing backpropogation.

Step 4: Compute average gradients across spatial dimensions for each feature map.

Step 5: Weight the feature maps by the computed gradients.

Step 6: Sum the weighted feature maps to generate an activation map.

Step 7: Apply ReLU to retain positive values in the activation map.

Step 8: Resize the activation map to the image dimensions.

Step 9: Overlay the map on the input image.

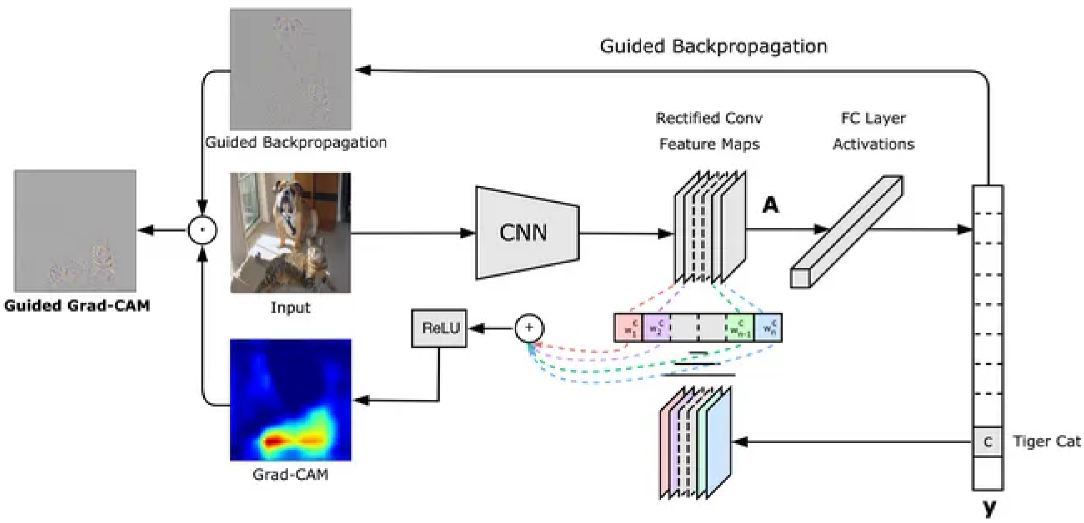


Figure 3.4: GradCAM

3.4 Data Flow Diagrams / Use Case Diagram

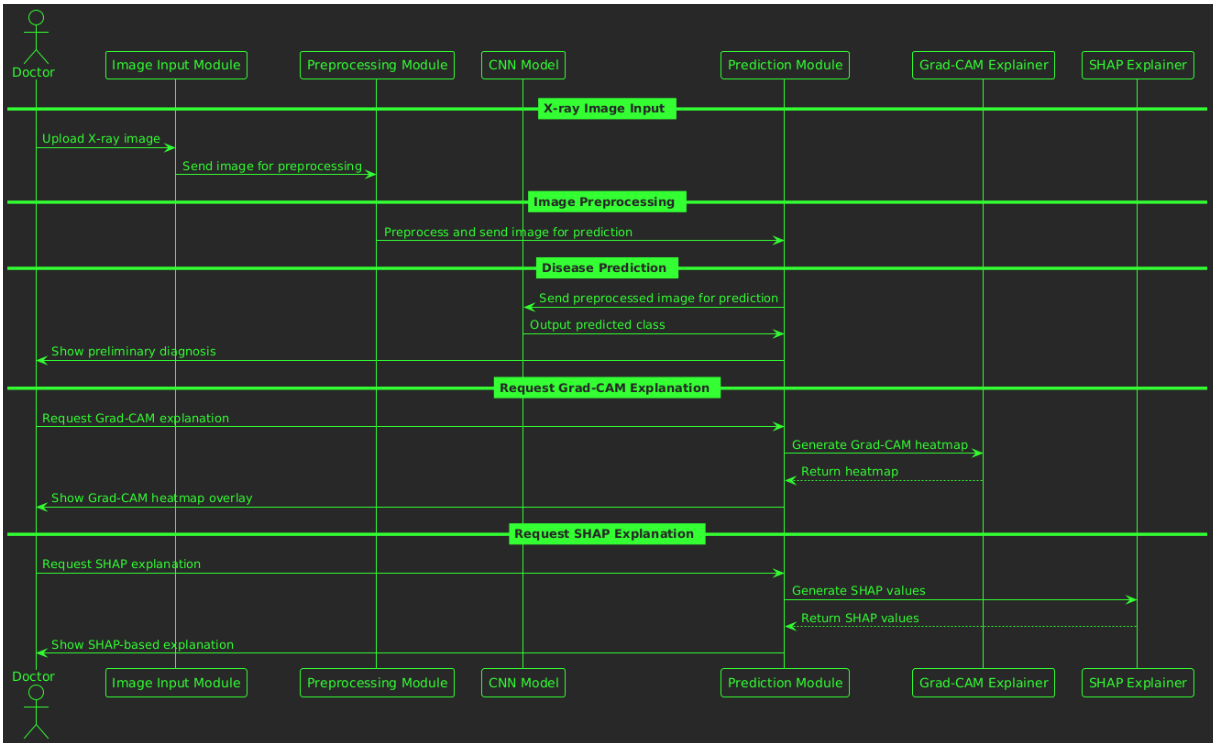


Figure 3.5: Data Flow Diagram

3.5 Tools and Technologies

3.5.1 Software Requirements

1. **Python 3.12:** Programming language for developing the AI models and application.

2. **PyTorch 2.5.1+cu121:** Deep learning frameworks used for training CNN.

3. **PySide6:** Framework for building the user interface.

3.5.2 Hardware Requirements

1. **Training:**

    - **GPU:** NVIDIA GeForce RTX 3050 or higher for accelerated model training.
    - **RAM:** Minimum 16GB for handling large datasets and model training.
    - **CPU:** Intel i7 or AMD Ryzen 7 for efficient processing.

- **Storage:** SSD with at least 100 GB of free space for the data.

2. **Execution:**

- **Processor:** Intel i5 or AMD Ryzen 5 for running the application.
- **RAM:** Minimum 8 GB for smooth operation.

3.6 Dataset Identified

The COVID-19 Chest X-ray dataset was compiled from multiple publicly available sources to address the limited availability of a large, specific dataset for COVID-19 diagnosis. The dataset consists of a diverse collection of X-ray images sourced from various platforms, ensuring comprehensive coverage of pulmonary conditions.

The dataset includes a total of 613 COVID-19 X-ray images obtained from GitHub, Radiopaedia, The Cancer Imaging Archive (TCIA), and the Italian Society of Radiology (SIRM). Additionally, an augmented dataset containing 912 pre-processed COVID-19 images was collected from Mendeley. To ensure balanced classification, the dataset also incorporates 1525 pneumonia X-ray images and 1525 normal X-ray images sourced from the Kaggle repository and the NIH dataset.

This dataset provides a diverse set of labeled X-ray images, facilitating deep learning-based diagnostic models for distinguishing between COVID-19, pneumonia, and normal cases.

3.7 Module Division and Work Breakdown

3.7.1 Module Division

Preprocessing

The module handles the acquisition and preparation of chest X-ray images. Data is collected from publicly available datasets, which include diverse pulmonary disease labels. Images are standardized to a uniform size and various data augmentation techniques like rotation and flipping are applied to enhance the model's robustness.

Feature Extraction and Detection Model

For analyzing chest X-ray images in this research, two CNN (Convolutional Neural Networks) models are used: InceptionV3 and ResNet-50. These models are built to automatically extract important features from the X-ray images and help in the detection of pulmonary diseases such as COVID-19 and pneumonia.

InceptionV3 is very good at extracting features at different scales and has parallel convolutional layers to capture spatial hierarchies and detailed specifics of images, and ResNet-50 makes use of deep residual learning to improve gradient flow and make the model learn complex patterns without degrading performance.

All features of interest must first be extracted before processing the input images of X-rays. This is done by passing the images through a system of convolution and pooling layers in the model. The features are then fed into layers that are fully interconnected for appropriate classification. In this case, the model has to predict if the image is normal, pneumonia, or covid positive. These architectures work together to provide reliable and accurate medical diagnosis by ensuring proper extraction of the features and swift detection of diseases.

Explainability Module

This module ensures interpretability of the system's predictions using techniques like SHAP, and Grad-CAM. These methods provide visual and numerical explanations, highlighting the regions and features contributing most to the model's decisions.

Post-Processing

The explanations and diagnostic results are presented in a way that is easy for clinicians to understand. To ensure reliability and continuous improvement, the system's performance is evaluated using metrics such as accuracy, precison and recall.

User Integration

The system features a user-friendly graphical interface developed in PySide, allowing doctors to upload X-rays, view results, and download reports.

3.7.2 Work Breakdown

Task	**Team Member**
Data Collection	Nandakishore T.J., Niranjana S. Nair, Nithin K.Cimu, Shanker Menon
Data Pre-processing	Shanker Menon
CNN	Shanker Menon, Nithin K. Cimu
SHAP	Niranjana S. Nair, Nithin K. Cimu
GradCAM	Nandakishore T.J.
User Interface	Nandakishore T.J., Niranjana S. Nair

Table 3.1: Team Work Breakdown

3.8 Key Deliverables

The system that was developed is a detection model that comprises the following features:

1. **Object Detection with CNN (Convolutional Neural Network)** They convolutions onto the Chest X-Ray Images and the diseases are identified with high accuracy.

