**PROJECT REPORT**

**ON**

**XAI-ASSISTED DEEP LEARNING MODELS FOR TRUSTWORTHY PREDICTION OF PNEUMONIA USING CHEST X-RAY IMAGES**

***Submitted by***

**ADITI MISHRA**

**2101110022**

**APARNA CHOUDHURY**

**2101110033**

**(7th Sem)**

Under The Esteemed Guidance Of

**Dr. ASHOK KUMAR BHOI**

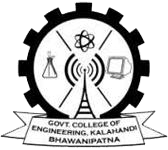
**(Asst. Professor)**

**In partial fulfilment for the award of the degree of**

BACHELOR OF TECHNOLGY

**In**

COMPUTER SCIENCE AND ENGINEERING



**GOVERNMENT COLLEGE OF ENGINEERING, KALAHANDI,**

**BHAWANIPATNA**

**NOVEMBER-2024**

**CERTIFICATE**

It is Certified that this project report **“XAI-ASSISTED DEEP LEARNING MODELS FOR TRUSTWORTHY PREDICTION OF PNEUMONIA USING CHEST X-RAY IMAGES”** is the bonafide work of **“ADITI MISHRA and APARNA CHOUDHURY ”** who carried out the project work under my supervision. It is Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

**SIGNATURE SIGNATURE**

### Dr. ASHOK BHOI Dr. GOPAL BEHERA

### SUPERVISER(Dept. Of CSE) (HEAD OF DEPARTMENT,CSE)

# DECLARATION

We declare that this written submission represents our ideas in our own words, also others ideas have been included. We have adequately cited and referenced the original sources that were included. We also declare that we have adhered to principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/act source in my submission.

We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal from the sources which have thus not been properly cited or from whom proper permission has been taken when needed.

**ADITI MISHRA**

**(Regd. No. - 2101110022)**

**APARNA CHOUDHURY**

**(Regd. No. - 2101110033)**

**ACKNOWLEDGEMENT**

We would like to express my heartfelt gratitude to Dr. Ashok Bhoi, assistant professor of the Department of Computer Science and Engineering, for his invaluable guidance and support throughout our exploration of the topic " xai-assisted deep learning models for trustworthy prediction of pneumonia using chest x-ray images."

His dedication to teaching and passion for the subject have been instrumental in shaping our understanding of x-ai. His willingness to answer my queries and engage in fruitful discussions has been truly inspiring. We are grateful for his encouragement and belief in my abilities, which motivated me to delve deeper into the intricacies of this subject.

Once again, I extend my heartfelt thanks to our principal sir Dr. Subranshu Shekhar Dash for his guidance and mentorship, which has played a vital role in shaping my academic journey in the field of data compression.

Sincerely,

Adtit Mishra - 2101110022

Aparna Choudhary - 2101110033

# ABSTRACT

This research delves into the realm of pneumonia detection in chest X-ray images, leveraging the power of Convolutional Neural Networks (CNNs). The utilization of diverse CNN architectures, namely Vgg16, AlexNet, LeNet, Inception, and Spinal, forms the crux of our investigation. Each architecture is carefully chosen for its unique characteristics, ranging from depth (Vgg16) to hierarchical feature extraction (AlexNet), simplicity (LeNet), and innovative parallel filtering (Inception). The unexpected star of the study, however, is the LeNet architecture, outshining others with a remarkable training accuracy of 95.42% and a validation accuracy of 93.71%.

A pivotal aspect of our research lies in the profound impact of superpixel methods on image segmentation, influencing the overall performance of our models. Quick Shift, with its mode- seeking segmentation scheme, and SLIC, a cluster-based algorithm, play integral roles in enhancing the precision of our predictions. Moreover, the study incorporates the Jaccard Coefficient as a robust metric for quantifying segmentation accuracy, providing a nuanced understanding of model performance.

In summary, this research amalgamates cutting-edge deep learning architectures, meticulous evaluation metrics, and sophisticated image segmentation techniques to advance pneumonia detection. The findings not only highlight the efficacy of LeNet in medical image analysis but also underscore the crucial role of superpixel methods in optimizing model performance. The inclusion of the Jaccard Coefficient further enriches the research landscape, offering a comprehensive quantitative measure for segmentation accuracy. This study contributes significantly to the field of medical diagnostics, specifically in the context of chest X-ray interpretation for pneumonia detection.

Student Signature Guide Signature

# TABLE OF CONTENTS

**CERTIFICATE…………………………………………..2**

# DECLARATION…………………………………………3

**ACKNOWLEDGEMENT……………………………….4**

# ABSTRACT………………………………………………5

# Introduction………………………………………………1

1. Existing Approaches 3
2. Problem Statement 7
3. System Framework 8
4. Methodology 11
   1. [Using only CNN Models 11](#_TOC_250002)
   2. [Using CNN with XAI Model 14](#_TOC_250001)
      1. [LIME Algorithm 15](#_TOC_250000)
      2. Super pixel Method-1 16
      3. Super pixel Method-2 17
      4. Super pixel Method-3 18
5. Simulation results 21
6. Conclusion and Future works 24
7. References 25

**LIST OF FIGURES**

1. FIGURE 1: Traditional vs. XAI-Based Pneumonia Detection 6
2. FIGURE 2: Schematic Diagram of X-AI 8
3. FIGURE 3: XAI Output Using Quick Shift Superpixel Method 21
4. FIGURE 4: XAI Output Using SLIC Superpixel Method 22
5. FIGURE 5: XAI Output Using Quick Shift with noise Superpixel Method 23

# INTRODUCTION

In the ever-evolving landscape of healthcare, the fusion of artificial intelligence (AI) and medical diagnostics emerges as a transformative force. At the forefront of this revolution is the pursuit of accurate and swift detection of pneumonia in chest X-ray images—a pursuit with profound implications for patient care, diagnostic efficiency, and the broader field of medical imaging.

## The Global Impact of Pneumonia

Pneumonia, a prevalent respiratory infection, continues to exact a substantial toll on global health. Timely diagnosis is critical, as delays can lead to complications and compromise patient outcomes. The integration of AI in pneumonia detection promises not only to expedite the diagnostic process but also to enhance accuracy, enabling healthcare professionals to initiate prompt and targeted interventions.

## Challenges in Traditional Diagnostic Approaches

Traditional methods of pneumonia detection, reliant on radiologist interpretations, face challenges in scalability and consistency. The subjectivity inherent in human diagnosis and the increasing volume of medical imaging data necessitate a paradigm shift towards automated, efficient, and reliable solutions. This study seeks to address these challenges through the application of advanced AI techniques.

## The Promise of AI in Medical Imaging

Within the realm of medical imaging, AI, and particularly deep learning, stands out as a beacon of hope. The capacity of deep learning models for intricate pattern recognition and image analysis has positioned them as invaluable tools in augmenting diagnostic capabilities. Harnessing the power of AI, our study focuses on developing a custom Convolutional Neural Network (CNN) tailored to dissect chest X-ray images with a precision attuned to the nuances of pneumonia detection.

## Motivation for Advanced AI-Based Approaches

Motivated by the urgent need for accuracy in pneumonia diagnosis, our study aims to contribute to the burgeoning field of AI in healthcare. The motivation stems not only from the desire to improve diagnostic efficacy but also to foster a deeper understanding of the interpretability of AI-driven recommendations. Transparency and interpretability become paramount, especially in the context of life-critical decisions.

## Embracing Explainable AI (XAI)

In navigating the complexities of AI adoption in medical diagnostics, our approach incorporates Explainable AI (XAI). Beyond achieving accurate predictions, the emphasis on XAI is driven by the need to demystify the decision-making process of complex models. The marriage of AI's computational prowess with interpretability ensures that healthcare professionals can trust and comprehend the diagnostic insights provided.

## The Significance of Pneumonia Detection:

Pneumonia, a prevalent respiratory infection, can affect individuals of all ages and, if not promptly identified and treated, can lead to severe health complications. The use of chest X- ray imaging to diagnose pneumonia has become a cornerstone of medical practice. It allows healthcare professionals to visualize pulmonary abnormalities, such as infiltrates or consolidations, which are indicative of the disease.

Automating the process of pneumonia detection through ML and DL models offers several significant advantages. These models are capable of rapid image analysis and can assist healthcare providers in cases of high patient volume. However, with great power comes a responsibility for interpretability and accountability.

# Existing Approaches (ML/DL based)

In the realm of pneumonia detection from chest X-ray images, the application of Machine Learning (ML) and Deep Learning (DL) models has transformed the diagnostic landscape. These approaches have shown great promise in automating the interpretation of radiographic images and significantly expediting the diagnosis process. This section provides an in-depth exploration of the existing ML/DL-based approaches in pneumonia detection.

## Machine Learning Approaches:

Machine Learning methods have long been applied to medical image analysis, including chest X-ray images. These approaches primarily rely on feature engineering and traditional algorithms to differentiate between normal and pneumonia-affected lung images. Key ML- based techniques include:

* Support Vector Machines (SVM): SVMs are used for binary classification tasks, and they work by finding a hyperplane that best separates data points into two classes. In the context of pneumonia detection, SVMs have been used with handcrafted image features to classify X-ray images.
* Random Forests: Random Forest is an ensemble learning method that combines the outputs of multiple decision trees to make a prediction. This approach is applied by extracting relevant image features and employing ensemble models for classification.
* Principal Component Analysis (PCA): PCA is used for dimensionality reduction by projecting image features onto a lower-dimensional space. It aids in compressing image information while preserving discriminative features for classification.

While these traditional ML approaches have demonstrated reasonable accuracy, they rely heavily on feature engineering, which requires domain expertise and may not capture the full complexity of radiographic images. Furthermore, they often lack the flexibility and adaptability seen in DL models.

## Deep Learning Approaches

Deep Learning, particularly Convolutional Neural Networks (CNNs), has ushered in a new era in chest X-ray image analysis. DL models have the capacity to automatically learn intricate patterns and features from raw images, eliminating the need for manual feature engineering. Key DL-based techniques for pneumonia detection include:

* Convolutional Neural Networks (CNNs): CNNs are specifically designed for image analysis and are characterized by convolutional layers that learn feature hierarchies from input data. In the context of pneumonia detection, CNNs have been shown to outperform traditional ML methods by a significant margin.
* Transfer Learning: Transfer learning leverages pre-trained CNN models (e.g., VGG, ResNet, Inception) trained on large image datasets and fine-tunes them on chest X-ray images. This approach is particularly valuable when limited labeled medical image data is available.

Deep Learning models have brought forth breakthroughs in pneumonia detection. They are capable of learning intricate features, identifying subtle patterns, and achieving high accuracy levels. However, one of the primary challenges associated with DL models is their inherent complexity, which often results in a lack of interpretability. The question of "why" a certain decision is made remains a concern.

## Survey on existing approaches:

**Sl No.**

## Author - Year Methodology Limitations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | Sazzad Yousuf Sourab, Md Ahasan Kabir  (2022) [5] | Machine Learning(SVM, Random Forest)  Deep learning(CNN , KNN) | Need of interpretable A.I | |
| 2 | Rohit Kundu, Ritacheta Das, Zong Woo Geem, Gi-Tae Han, Ram Sarkar  (2021) [4] | Deep learning(CNN , KNN, ANN) | Dependency Bounding Annotations | on Box |
| 3 | Amer Kareem, Haiming Liu & Paul Sant  (2022) [3] | Deep learning(CNN) | lower interpretability,high computational complexity | |
| 4 | Nahida Habib, Md. Mahmodul Hasan, Mohammad Motiur Rahman (2020) [2] | Machine Learning(SVM, PCA) Deep learning(Logistic Regression) | Need of visualization for better  understanding | |
| 5 | Yusuf Brima, Marcellin Atemkeng, Stive Tankio Djiokap, Jaures Ebiele and Franklin Tchakounté (2021) [1] | Deep learning(Transfer Learning) | Opaqueness and Lack of Interpretability | |

**Challenges in ML/DL Approaches:**

Despite their considerable success, both traditional ML and DL-based approaches face common challenges in pneumonia detection:

* Interpretability: The black-box nature of DL models poses difficulties in understanding their decision-making processes. In a medical context, knowing why a particular diagnosis was made is essential.
* Data Imbalance: Datasets for pneumonia detection are often imbalanced, with a more significant number of normal images than pneumonia-affected images. This imbalance can impact model performance.
* Robustness and Generalization: Ensuring that models generalize well to diverse patient populations and conditions is a considerable challenge. Overfitting and underfitting are common concerns.
* Data Augmentation: Effective data augmentation techniques are needed to mitigate the shortage of labelled medical images for model training.

In response to these challenges, the integration of explainable Artificial Intelligence (XAI) techniques, such as LIME, becomes increasingly relevant to enhance the interpretability of ML/DL models for pneumonia detection.

## Motivation for AI-Based Pneumonia Detection with XAI:

### Addressing Interpretability in Life-Critical Diagnoses:

In the realm of medical diagnoses, especially in critical conditions like pneumonia, the need for transparent and interpretable diagnostic outcomes cannot be overstated. The decisions made in healthcare are often life-critical, and the ability to understand and trust the diagnostic process is paramount. It is in this context that the incorporation of Explainable AI (XAI) takes center stage.