2. **Explainability Module:** Deep advanced techniques have been inculcated for explainability. Some are - Grad-CAM and SHAP on this model in order to present a more readable and clear trans-perv outcome that explains through its very reason deciding decision that makes facilitation more smooth. Doctors check predictions and take much of the uncertainty of the decision making model.

3. **Predictive Classification:**It predicts the probability of pneumonia and COVID-19 based on chest X-ray image analysis and gives doctors all the information that will be used to make a diagnosis for the infection,

4. **Doctor Review Interface:** Presents model results, bounding box visualization, and explanations through an easy-to-use interface so that doctors can simply review results, validated diagnoses, and correct decisions.

5. **Secure Database:** A secure database storing all images of chest X-rays, model predictions, coordinates of the bounding box, explanations, confirmed diagnoses, and any other appropriate information so that data can be kept confidential and the rules on data governance and regulation in healthcare strictly followed.

3.9 Summary of the Chapter

This chapter includes a general overview of the system architecture, detailing various modules of the system. These modules are elaborative on showing the technologies used in implementation and are based on tools and frameworks used, and that of the selection of the data set made for training/testing and validation models for detection work. It even breaks down the work structure, providing information concerning the activities involved in the production of the system. This gives a detailed view of how tasks are divided across different members of the group and provides an estimated timeline for the completion of each phase, thereby ensuring a systematic approach to the development process.

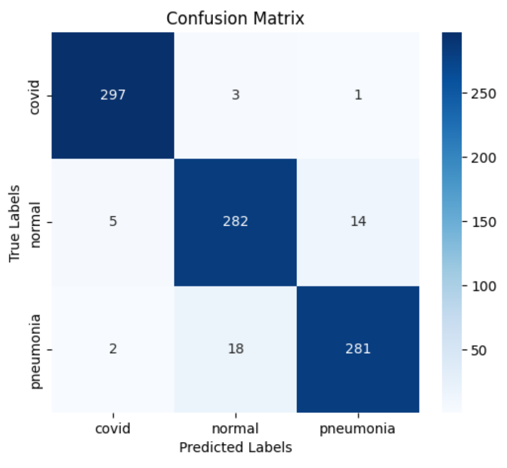
Chapter 4

Results and Discussions

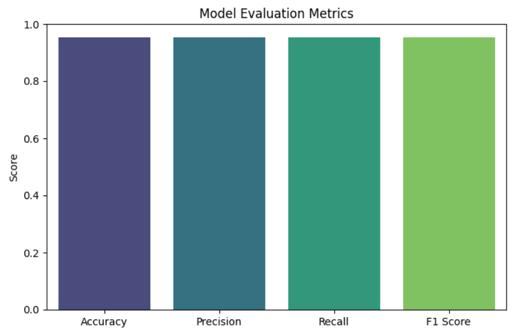
4.1 CNN Model Performance Metrics

This section provides a thorough comparison of three convolutional neural network (CNN) models: InceptionV3, ResNet-50, and the proposed custom CNN, in terms of their classification accuracy on chest X-ray images. The comparison involves confusion matrices and standard evaluation metrics like accuracy, precision, recall, and F1-score.

4.1.1 InceptionV3



(a) Inception Confusion Matrix



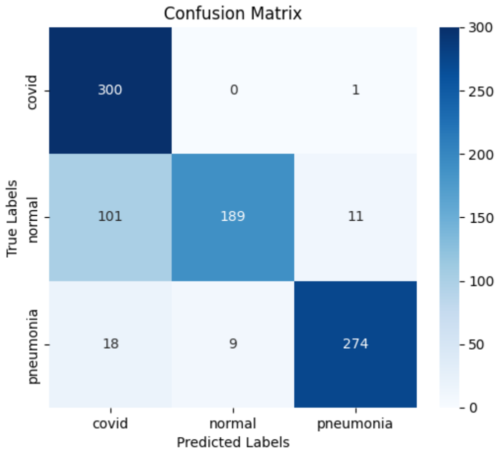
(b) Inception Evaluation Metrics

Figure 4.1: Inception Model Evaluation Results

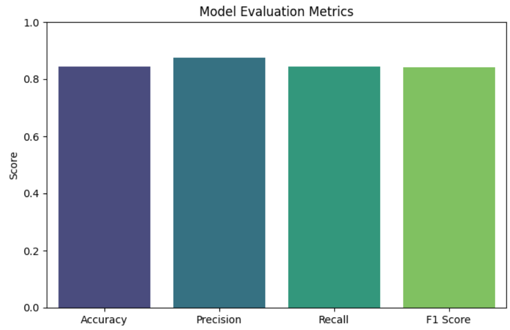
- Figure 4.1(a) shows the confusion matrix of the InceptionV3 model, which reflects its ability to classify between different classes accurately.
- As shown in Figure 4.1(b) and Table 4.1, InceptionV3 works best overall compared to the other models with an accuracy of 96.57%, precision of 0.9678, recall of 0.9657,

and F1-score of 0.9656.

4.1.2 ResNET-50



(a) ResNET-50 Confusion Matrix

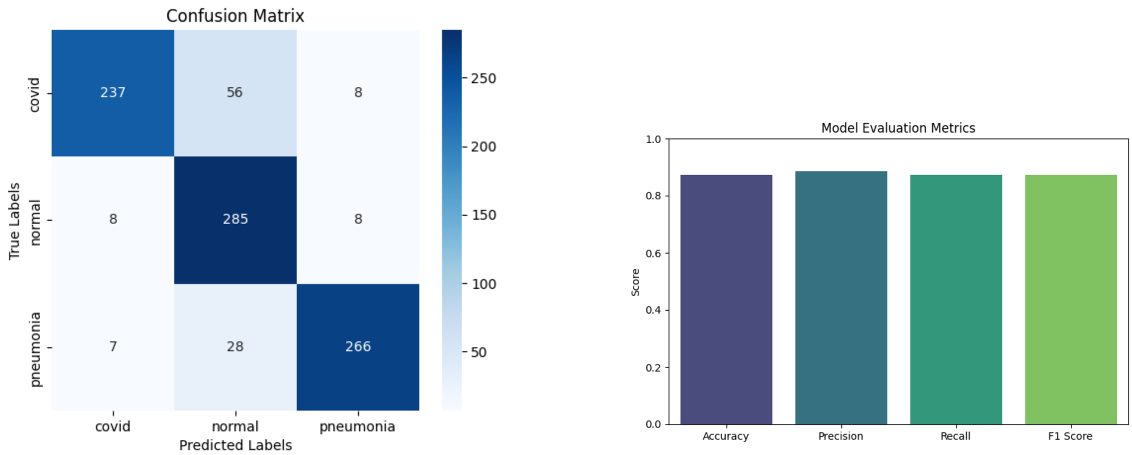


(b) ResNET-50 Evaluation Metrics

Figure 4.2: ResNET-50 Model Evaluation Results

- The ResNet-50 confusion matrix in Figure 4.2(a) indicates a high performance, although slightly less than InceptionV3.
- As shown in Figure 4.2(b), ResNet-50 has an accuracy of 90.92%, precision of 0.9150, recall of 0.9092, and an F1-score of 0.9089.
- The results indicate that ResNet-50 continues to perform strongly with only a small decline in metrics, so it is a good alternative where computational efficiency is more important than absolute performance.

4.1.3 Proposed custom CNN



(a) Proposed CNN Confusion Matrix (b) Proposed CNN Evaluation Metrics

Figure 4.3: Proposed CNN Model Evaluation Results

- The proposed custom CNN is tested in Figure 4.3, where sub-figure (a) displays its confusion matrix and sub-figure (b) its metrics.
- As Table 4.1 indicates, the suggested model has an accuracy of 87.26
- Although it performs worse than the pre-trained models, the custom CNN still provides competitive performance and is especially useful in low-computational-resource environments or where a light, interpretable model is needed.