* **Ensuring Trustworthiness:** In life-critical diagnoses, trust in the diagnostic process is not just a desirable attribute but an absolute necessity. The opacity of AI models, especially deep learning models, can be a significant hurdle. XAI offers the means to make AI-assisted diagnostic decisions understandable and trustworthy. It bridges the gap between complex machine learning algorithms and clinical interpretability, thereby instilling confidence in the decision-making process.
* **Enhancing Clinician Confidence:** One of the primary motivations for integrating visualization and XAI into pneumonia detection is to bolster clinicians' confidence in the recommendations provided by AI systems.
* **Efficiency and Accuracy:** An AI-driven diagnostic system that provides transparent and interpretable results can have a profound impact on clinical decision-making. Clinicians who can confidently comprehend the basis for a particular diagnosis are better equipped to make efficient and accurate decisions. This enhanced confidence in AI-driven recommendations is a catalyst for more effective diagnoses, potentially saving lives through swift and precise medical interventions.
* **Mitigating the Limitations of Deep Learning:** While deep learning models are incredibly powerful, they are not without their limitations. In medical diagnosis, deep learning can sometimes fall short in providing the visual insights that are crucial for accurate decision-making. The integration of XAI offers a solution to this challenge.
* **Enhancing Visual Information:** Deep learning models excel at pattern recognition but may struggle to provide the visual information required for nuanced medical diagnoses. Embracing XAI allows us to bridge this gap by offering medical professionals the necessary visual insights to make informed decisions. This amalgamation of deep learning's computational prowess with the interpretability offered by XAI contributes to improved patient outcomes by ensuring that no critical detail goes unnoticed.

This motivation section highlights the critical need for XAI in pneumonia detection, emphasizing the life-critical nature of medical diagnoses and the potential to enhance clinician confidence while mitigating the limitations of deep learning in medical imaging.

## Visualizing the Contrast: Traditional vs. XAI-Based Pneumonia Detection:

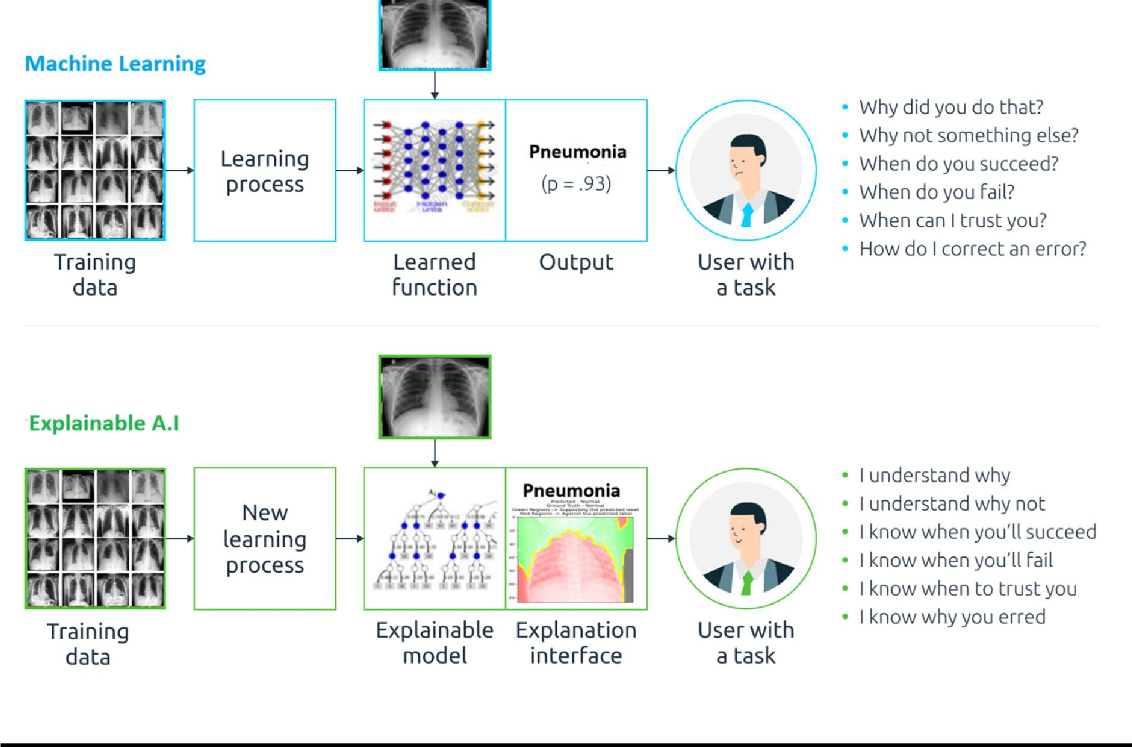


FIGURE 1: Traditional vs. XAI-Based Pneumonia Detection

# Problem Statement

The current landscape of pneumonia detection relies heavily on Convolutional Neural Networks (CNNs). While these models demonstrate promising accuracy, their inherent complexity raises concerns about interpretability and transparency. The "black-box" nature of CNNs poses challenges in understanding the decision-making processes crucial for critical applications in medical diagnosis.

Despite the advancements in CNN architectures, there remains a need to address the following key issues:

**Interpretability Challenges:** Existing CNN models lack clear interpretability, making it challenging for medical professionals to trust and understand the rationale behind their predictions.

**Enhancing Model Transparency:** The intricate layers of CNN architectures often hinder the extraction of meaningful insights, particularly in the context of life-critical diagnoses like pneumonia.

**Comparison of Superpixel Methods:** The choice of super pixel algorithms within the eXplainable Artificial Intelligence (XAI) framework, specifically LIME, introduces a need for a comprehensive comparison. This comparison aims to identify the superior super pixel method, either Quick Shift or SLIC, concerning interpretability and performance in pneumonia detection.

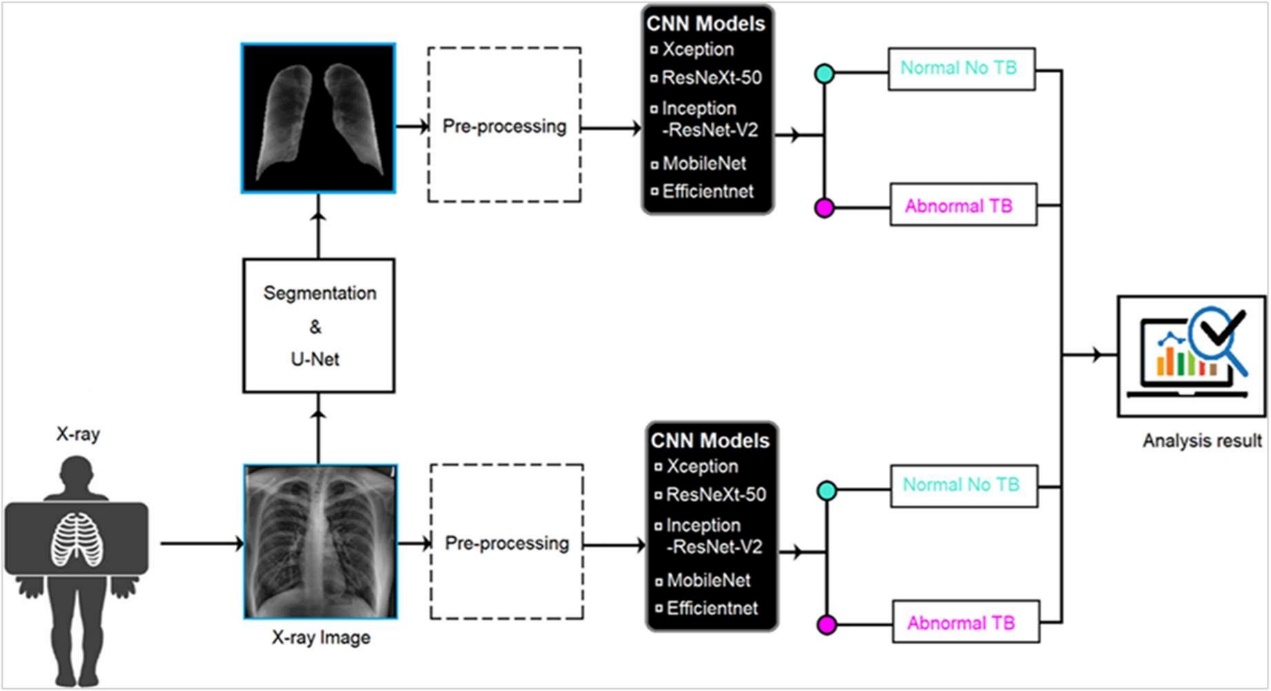
In light of these challenges, our project seeks to explore the integration of CNN models, including Vgg16, AlexNet, LeNet, Inception, and Spinal for pneumonia detection. Additionally, we aim to enhance model interpretability through the incorporation of eXplainable AI (XAI), leveraging the LIME algorithm with two distinct super pixel methods.

# System Framework

The system framework intricately integrates Convolutional Neural Networks (CNNs) and eXplainable Artificial Intelligence (XAI) techniques, presenting a holistic approach to pneumonia detection with a focus on interpretability.

## Schematic Diagram

The schematic diagram provides a holistic view of the entire pneumonia detection system, encompassing the integration of CNN architectures, XAI techniques, and the data processing pipeline.



**X-AI**

CNN MODEL

Segmenta tion

CNN MODEL

FIGURE 2: Schematic Diagram of X-AI

In this schematic diagram:

### Convolutional Neural Networks (CNNs):

* Architectures:
  + Vgg16: Employing the Vgg16 architecture known for its depth and performance in image classification tasks.
  + AlexNet: Utilizing AlexNet, a pioneering deep learning model recognized for its success in the ImageNet Large Scale Visual Recognition Challenge.
  + LeNet: Incorporating LeNet, an early yet effective CNN architecture suitable for image recognition.
  + Inception: Leveraging the Inception architecture designed for improved efficiency in capturing complex features.
  + Spinal: Introduced for diversified feature extraction, specializing in intricate patterns to enhance ensemble performance.

### eXplainable Artificial Intelligence (XAI):

* LIME Integration:
  + Seamless integration of Local Interpretable Model-agnostic Explanations (LIME) for enhanced model interpretability.
  + Detailed analysis of LIME's working mechanism, emphasizing its role in providing local explanations for individual predictions.
* Superpixel Methods:
  + Incorporating Quick Shift and SLIC (Simple Linear Iterative Clustering) within LIME.
  + Comparative assessment of the two superpixel algorithms to determine their impact on interpretability.

### Data Processing:

* Data Collection:
  + Compiling a diverse and extensive dataset of chest X-ray images sourced from reputable medical repositories.
  + Rigorous quality checks to ensure the inclusion of representative samples.
* Preprocessing:
  + Implementing standardized preprocessing techniques to optimize images for CNN training.
  + Addressing challenges such as class imbalance through strategic preprocessing strategies.

### Model Integration:

* Pipeline Overview:
  + Commencing with the input of chest X-ray images, the pipeline undergoes preprocessing and segmentation to enhance region-specific information.
  + Navigating through the varied CNN architectures, each contributing unique perspectives in pneumonia detection.
  + Culminating with eXplainable insights facilitated by LIME, offering transparency into the decision-making process.

The schematic diagram provides a visual representation of the overall workflow, emphasizing the incorporation of the XAI component and LIME for interpretability.

## Description of Steps:

The implementation of the XAI-based approach using LIME for pneumonia detection in chest X-ray images involves a sequence of well-defined steps:

### Data Preparation:

Begin by organizing and preparing the data for pneumonia detection. This involves specifying the paths to the training, testing, and validation datasets. The data should be stored in a structured manner to facilitate loading.

### Model Selection:

Choose a custom Convolutional Neural Network (CNN) model for pneumonia detection. Create the model using TensorFlow and Keras.

### Data Split:

Split the dataset into training and testing sets for model training and evaluation.

### Model Training:

Train the custom CNN model using the training data and validate it with the testing data.

### LIME Integration:

Integrate LIME for model interpretability. Additional code specific to LIME is required and may include importing LIME and configuring it.

### Image Selection:

Choose a representative chest X-ray image from the testing dataset for explanation. No code is needed; this is a manual selection process.

### Image Perturbation:

Apply perturbations to the selected image, creating variations of the original image. This is typically handled by LIME.

### Model Prediction:

Make predictions for each perturbed image using the trained CNN model.

### Explanation Generation:

Use LIME to generate explanations for the selected image. LIME code is typically used for this step.

### Visualization:

Create visualizations that overlay LIME-generated explanations on the original chest X-ray image. Visualization code depends on the library or method used

### Interpretation:

Interpret LIME-generated explanations to understand which regions in the image influenced the model's decision. This is a manual interpretation process.

# Methodology

Our methodology is meticulously designed to delve into the realms of pneumonia detection, employing a strategic combination of Convolutional Neural Networks (CNNs) and eXplainable Artificial Intelligence (XAI) techniques. This section delineates the experimental procedures undertaken to ensure a comprehensive evaluation.

### Experimental Setup

Before delving into the specifics, it's imperative to outline the experimental setup. We carefully selected five distinct CNN architectures—Vgg16, AlexNet, LeNet, Inception and Spinal—to gauge their efficacy in pneumonia prediction. Additionally, the integration of eXplainable AI adds a layer of transparency to our models, fostering interpretability.

### Model Configurations

The methodology unfolds with a focused examination of two primary configurations: utilizing only CNN models for prediction and incorporating CNNs with XAI models. Each configuration is meticulously examined to draw nuanced comparisons and insights.

## Using only CNN models:

### Introduction to CNN models:

In our pursuit of enhancing pneumonia prediction accuracy in chest X-ray images, we leverage the capabilities of several state-of-the-art Convolutional Neural Network (CNN) architectures. Each of these models brings a unique set of features and complexities, contributing to the overall effectiveness of our predictive system.

### Vgg16 (Visual Geometry Group 16):

Vgg16, developed by the Visual Geometry Group, is renowned for its depth and performance in image classification tasks. With 16 weight layers, this architecture excels at capturing intricate patterns and features within images.