Model	Accuracy	Precision	Recall	F1-Score
Proposed CNN	87.26%	0.8854	0.8726	0.8735
ResNET-50	90.92%	0.9150	0.9092	0.9089
InceptionV3	96.57%	0.9678	0.9657	0.9656

Table 4.1: Evaluation metrics of Proposed CNN model

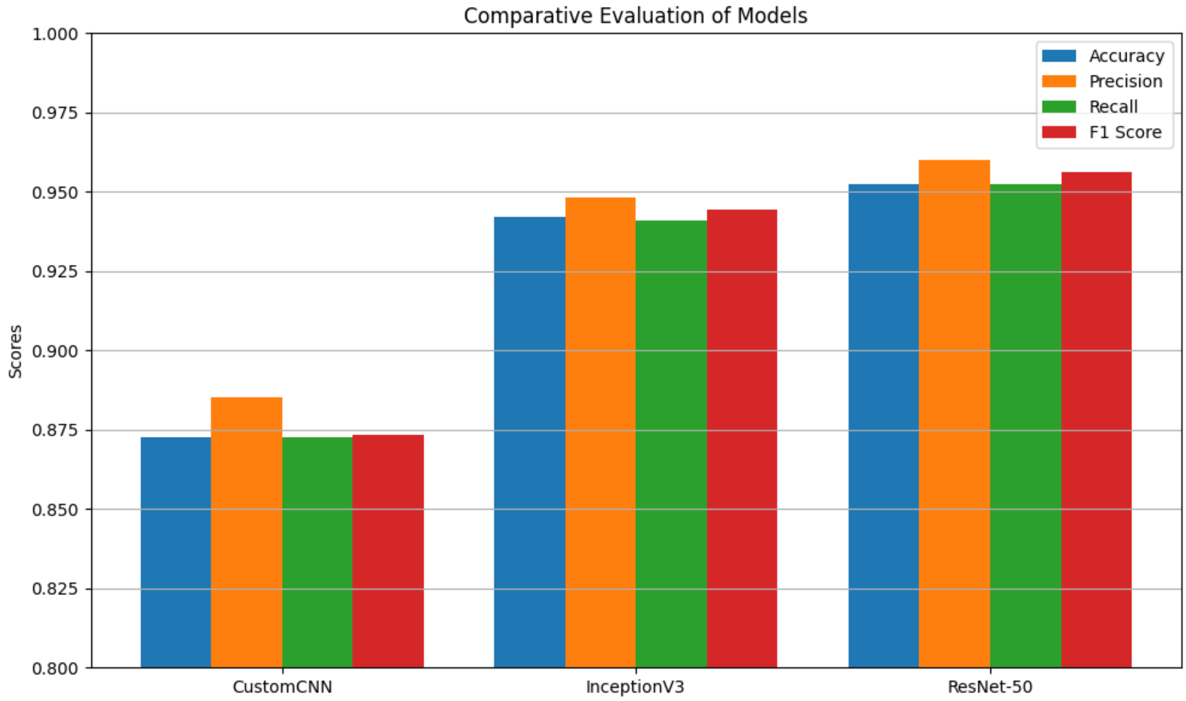


Figure 4.4: Model Comparisons

4.2 Model Evaluation Summary

The performance of the suggested custom Convolutional Neural Network (CNN) model was evaluated through traditional evaluation measures such as Accuracy, Precision, Recall, and F1-score. These measures reflect a comprehensive analysis of how well the model classifies chest X-ray images correctly into their corresponding categories.

On the training set, the model obtained an accuracy of 89.89%, precision of 0.9136, recall of 0.8989, and an F1-score of 0.9004. These metrics confirm that the model was able to learn strong features during training and accurately identified the labeled samples.

When tested on the validation dataset, the model achieved an accuracy of 87.26%, precision of 0.8854, recall of 0.8726, and an F1-score of 0.8735. These numbers, while marginally lower than the training data, are very close to it, indicating that the model is generalizing very well and not overfitted.

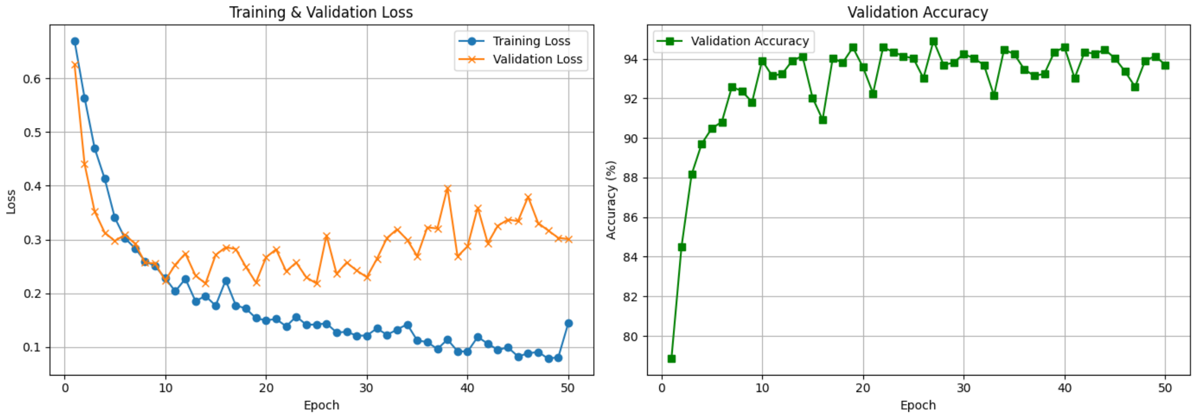


Figure 4.5: Training/Validation Loss & Validation Accuracy

4.3 Analysis and Interpretation

While the pre-trained models, including ResNet-50 and InceptionV3, had greater overall validation set accuracy (around 95.24% and 96.57% respectively), a closer inspection indicates that the suggested Custom CNN model provided superior generalization performance. This was attested to by a less significant difference between the training and validation scores — an indicator of lower overfitting. Even though the custom model had a slightly lower validation accuracy of 87.26%, its balanced precision (0.8854), recall (0.8726), and F1-score (0.8735) indicate more even performance across classes. However, the higher-performing pre-trained models might have overfitted just a bit to the training data, as indicated by training accuracy values close to 99.84

. Thus, the moderate complexity of the custom model allowed it to provide more consistent predictions on new data, better serving the purpose of the project to develop a strong disease classification system.

4.4 Chapter Summary

Here, the accuracy of three models, i.e., InceptionV3, ResNet-50, and a custom CNN, was compared in terms of accuracy, precision, recall, and F1-score in this chapter. Even though InceptionV3 and ResNet-50 were more accurate on the validation set, the Custom CNN better generalized. Its training and validation accuracy were near equal, i.e., less overfitted and generalizing well on new data. This renders the Custom CNN appropriate for real chest X-ray classification in real-world scenarios despite it being lightweight.

Chapter 5

Conclusion & Future Scope

5.1 Conclusion

This project integrates XAI in the diagnosis of pulmonary disease, thus filling the gap between state-of-the-art AI models and real-world clinical applications. Using advanced deep learning techniques this system is able to accurately and efficiently identify pneumonia and COVID-19 in chest radiographs. Explainability methods such as Grad-CAM, and SHAP further enhance the system's transparency, enabling medical professionals to understand and trust the AI's decision-making process. The graphical user interface (GUI) provides an interactive platform for clinicians, streamlining their workflow and improving patient care. This research emphasizes not only diagnostic accuracy but also the critical role of interpretability in fostering acceptance of AI systems in healthcare, making it a significant step towards smarter, patient-centered diagnostics.

5.2 Future Scope

- The framework developed in this project can be extended to other domains of medical imaging such as CT scans, MRIs, and ultrasound images. This will widen its scope more than pulmonary diseases to include neurological, cardiac, and orthopedic conditions.
- This model can be optimized for real-time inference, thereby allowing its use in emergency and outpatient settings, where rapid diagnostics are critical.
- Continued emphasis on ethical AI practices, such as bias reduction in training datasets and fairness across demographic groups, will strengthen the credibility and inclusivity of the system in clinical environments.

References

[1] M. Singh, S. Bansal, S. Ahuja, R. Dubey, B. Panigrahi, and N. Dey, "Transfer learning based ensemble support vector machine model for automated covid-19 detection using lung computerized tomography scan data," 05 2020.

[2] M. Karim, T. Döhmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "Deepcovidexplainer: Explainable covid-19 diagnosis from chest x-ray images," in *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2020, pp. 1034–1037.

[3] L. O. Teixeira, R. M. Pereira, D. Bertolini, L. S. Oliveira, L. Nanni, G. D. C. Cavalcanti, and Y. M. G. Costa, "Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images," *Sensors*, vol. 21, no. 21, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/21/7116

[4] P. Szepesi and L. Szilagyi, "Detection of pneumonia using convolutional neural networks and deep learning," *Biocybernetics and Biomedical Engineering*, vol. 42, 08 2022.

[5] M. Sahin, H. Ulutaş, E. Yuce, and M. Erkoç, "Detection and classification of covid-19 by using faster r-cnn and mask r-cnn on ct images," *Neural Computing and Applications*, vol. 35, pp. 1–15, 03 2023.

[6] Z. Naz, M. U. G. Khan, T. Saba, A. Rehman, H. Nobanee, and S. A. Bahaj, "An explainable ai-enabled framework for interpreting pulmonary diseases from chest radiographs," *Cancers*, vol. 15, no. 1, 2023. [Online]. Available: https://www.mdpi.com/2072-6694/15/1/314

[7] M. M. Ahsan, K. D. Gupta, M. M. Islam, S. Sen, M. L. Rahman, and M. Shakhawat Hossain, "Covid-19 symptoms detection based on nasnetmobile with explainable ai using various imaging modalities," *Machine Learning and Knowledge Extraction*, vol. 2, no. 4, pp. 490–504, 2020.

[8] L. Zou, H. L. Goh, C. J. Y. Liew, J. L. Quah, G. T. Gu, J. J. Chew, M. P. Kumar, C. G. L. Ang, and A. W. A. Ta, "Ensemble image explainable ai (xai) algorithm for severe community-acquired pneumonia and covid-19 respiratory infections," *IEEE Transactions on Artificial Intelligence*, vol. 4, no. 2, pp. 242–254, 2023.