### AlexNet:

A pioneering model in the realm of deep learning, AlexNet gained prominence for its remarkable success in the ImageNet Large Scale Visual Recognition Challenge. Its design focuses on extracting hierarchical features, making it adept at image classification.

### LeNet:

Although an earlier CNN architecture, LeNet remains effective in image recognition tasks. Its simplicity and efficiency make it suitable for our purpose, contributing to the diverse set of models employed for pneumonia prediction.

### Inception:

Leveraging the Inception architecture, we prioritize improved efficiency in capturing complex features within chest X-ray images. Inception's innovative design incorporates multiple filter sizes in parallel, enhancing its ability to discern various patterns.

* + - **Spinal:** Uniquely designed for diverse feature extraction, Spinal plays a key role in enhancing ensemble performance by focusing on intricate patterns within medical images.

### Significance of Model Selection:

These chosen CNN architectures are not arbitrary; their selection is a strategic choice based on their proven track records in image classification and feature extraction. The depth, efficiency, and unique characteristics of Vgg16, AlexNet, LeNet, and Inception collectively contribute to a robust framework for pneumonia prediction. Through our systematic evaluation, we aim to discern how each model performs individually and their collective efficacy in enhancing diagnostic accuracy.

### Model Training:

**Overview:**

Our model training involves the implementation of five distinct Convolutional Neural Network (CNN) architectures—Vgg16, AlexNet, LeNet, Inception, and Spinal. Each architecture contributes unique features and complexities, catering to the varied aspects of chest X-ray image classification.

**Hyperparameters and Optimization:**

1. **Vgg16:**
   * + **Hyperparameters:** Standard configurations with a focus on depth.
     + **Optimization:** Utilizing the Adam optimizer with a learning rate suitable for convergence.

### AlexNet:

* + - **Hyperparameters:** Defined kernel sizes, strides, and filter depths.
    - **Optimization:** Leveraging the Adam optimizer with a carefully tuned learning rate.

### LeNet:

* + - **Hyperparameters:** Configured for simplicity with distinct convolutional and pooling layers.
    - **Optimization:** Using the Adam optimizer to facilitate training efficiency.

### Inception:

* + - **Hyperparameters:** Employing varying kernel sizes and filter depths within inception blocks.
    - **Optimization:** Adam optimizer with a learning rate conducive to effective training.

### Spinal:

* + - **Hyperparameters:** Tailored for ensemble learning, Spinal incorporates

diverse configurations to enhance feature extraction across multiple paths.

* + - **Optimization:** The Adam optimizer is employed, fine-tuned to accommodate the unique architecture of Spinal for optimal convergence.

### Unique Considerations:

* + - **Transfer Learning:** All models are initialized with pre-trained weights on ImageNet, harnessing knowledge from diverse datasets.
    - **Freezing Layers:** Certain layers are frozen during training to retain pre-learned features and optimize model performance.

### Evaluation Metrics:

The training process involves a rigorous evaluation of multiple performance metrics:

* + - **Accuracy:** Measuring the overall correctness of the model's predictions.
    - **Loss:** Assessing the divergence between predicted and actual values, guiding the optimization process.

### Model Robustness:

To ensure model robustness, early stopping criteria are applied to halt training if the validation loss fails to improve for a predefined number of epochs. This precaution prevents overfitting and encourages the generalization of the models to new data.

Through these training methodologies, we aim to develop highly effective models capable of accurately classifying chest X-ray images for pneumonia detection.

### Model table for pneumonia model:

The following table summarizes the performance metrics for each CNN architecture, showcasing accuracy and loss during both training and testing phases:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **CNN**  **Methods** | **Training Accuracy** | **Training Loss** | **Test Accuracy** | **Test Loss** | **Validation Accuracy** | **Validation Loss** |
| **VGG16** | 94.92% | 0.1367 | 93.85% | 0.1516 | 93.84% | 0.1601 |
| **LENET** | **97.01%** | **0.0856** | **95.42%** | **0.2090** | **93.71%** | **0.1656** |
| **INCEPTIO N** | 95.74% | 0.0977 | 93.51% | 0.2695 | 92.88% | 0.2220 |
| **SPINAL** | 94.15% | 0.1593 | 92.49% | 0.2205 | 92.20% | 0.2080 |
| **ALEXNET** | 94.71% | 0.0965 | 94.26% | 0.1600 | 93.65% | 0.1701 |

### Observations:

1. **Accuracy Trends:**
   * + LeNet demonstrates the highest training and testing accuracy, indicating robust learning.
     + Inception follows closely, showcasing consistent performance.
     + Vgg16, Alexnet and Spinal exhibit slightly lower accuracy, indicating potential room for improvement.

### Loss Analysis:

* + - LeNet and AlexNet achieves the lowest training and testing loss, indicating effective convergence.
    - Inception and Vgg16 performs well, with lower losses, demonstrating efficient learning.
    - Spinal and Inception show slightly higher losses, suggesting potential areas for optimization.

This table provides a snapshot of the models' performance, guiding further analysis and optimization efforts.

## Using CNN with XAI Model:

* + - In the quest for enhanced interpretability in medical image analysis, we integrate Explainable AI (XAI) techniques with our Convolutional Neural Network (CNN) models.
    - Model interpretability is crucial in medical contexts, allowing clinicians to trust and understand the decision-making process.

### LIME Algorithm Integration:

* + - Our approach incorporates the Local Interpretable Model-agnostic Explanations (LIME) algorithm to provide insights into individual predictions.
    - LIME functions by perturbing input images and observing the impact on model predictions, offering localized interpretability.

### Demonstration:

* + - To illustrate the effectiveness of XAI, we present visual demonstrations using LIME- generated images.
    - These images highlight specific regions within X-rays that significantly influence the CNN's predictions.

### Analysis and Insights:

* + - Through XAI, we gain valuable insights into the decision-making process of the CNN model.
    - Notable regions contributing to pneumonia predictions are identified, aiding in the understanding of model behavior.

### Super Pixels Overview:

* + - In addition to LIME, our XAI integration involves leveraging superpixel methods (e.g., Quick Shift and SLIC) within the interpretability framework.
    - Super pixels provide a higher-level representation of image regions, contributing to the overall interpretability of the CNN model.

### LIME-based Technique

Local Interpretable Model-agnostic Explanations (LIME) is a powerful and versatile XAI technique employed to enhance the interpretability of the pneumonia detection model. LIME addresses the critical challenge of understanding the decisions made by complex ML/DL models, making it a valuable tool in medical image analysis.

## LIME Algorithm:

The LIME algorithm is designed to generate interpretable and locally faithful explanations for machine learning classifiers. It operates on a specific input instance and aims to approximate the decision boundary of the underlying model in the vicinity of that instance.

## Pseudo-code Implementation

Below is a pseudo-code representation of the LIME algorithm used in our approach:

Input: classifier 'f', input sample 'x', number of superpixels 'n', number of features to pick 'm' Output: explainable coefficients from the linear model

1: y ← f.predict(x) 2: for i in n do

3: pi ← Permute(x) . Randomly pick superpixels 4: z ← f.predict(pi)

5: dist ← |y − z| 6: end for

7: simscore ← SimilarityScore(dist) 8: p ← Pick(p, simscore, m)

9: L ← LinearModel.fit(p, m, simscore) 10: return L.weight

return L.weights: Return the weights of the linear model. These weights provide an interpretable explanation of how the selected superpixels influence the classifier's prediction for the input sample.

## Considerations and Fine-tuning

It's important to note that the success of the LIME algorithm relies on thoughtful consideration and fine-tuning of parameters such as the number of superpixels and features. Additionally, the specifics of the permutation process and similarity score calculation should align with the guidelines outlined in the original LIME paper or relevant library documentation.

## Super pixel Method-1 (QuickShift)

The Quickshift method is a 2D image segmentation algorithm that is part of the family of local mode-seeking algorithms1. It operates in a 5D space consisting of color information and image location1.

Quickshift arranges all of the data points into a tree where parents in the tree are the nearest neighbors in the feature space which increase the estimate of the density. By imposing a limit on the distance between nearest neighbors (max\_dist), we decrease the amount of computation required to search for the nearest neighbors.

One of the benefits of Quickshift is that it actually computes a hierarchical segmentation on multiple scales simultaneously.

### Quick-Shift (QS) is an algorithm LIME uses by default.

Quickshift operates in a 5-dimensional space, considering the pixels as samples over this space which includes 3 color dimensions and 2 spatial dimensions.

## Algorithm

1. Input:
   * Image *I*
   * Bandwidth parameter *r* (controls the size of the neighborhood)
   * Optional parameters: kernel size, max distance, etc. 2.Initialize:
   * Choose initial cluster centers. This can be done, for example, by sampling pixels at a regular grid.
2. Iterative Update:
   * Repeat until convergence:
     + For each cluster center Ck :
     + Assign pixels from a square neighborhood around Ck based on a distance measure.
     + Update the cluster center based on the assigned pixels.
3. Convergence Check:
   * Check for convergence, typically by measuring the change in cluster centers or other convergence criteria.
4. Enforce Connectivity:
   * After convergence, enforce connectivity to ensure that segmented regions form coherent structures.

## Super pixel Method- 2 (SLIC)

SLIC, or Simple Linear Iterative Clustering, is a method for image segmentation. It is a popular superpixel algorithm and is used in a variety of image processing tasks.

SLIC works by converting the image into a 5D space composed of color and image location information. It then clusters pixels in this 5D space using a method similar to k-means. The result is a set of “superpixels” which are groups of pixels that share similar characteristics.

SLIC is often used in tasks where reducing the complexity of the image can be beneficial, such as object detection or semantic segmentation.

SLIC uses the well-known K-Means algorithm as a basis, but there are essential differences:

* The search space (2S×2S) is limited proportional to the size of the superpixel (S × S). This significantly reduces the number of distance calculations.
* In addition, the complexity is independent of the number of superpixels k, whereby SLIC has a complexity of O(N).
* Furthermore, a weighted distance measure (see equation 1) combines the spatial (ds) and color (dc) proximity.
* In addition, the control of compactness and size of the superpixels is ensured by a parameter (m).



## Algorithm

1: Initialize cluster centers Ck by sampling pixels at regular grid steps S.

2: Perturb cluster centers in an n, x n, neighborhood, to the lowest gradient position. 3: repeat

4: for each cluster center Ck do

5: Assign the best matching pixels from a 2S x 2S square neighborhood around

the cluster center according to the distance measure (Eq. 1). 6: end for

7: Compute new cluster centers and residual error E {LI distance between previous centers and recomputed centers}

8: until E < threshold

9: Enforce connectivity.

## Super pixel Method-3 (Quick Shift with Noise)

The QuickShift algorithm, known for its efficiency in superpixel segmentation, was adapted to accommodate the deliberate addition of noise to the image. Our implementation involves preprocessing steps to handle noise, followed by the application of the QuickShift algorithm to segment the image into superpixels. The integration of noise introduces perturbations to pixel values, enhancing the algorithm's robustness and providing a more realistic representation of real-world image data. Post-processing steps were applied to refine the segmentation results, and parameter tuning was performed to optimize the algorithm's performance in the presence of noise.

## Algorithm

1. Initialization with Noise:
   * Start by initializing cluster centers. Instead of placing them precisely on a regular grid, introduce noise by perturbing the initial positions randomly. This helps to inject variability into the clustering process.
2. Iterative Update:

* Repeat the following steps until convergence:
  + For each cluster center (Ck):
    - Assign pixels to the cluster based on a distance measure, considering both color and spatial information.
    - Update the cluster center by computing the mean position of the assigned pixels.

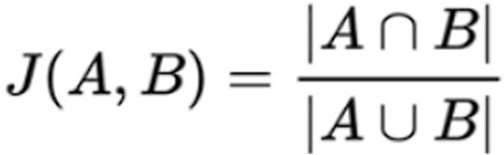
1. Convergence Check:
   * Check for convergence by monitoring changes in cluster centers or other convergence criteria.
2. Enforce Connectivity:

- After convergence, enforce connectivity to ensure that the segmented regions form coherent structures.

By incorporating noise during initialization, **the algorithm becomes more robust and less sensitive to minor variations in the data.** This enhanced QuickShift with noise approach aims to produce more flexible and adaptive segmentation results, particularly useful in scenarios where the data may contain irregularities or uncertainties.

## Jaccard Coefficient

The Jaccard coefficient, also known as the Jaccard index, is a measure of similarity between two sets. It is defined as the size of the intersection of the sets divided by the size of the union of the sets. The Jaccard coefficient is commonly used in various fields, including statistics, data analysis, and information retrieval. Mathematically, it can be expressed as:



where:

* Set A: The set of pixels belonging to a ground truth segmentation mask.
* Set B: The set of pixels obtained from the predicted segmentation.

-  is the number of correctly segmented pixels, indicating the overlap between the ground truth and the algorithm's prediction.

-  is the total number of pixels considered in the evaluation, encompassing both correctly and incorrectly segmented regions.

The Jaccard coefficient produces a value between 0 and 1. A value of 1 indicates that the sets are identical, while a value of 0 means that the sets have no elements in common.

In the context of image segmentation or clustering evaluation, the Jaccard coefficient is often used to measure the similarity between the ground truth set of pixels and the set of pixels obtained from an algorithmic segmentation. This can provide insights into the accuracy and effectiveness of the segmentation algorithm by quantifying the overlap between the predicted and true segmentations.