[9] A. S. Albahri, R. A. Hamid, J. K. Alwan, Z. Al-Qays, A. Zaidan, B. Zaidan, A. Albahri, A. H. AlAmoodi, J. M. Khlaf, E. Almahdi *et al.*, "Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (covid-19): a systematic review," *Journal of medical systems*, vol. 44, pp. 1–11, 2020.

[10] O. Albahri, A. Zaidan, A. Albahri, B. Zaidan, K. H. Abdulkareem, Z. Al-qaysi, A. Alamoodi, A. Aleesa, M. Chyad, R. Alesa, C. Lim, M. M. Lakulu, A. Ibrahim, and N. A. Rashid, "Systematic review of artificial intelligence techniques in the detection and classification of covid-19 medical images in terms of evaluation and benchmarking: Taxonomy analysis, challenges, future solutions and methodological aspects," *Journal of Infection and Public Health*, vol. 13, no. 10, pp. 1381–1396, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187603412030558X

[11] Y. Pathak, P. Shukla, A. Tiwari, S. Stalin, S. Singh, and P. Shukla, "Deep transfer learning based classification model for covid-19 disease," *IRBM*, vol. 43, no. 2, pp. 87–92, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1959031820300993

[12] L. Wang and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," 2020. [Online]. Available: https://arxiv.org/abs/2003.09871

[13] W. H. Self, D. M. Courtney, C. D. McNaughton, R. G. Wunderink, and J. A. Kline, "High discordance of chest x-ray and computed tomography for detection of pulmonary opacities in ed patients: implications for diagnosing pneumonia," *The American Journal of Emergency Medicine*, vol. 31, no. 2, pp. 401–405, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735675712004639

[14] J. Krause, A. Perer, and K. Ng, "Interacting with predictions: Visual inspection of black-box machine learning models," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 5686–5697. [Online]. Available: https://doi.org/10.1145/2858036.2858529

[15] Y. Li, Z. Zhang, C. Dai, Q. Dong, and S. Badrigilan, "Accuracy of deep learning for automated detection of pneumonia using chest x-ray images: A systematic review and meta-analysis," *Computers in Biology and Medicine*, vol. 123, p. 103898, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001048252030247X

[16] S. Podder, S. Bhattacharjee, and A. Roy, "An efficient method of detection of covid-19 using mask r-cnn on chest x-ray images," *AIMS Biophys*, vol. 8, no. 3, pp. 281–290, 2021.

Appendix A: Presentation

Final Presentation

Explainable AI for Pulmonary Disease Diagnosis

Team Members:
Nandakishore T J (U2102145)
Niranjana S Nair (U2102159)
Nithin Kurisingal Cimu (U2102161)
Shanker Menon (U2103193)

Project Guide:
Ms. Dincy Paul
Assistant Professor
Department of CSE

Contents

Introduction

- Pulmonary diseases such as pneumonia and COVID-19 pose serious health risks and require timely and accurate diagnosis. Chest X-ray imaging remains one of the most accessible diagnostic tools, but interpreting these images often requires expert radiological analysis.
- The goal is to support clinicians with accurate diagnostics and interpretable insights, improving decision-making and fostering trust in AI-assisted medical tools.

Problem Definition

To develop a diagnostic tool to provide accurate and interpretable classification of pulmonary diseases from chest radiographs, thereby improving clinical decision-making.

Novelty and Innovatives

- **Improve Diagnostic Accuracy and Early Detection**
  Explainable AI (XAI) helps detect subtle patterns in pulmonary disease data, enabling early diagnosis of conditions like Covid or Pneumonia, which can lead to better patient outcomes and more effective treatment.
- **Transparency and Trust in AI**
  In healthcare, it's crucial for doctors and patients to understand how AI systems reach their decisions. XAI provides this transparency, making AI predictions more trustworthy and actionable for clinicians.

Project Objective

- Accurately diagnose pulmonary diseases using AI models.
- Provide transparent and interpretable insights into the AI's decision-making process.
- Ensure trust and actionable results for clinicians and patients.
- Enhance collaboration between AI systems and healthcare professionals.

Scope of Implementation

- Disease Focus: Limited to pulmonary diseases like pneumonia and COVID-19 based on available labeled datasets.
- Image Modality: Only Chest X-ray images (not CT scans or MRIs).
- Model Types Used: Implementation of custom CNN, ResNet-50, and InceptionV3.
- Explainability Tools: Visual explanations using Grad-CAM and SHAP to support transparency in diagnosis.
- Prediction Output: Predicts disease class and provides region-based + feature-level explanations.
- User Interaction: Interface or CLI allowing doctors to upload X-rays and view prediction + explanations.

Literature Survey

Title	Dataset	Methodology	Result	Advantages	Disadvantages
Naz, Z. et al. An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs[1]	COVID-CT dataset and COVIDNet	A pretrained ResNet model was used with tools like Grad-CAM, SHAP, and LIME, evaluated using AUC-ROC.	Achieved 93–97% accuracy	Provides visual and feature-level explanations using multiple XAI tools; supports radiologists	Tailored to specific diseases and dependent on the data
Singh et al. Transfer learning–based ensemble SVM for COVID-19 detection[2]	COVID-CT dataset, the Italian Society of Medical and Interventional Radiology	An ensemble combining VGG16 CNN and SVM	Achieved 95.7% accuracy	Effective use of transfer learning; strong classification performance	High computational cost and limited generalization

Literature Survey

Title	Dataset	Methodology	Result	Advantages	Disadvantages
Patrik Szepesi et al. Detection of pneumonia using CNN and deep learning[3]	Lung X-ray images taken from Guangzhou Women and Children's Medical Center	An enhanced VGG-16 CNN	97.2% accuracy; 97.3% recall; 97.4% precision	Competitive metrics; good recall and precision	Pediatric-focused, data-dependent, and reliant on the quality of Kaggle data
M. R. Karim et al. "DeepCOVIDExplainer: Explainable COVID-19 Diagnosis from Chest X-ray Images"[4]	Korean Journal of Radiology	VGG-16/19, ResNet, DenseNet, with interpretability methods like Grad-CAM, Grad-CAM++, and LRP	96.12% PPV; reliable with XAI	Robust ensemble; strong interpretability methods	Risk of overfitting and increased complexity

Literature Survey

Title	Dataset	Methodology	Result	Advantages	Disadvantages
Teixeira et al. Impact of Lung Segmentation on the Diagnosis and Explanation of COVID-19 in Chest X-ray Images. Sensors 2021[5]	COVID-19 Chest X-ray datasets	U-Net CNN for lung segmentation, combined with VGG, ResNet, Inception, and interpretability methods like Grad-CAM and LIME	F1-Score of 0.92 and 0.94	Effective segmentation with the inclusion of XAI techniques	Cross-dataset challenges and potential data bias
Sahin et al. Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images[6]	CT images from Yozgat Bozok University Faculty of Medicine	VGG-16/19, ResNet, DenseNet, with interpretability methods like Grad-CAM, Grad-CAM++, and LRP	97.72% accuracy with ResNET-50	93.86% accuracy with Mask R-CNN achieving a mAP of 97.72%, demonstrating robustness to dataset variations	Limited focus on interpretability

Methodology

- **Dataset Collection:**
  - Dataset collected from kaggle:
    - Covid-GAN and Covid-Net mini Chest X-ray
    - COVID-19 Xray Dataset
    - COVID19_Pneumonia_Normal_Chest_Xray_PA_Dataset
    - Chest Xray for covid-19 detection
- **Data Preprocessing**:
  - Collect and preprocess chest X-ray data.
  - Apply image augmentation for variability.
- **CNN Model**:
  - Use trained CNN to detect and classify pulmonary diseases.

Methodology

Pre Trained CNN Models

- **Resnet50**:
  - ResNet-50 is a deep learning model with 50 layers.
  - Uses skip connections to prevent training issues.
  - Extracts low-level features (edges, textures) and high-level features (lung abnormalities).
  - Pretrained on ImageNet, making it easier to adapt for medical images

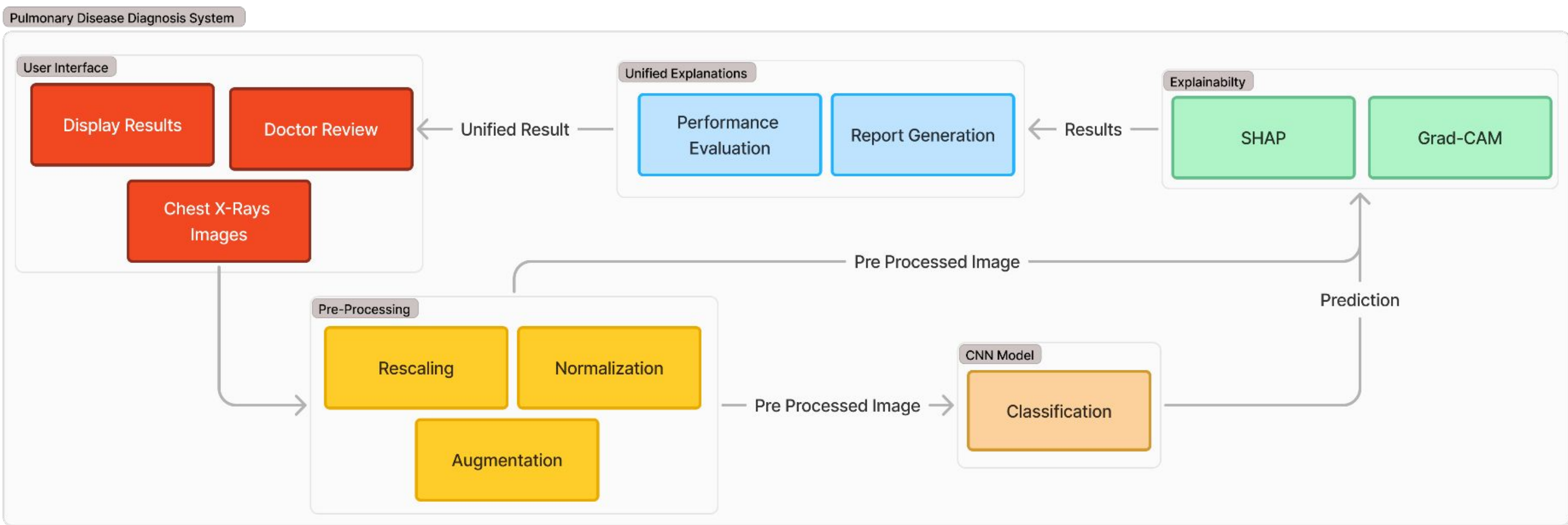
Methodology

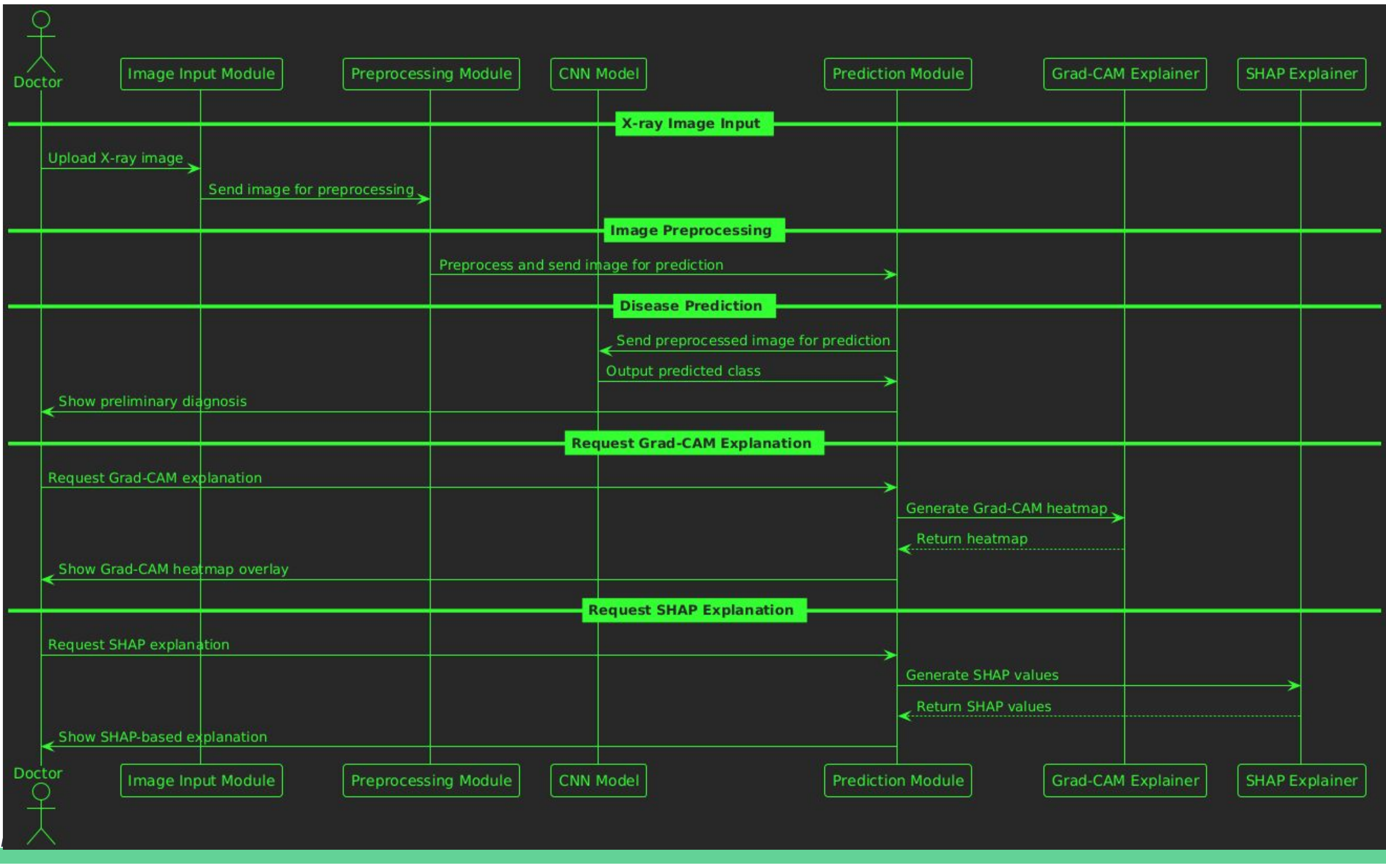
- **InceptionV3:**
  - InceptionV3 is a deep learning model with 48 layers.
  - More efficient than traditional CNNs due to factorized convolutions.
  - Includes an auxiliary classifier during training to improve gradient flow.
  - Two outputs during training (main classifier + auxiliary classifier) help with better learning.
  - Single output during testing, as the auxiliary classifier is removed.
  - Pretrained on ImageNet, making it adaptable for medical image analysis.

Methodology

- **Explainability with SHAP and Grad-CAM:**
  - SHAP: Analyze feature contributions (image regions) to the model's predictions.
  - Grad-CAM: Generates heatmaps to visualize which areas of the image the CNN focused on when making a decision.
- **Evaluation:**
  - Assess model performance using metrics (Accuracy, F1, etc.).
  - Validate explanations with clinical experts.
- **User Interface:**
  - Create a visual interface for doctors to view predictions and explanations.

System Architecture





Modules

1. **Image Input Module**

➔ This module is responsible for collecting medical imaging data, specifically lung X-ray scans, from the doctor using the GUI.

2. **Image Pre-Processing Module**

➔ Once the X-ray images are acquired, they need to be preprocessed to prepare them for the CNN model.

- This includes tasks like resizing, normalization and data augmentation (e.g., rotation, flipping) to improve model generalization during training.
- Resizing and normalization is done for other uses like explanation.

3. Convolutional Neural Network (CNN) Module

- This module is the core of the detection system where the preprocessed images are fed into a CNN model.
- The CNN first extracts important features from the image.
- These features help the model to find and focus on key areas that may show signs of disease.
- Then, the CNN classifies what's inside those areas, like detecting pneumonia or COVID.
- The model learns this by training on images with known disease labels.

4. Explainability Module

- Explainability techniques such as SHAP and Grad-CAM are used to interpret the CNN predictions.
- SHAP (Shapley Additive Explanations) is based on Shapley values, where each feature's contribution to the prediction is calculated by considering all possible combinations of features.
- Gradients are partial derivatives that indicate how much a change in a certain parameter will affect the model's output.
- In Grad-CAM, the gradients represent how important each feature map is for the model's prediction of a particular class.

5. **Post-Processing Module**
➔ The post-processing module processes the predictions and the explanations provided by the CNN and Explainability modules.
➔ It prepares the output and generates the report of the diagnosis.

6. **User Interface Module**
➔ This module provides a graphical user interface (GUI) for doctors, radiologists, or healthcare professionals to interact with the system.
➔ It displays the visual explanations and other relevant information in a user-friendly manner.

7. **Performance Evaluation Module**
➔ This module evaluates the system's performance based on various metrics such as accuracy, precision, recall, F1 score, etc.
➔ It compares the predicted results against ground truth labels to track the system's diagnostic capabilities.

Assumptions

- **Availability of Quality Data**
  High-quality, representative X-ray scan data for various pulmonary diseases is available for training, validation, and testing.
- **Preprocessing Effectiveness**
  Image preprocessing techniques will retain important diagnostic information and improve model performance.
- **Explainability Methods Provide Meaningful Insights**
  SHAP, and Grad-CAM will generate interpretable and medically useful explanations that are understandable and trusted.

Assumptions

- **Sufficient Computational Resources**
  Adequate computational power (e.g., GPUs) is available to train the cnn and run explainability algorithms without performance bottlenecks.
- **Generalization of the AI Model**
  The CNN model will generalize well to new, unseen data, ensuring robust performance across different patient cases and disease types.