**Superpixel model table:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Superpixel**  **method** | **Mean**  **Value** | **Variance** | **Standard**  **deviation** | **Jaccard**  **Coefficient** |
| **Quick-Shift** | 0.686 | 0.02 | 0.020 | 0.8954 |
| **SLIC** | 0.5010 | 0.04 | 0.205 | 0.8954 |
| **QuickShift**  **with Noise** | **0.691** | **0.0015** | **0.0221** | **0.9679** |

# Simulation results

In this project, we undertook an extensive exploration of pneumonia detection leveraging state- of-the-art Convolutional Neural Network (CNN) architectures, including LeNet, AlexNet, VGG16, Spinal, Inception, and a custom-designed method.Following the CNN-based pneumonia detection, we applied Explainable AI (XAI) techniques, specifically LIME (Local Interpretable Model-agnostic Explanations), to elucidate the decision-making processes of the models.Notably, we employed three distinct superpixel methods, namely Quick Shift, SLIC and Quick Shift with noise, to enhance the interpretability of the models.The simulation of results focuses on presenting output images alongside their corresponding feature importance, highlighting the regions that significantly contribute to the pneumonia detection decisions. Additionally, we introduced a novel aspect to the superpixel methodology by incorporating noise during the Quick Shift process. This unique approach aims to assess the models' robustness and sensitivity to variations in image data. The amalgamation of diverse CNN architectures, multiple superpixel methods, and the integration of noise in the XAI process allows for a comprehensive evaluation of the pneumonia detection system's performance, shedding light on its interpretability and potential clinical applicability.

The forthcoming visualizations and analyses aim to elucidate the intricate details of the models' decision boundaries, providing insights crucial for both understanding and refining the pneumonia detection system:

**XAI Output Using Quick Shift Superpixel Method**

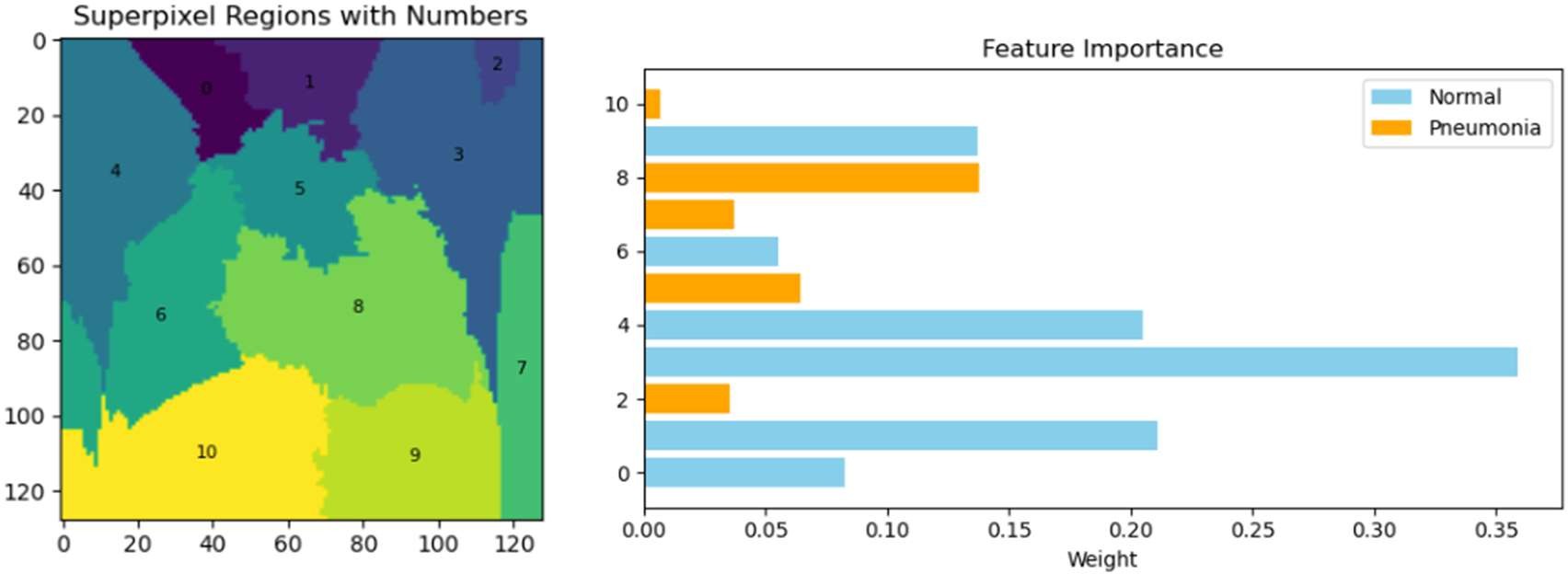
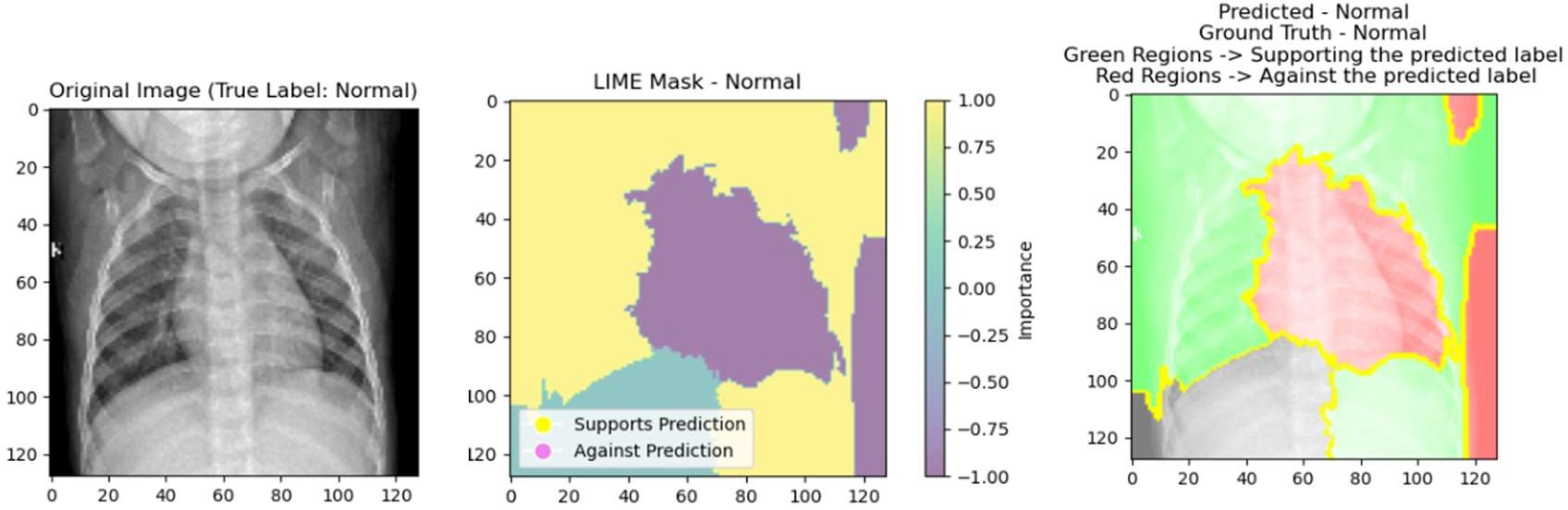
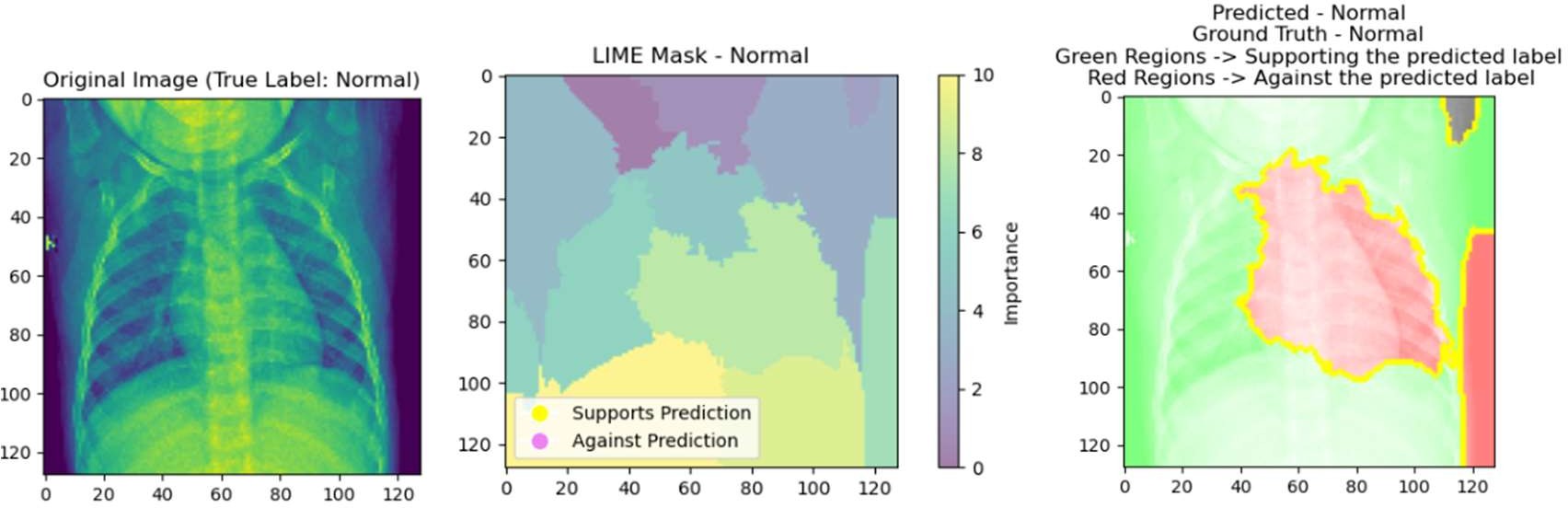