Work Breakdown and Responsibilities

Nandakishore T J:

GRADCAM,

CNN Model Training

Niranjana S Nair:

GUI,

SHAP

Nithin K Cimu:

SHAP,

CNN Model Training

Shanker Menon:

Data Preprocessing,

CNN Model Training

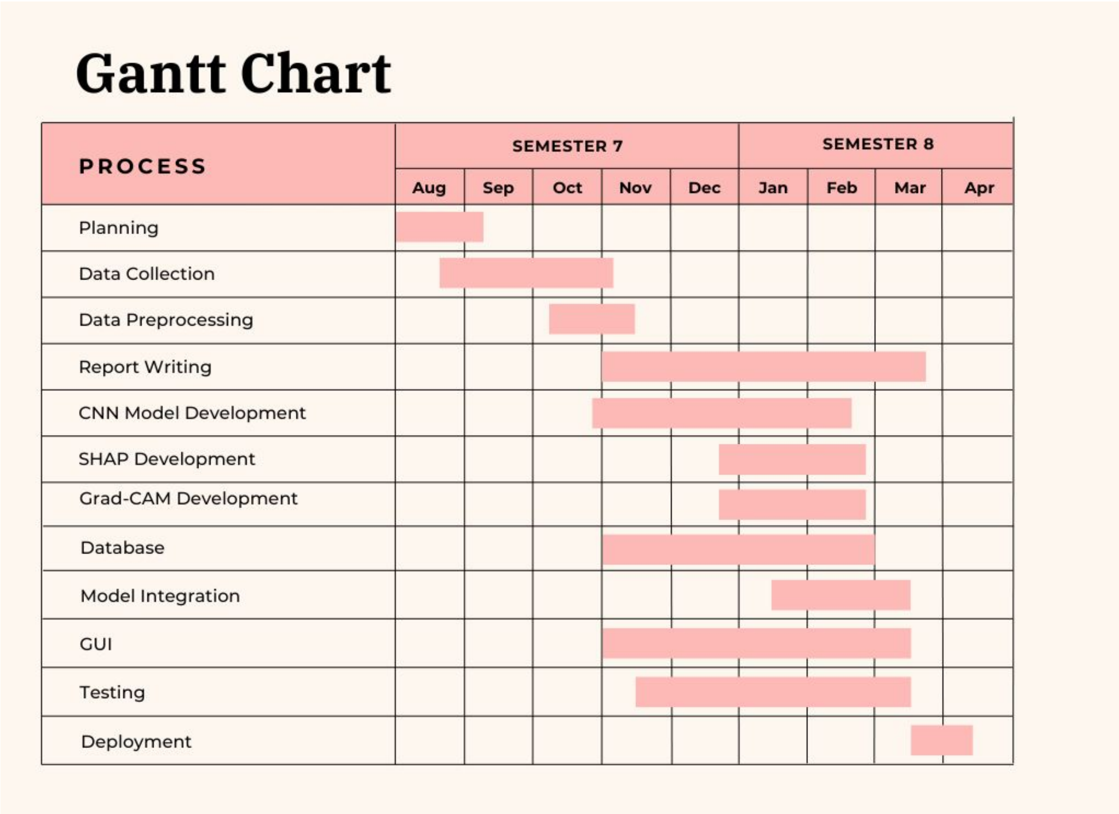
Software Requirements

- **Python 3.12:** Programming language for developing the AI models and application.
- **PyTorch 2.5.1+cu121:** Deep learning frameworks used for training CNN.
- **PySide6:** Framework for building the user interface.

Hardware Requirements

- **Training**:
    - **GPU**: NVIDIA GeForce RTX 3060 or higher for accelerated model training.
    - **RAM**: Minimum 16GB for handling large datasets and model training.
    - **CPU**: Intel i7 or AMD Ryzen 7 for efficient processing.
    - **Storage**: SSD with at least 100 GB of free space for the data.

- **Execution**:
    - **Processor**: Intel i5 or AMD Ryzen 5 for running the application.
    - **RAM**: Minimum 8 GB for smooth operation.



Risks and Challenges

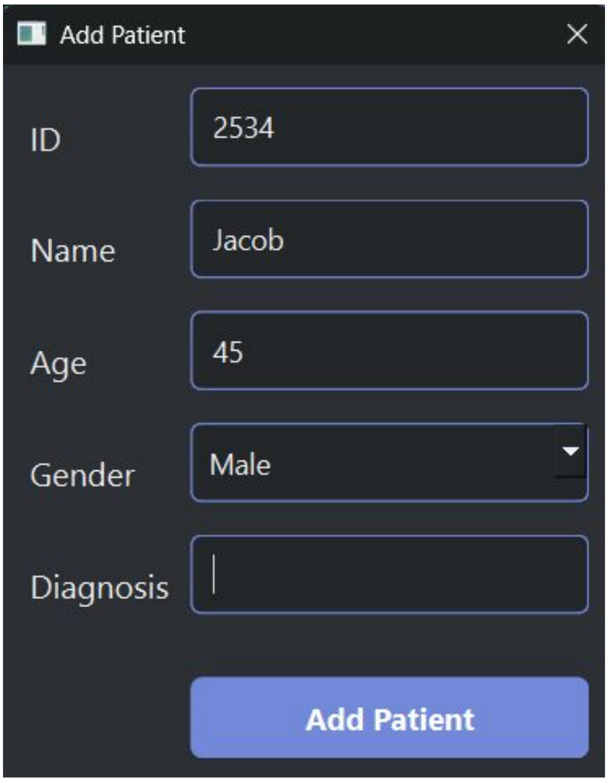
- **Data Quality and Availability**
  Insufficient or poor-quality X-ray images can hinder model performance, leading to inaccurate predictions.
- **Model Overfitting**
  The AI model may overfit to the training data, resulting in poor generalization to unseen data.
- **Integration with Clinical Workflow**
  - Resistance from medical staff
  - Require significant changes in current practices, impacting adoption and usage.

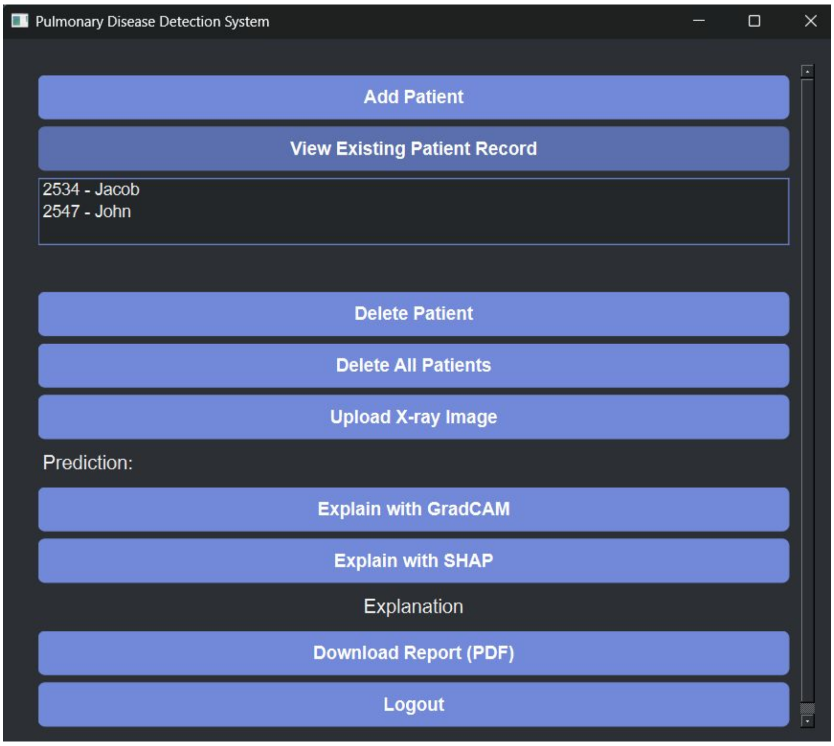
Risks and Challenges

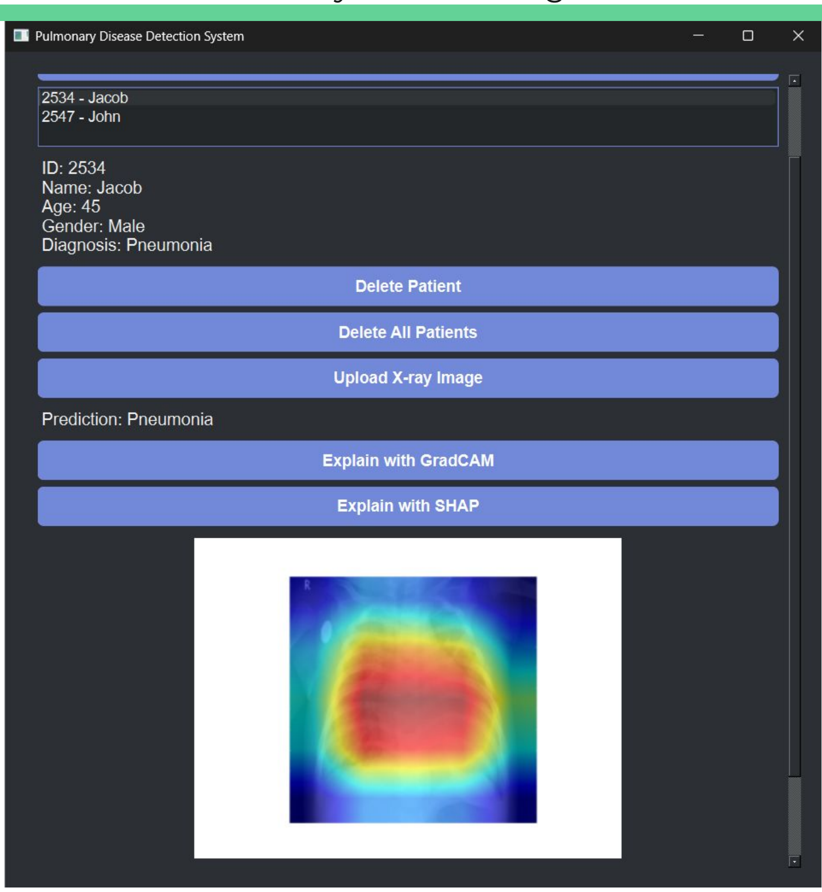
- **Ethical Concerns**
  - Bias in training data that could lead to unequal treatment across different patient populations.
- **Maintenance and Updates**
  - Keep the system aligned with the latest medical knowledge and technological advancements
  - Resource-intensive and require ongoing support.

Results

- Python App to predict the disease and give the explanations for the prediction has been developed.
- CNN models like ResNET-50, InceptionV3 and a custom model have been trained to predict the disease.
- SHAP and Grad-CAM have been implemented to give explanations for the predictions


Add Patient
ID
2534
Name
Jacob
Age
45
Gender
Male
Diagnosis
Add Patient


Pulmonary Disease Detection System
Add Patient
View Existing Patient Record
2534 - Jacob
2547 - John
Delete Patient
Delete All Patients
Upload X-ray Image
Prediction:
Explain with GradCAM
Explain with SHAP
Explanation
Download Report (PDF)
Logout


Pulmonary Disease Detection System
2534 - Jacob
2547 - John
ID: 2534
Name: Jacob
Age: 45
Gender: Male
Diagnosis: Pneumonia
Delete Patient
Delete All Patients
Upload X-ray Image
Prediction: Pneumonia
Explain with GradCAM
Explain with SHAP

Pulmonary Disease Detection System

Patient Diagnosis Report

Patient Information

ID: 2534
Name: Jacob
Age: 45
Gender: Male
Diagnosis: Pneumonia

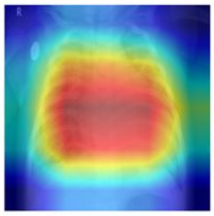
Diagnosis Details

Diagnosis: Pneumonia

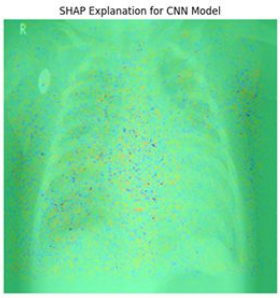
Original X-ray Image

AI Explanation Visualizations

Grad-CAM Explanation:



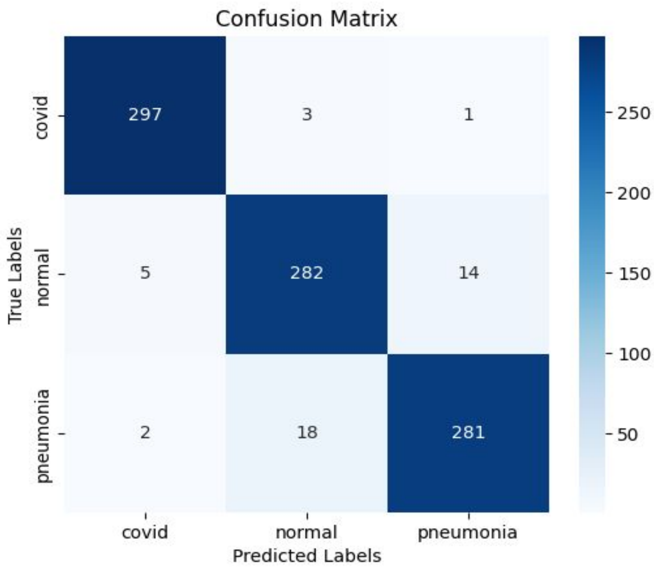
SHAP Explanation:



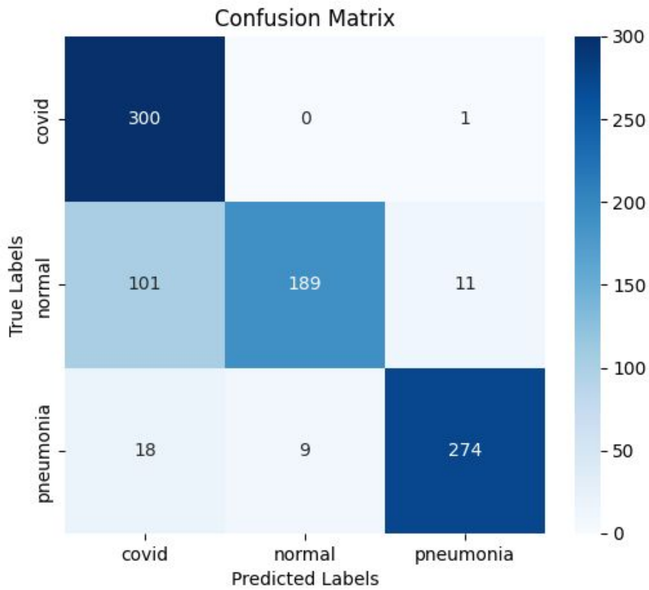
Model Evaluation Metrics

Model	Accuracy	Precision	Recall	F1 Score
ResNET-50	90.92%	0.9150	0.9092	0.9089
InceptionV3	96.57%	0.9678	0.9657	0.9656
Proposed Model	87.26%	0.8854	0.8726	0.8735

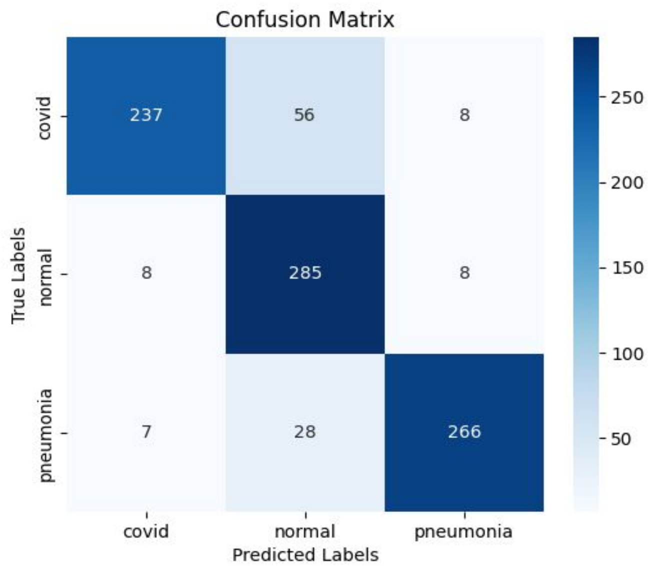
ResNET-50



InceptionV3



Proposed CNN Model



Future Work

- **Ensemble Learning**: Combine predictions from multiple models (e.g., ResNet + Inception + custom CNN) to improve robustness and accuracy.
- **Layer-wise Relevance Propagation (LRP)**: Add LRP for deeper interpretability, especially in identifying how pixel-level decisions contribute.
- **Web/Cloud Deployment**: Host the model with explanations as a web-based app for broader accessibility.
- **Lightweight Models**: Introduce efficient models like MobileNet or EfficientNet for faster inference in real-time or embedded environments.

Conclusion

- Utilizes CNN for accurate classification of pulmonary diseases from chest radiographs.
- Incorporates Explainable AI techniques such as:
  - SHAP to quantify feature contributions.
  - Grad-CAM to highlight relevant regions in the radiographs.
- Provides transparent and interpretable insights into the model's decisions.
- Aids healthcare professionals in making informed clinical decisions with greater confidence.

References

- Naz Z, Khan MUG, Saba T, Rehman A, Nobanee H, Bahaj SA. An Explainable AI-Enabled Framework for Interpreting Pulmonary Diseases from Chest Radiographs. Cancers (Basel). 2023 Jan 3;15(1):314. doi: 10.3390/cancers15010314. PMID: 36612309; PMCID: PMC9818469.
- Zafar, M.R.; Khan, N. Deterministic Local Interpretable Model-Agnostic Explanations for Stable Explainability. Mach. Learn. Knowl. Extr. 2021, 3, 525-541. https://doi.org/10.3390/make3030027
- Singh, M., Bansal, S., Ahuja, S. et al. Transfer learning–based ensemble support vector machine model for automated COVID-19 detection using lung computerized tomography scan data. Med Biol Eng Comput 59, 825–839 (2021). https://doi.org/10.1007/s11517-020-02299-2

References

- Teixeira, L.O.; Pereira, R.M.; Bertolini, D.; Oliveira, L.S.; Nanni, L.; Cavalcanti, G.D.C.; Costa, Y.M.G. Impact of Lung Segmentation on the Diagnosis and Explanation of COVID-19 in Chest X-ray Images. Sensors 2021, 21, 7116. https://doi.org/10.3390/s21217116
- Szepesi, Patrik & Szilagyi, Laszlo. (2022). Detection of pneumonia using convolutional neural networks and deep learning. Biocybernetics and Biomedical Engineering. 42. 10.1016/j.bbe.2022.08.001.
- Sahin, M.E., Ulutas, H., Yuce, E. et al. Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images. Neural Comput & Applic 35, 13597–13611 (2023). https://doi.org/10.1007/s00521-023-08450-y

Thank You

Appendix B: Vision, Mission, Programme Outcomes and Course Outcomes

Vision, Mission, Programme Outcomes and Course Outcomes

Institute Vision

To evolve into a premier technological institution, moulding eminent professionals with creative minds, innovative ideas and sound practical skill, and to shape a future where technology works for the enrichment of mankind.

Institute Mission

To impart state-of-the-art knowledge to individuals in various technological disciplines and to inculcate in them a high degree of social consciousness and human values, thereby enabling them to face the challenges of life with courage and conviction.

Department Vision

To become a centre of excellence in Computer Science and Engineering, moulding professionals catering to the research and professional needs of national and international organizations.

Department Mission

To inspire and nurture students, with up-to-date knowledge in Computer Science and Engineering, ethics, team spirit, leadership abilities, innovation and creativity to come out with solutions meeting societal needs.

Programme Outcomes (PO)

Engineering Graduates will be able to:

1. Engineering Knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. Conduct investigations of complex problems: Use research-based knowledge including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. Modern Tool Usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal, and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

7. Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8. Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. Individual and Team work: Function effectively as an individual, and as a member or leader in teams, and in multidisciplinary settings.

10. Communication: Communicate effectively with the engineering community and with society at large. Be able to comprehend and write effective reports documentation. Make effective presentations, and give and receive clear instructions.

11. Project management and finance: Demonstrate knowledge and understanding of engineering and management principles and apply these to one's own work, as a member and leader in a team. Manage projects in multidisciplinary environments.

12. Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and lifelong learning in the broadest context of technological change.

Programme Specific Outcomes (PSO)

A graduate of the Computer Science and Engineering Program will demonstrate:

PSO1: Computer Science Specific Skills
The ability to identify, analyze and design solutions for complex engineering problems in multidisciplinary areas by understanding the core principles and concepts of computer science and thereby engage in national grand challenges.

PSO2: Programming and Software Development Skills
The ability to acquire programming efficiency by designing algorithms and applying standard practices in software project development to deliver quality software products meeting the demands of the industry.

PSO3: Professional Skills
The ability to apply the fundamentals of computer science in competitive research and to develop innovative products to meet the societal needs thereby evolving as an eminent researcher and entrepreneur.

Course Outcomes (CO)

CO	Description
CO1	Identify academic documents from the literature which are related to her/his areas of interest (Cognitive knowledge level: **Apply**).
CO2	Read and apprehend an academic document from the literature which is related to her/his areas of interest (Cognitive knowledge level: **Analyze**).
CO3	Prepare a presentation about an academic document (Cognitive knowledge level: **Create**).
CO4	Give a presentation about an academic document (Cognitive knowledge level: **Apply**).
CO5	Prepare a technical report (Cognitive knowledge level: **Create**).

Appendix C: CO-PO-PSO Mapping

COURSE OUTCOMES:

After completion of the course, the student will be able to:

SL.NO	DESCRIPTION	Bloom's Taxonomy Level
CO1	Model and solve real-world problems by applying knowledge across domains (Cognitive knowledge level:Apply).	Level3: Apply
CO2	Develop products, processes, or technologies for sustainable and socially relevant applications. (Cognitive knowledge level:Apply).	Level 3: Apply
CO3	Function effectively as an individual and as a leader in diverse teams and comprehend and execute designated tasks. (Cognitive knowledge level:Apply).	Level 3: Apply
CO4	Plan and execute tasks utilizing available resources within timelines, following ethical and professional norms (Cognitive knowledge level:Apply).	Level 3: Apply
CO5	Identify technology/research gaps and propose innovative/creative solutions (Cognitive knowledge level:Analyze).	Level 4: Analyze
CO6	Organize and communicate technical and scientific findings effectively in written and oral forms (Cognitive knowledge level:Apply).	Level 3: Apply

CO-PO AND CO-PSO MAPPING

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	2	3	3	2	2	1	1	2	2	1	2	3	2	2
CO2	3	3	3	2	3	2	2	2	2	2	2	2	3	3	3
CO3	2	1	2	1	2	1	1	2	3	3	2	2	2	2	3
CO4	2	2	2	2	2	3	2	3	3	3	2	3	2	2	2
CO5	3	3	3	3	3	2	2	1	2	2	2	2	3	3	3
CO6	2	2	2	2	2	2	1	1	3	3	2	3	2	2	3

3/2/1: High/Medium/Low

JUSTIFICATIONS FOR CO-PO MAPPING

Mapping	Level	Justification
101003/CS822U.1-PO1	M	Ability to apply fundamental knowledge of mathematics, science, and engineering to model and solve real-world problems.
101003/CS822U.1-PO2	M	Capability to analyze real-world problems, review research literature, and develop substantiated conclusions.
101003/CS822U.1-PO3	M	Skills to design and develop solutions for practical applications based on engineering principles.
101003/CS822U.1-PO4	M	Competence in conducting investigations and interpreting data to solve engineering challenges.
101003/CS822U.1-PO5	H	Proficiency in utilizing modern engineering tools and techniques to analyze and address real-world problems.
101003/CS822U.1-PO6	M	Awareness of the societal impact of engineering solutions and the ethical responsibilities of professionals.
101003/CS822U.1-PO7	M	Understanding of environmental and sustainability considerations in engineering applications.
101003/CS822U.1-PO8	L	Adherence to ethical and professional norms in engineering practices.
101003/CS822U.1-PO9	L	Capability to work independently and collaborate effectively within multidisciplinary teams.
101003/CS822U.1-PO10	M	Ability to communicate technical concepts and solutions effectively in oral and written formats.
101003/CS822U.1-PO11	H	Application of engineering and management principles in project development and implementation.
101003/CS822U.1-PO12	H	Recognition of the need for continuous learning to stay updated with evolving technologies.
101003/CS822U.2-PO1	H	Systematic approach to planning, developing, testing, and implementing solutions in computing domains.

101003/CS822U.2-PO2	H	Mathematical and engineering fundamentals applied to problem identification and solution design.
101003/CS822U.2-PO3	H	Formulation and systematic analysis of project requirements to ensure effective solutions.
101003/CS822U.2-PO5	H	Use of a structured approach in solving complex computational and engineering problems.
101003/CS822U.2-PO6	H	Consideration of technical and societal aspects while developing solutions.
101003/CS822U.2-PO7	H	Application of sustainable engineering principles in project execution.
101003/CS822U.2-PO8	M	Emphasis on ethical considerations and responsible engineering practices.
101003/CS822U.2-PO9	H	Professional conduct in project execution while adhering to ethical norms.
101003/CS822U.2-PO11	H	Effective communication through reports, presentations, and clear instructions.
101003/CS822U.2-PO12	M	Team-based learning approach enhances problem-solving and collaboration skills.
101003/CS822U.3-PO9	H	Team projects encourage independent thinking and lifelong learning.
101003/CS822U.3-PO10	H	Application of algorithm design and development skills to project execution.
101003/CS822U.3-PO11	H	Effective problem-solving strategies improve the quality of solutions in various domains.
101003/CS822U.3-PO12	H	Use of fundamental engineering concepts for problem-solving and decision-making.
101003/CS822U.4-PO5	H	Problem identification and solution formulation using technical knowledge.
101003/CS822U.4-PO8	H	Consideration of safety, health, and ethical factors in project execution.
101003/CS822U.4-PO9	H	Use of research-based knowledge to analyze and interpret experimental results.

101003/CS822U.4-PO10	H	Selection and application of modern engineering tools for problem-solving.
101003/CS822U.4-PO11	M	Engineering solutions addressing societal and environmental concerns.
101003/CS822U.4-PO12	H	Understanding the need for sustainable development in engineering solutions.
101003/CS822U.5-PO1	H	Adherence to ethical principles and professional responsibilities.
101003/CS822U.5-PO2	M	Effective communication of engineering concepts and documentation.
101003/CS822U.5-PO3	H	Integration of engineering and management principles in project execution.
101003/CS822U.5-PO4	H	Emphasis on continuous learning for technological advancements.
101003/CS822U.5-PO5	M	Skill enhancement in programming, analysis, and algorithm development.
101003/CS822U.5-PO12	M	Application of computing and IT skills in solving industry-relevant problems.
101003/CS722U.6-PO5	M	Development of systematic approaches for designing and testing solutions.
101003/CS822U.6-PO8	H	Collaboration within teams to solve complex engineering problems.
101003/CS822U.6-PO9	H	Effective teamwork in research, analysis, and solution development.
101003/CS822U.6-PO10	M	Designing engineering components and systems to meet specific requirements.
101003/CS822U.6-PO11	M	Application of research methodologies in data analysis and system evaluation.
101003/CS822U.6-PO12	H	Ethical responsibility and professionalism in engineering practices.
101003/CS822U.1-PSO1	H	Application of computer science principles to solve industry-relevant problems.

101003/CS822U.2-PSO2	M	Development of sustainable and socially relevant applications.
101003/CS822U.3-PSO3	H	Collaboration and teamwork skills improve professional competencies.
101003/CS822U.4-PSO3	H	Effective planning and scheduling lead to better project management.
101003/CS822U.5-PSO1	H	Application of computational knowledge to create innovative solutions.
101003/CS822U.6-PSO3	H	Communication and documentation of technical findings enhance professional growth.