FIGURE 3: XAI Output Using Quick Shift Superpixel Method

In employing the Quick Shift superpixel method for our pneumonia detection project, the XAI output reveals a detailed understanding of the Convolutional Neural Network's (CNN)

decision-making process. The original chest X-ray serves as the foundation, while the LIME mask highlights influential areas guiding the model's prediction. The model's output is visually presented for validation. The accompanying superpixel image, annotated with numbers and feature weights, elucidates the specific regions crucial to pneumonia detection. Despite comparable performance with SLIC in Jaccard Coefficient, Quick Shift stands out, demonstrating its efficacy in providing interpretable and informative visualizations

**XAI Output Using SLIC Superpixel Method**



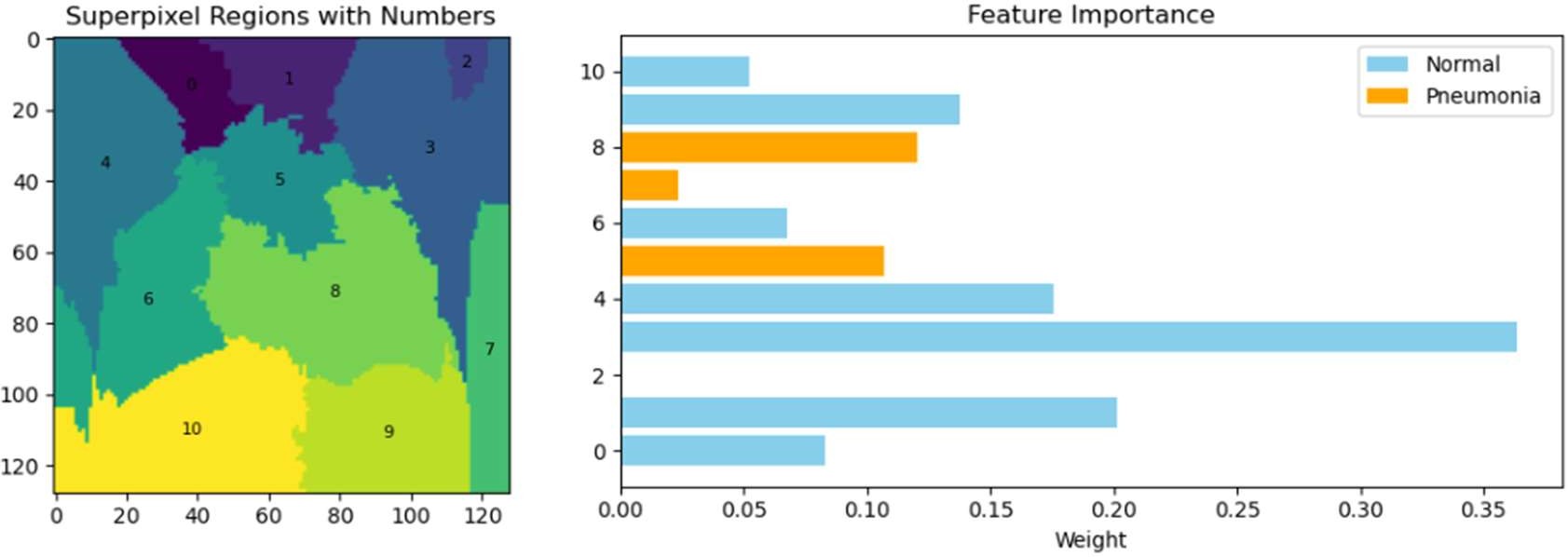


FIGURE 4: XAI Output Using SLIC Superpixel Method

In the case of SLIC (Simple Linear Iterative Clustering), the XAI output mirrors that of Quick Shift, showcasing transparency and interpretability in pneumonia detection. The original X- ray, LIME mask, and model prediction images contribute to a comprehensive view of the model's decision process. The superpixel image, complete with assigned numbers and feature weights, accentuates the areas significant for pneumonia detection. Despite yielding similar performance to Quick Shift in Jaccard Coefficient, SLIC remains a reliable method for producing interpretable visualizations in our project.

**XAI Output Using Quickshift with Noise Superpixel Method**

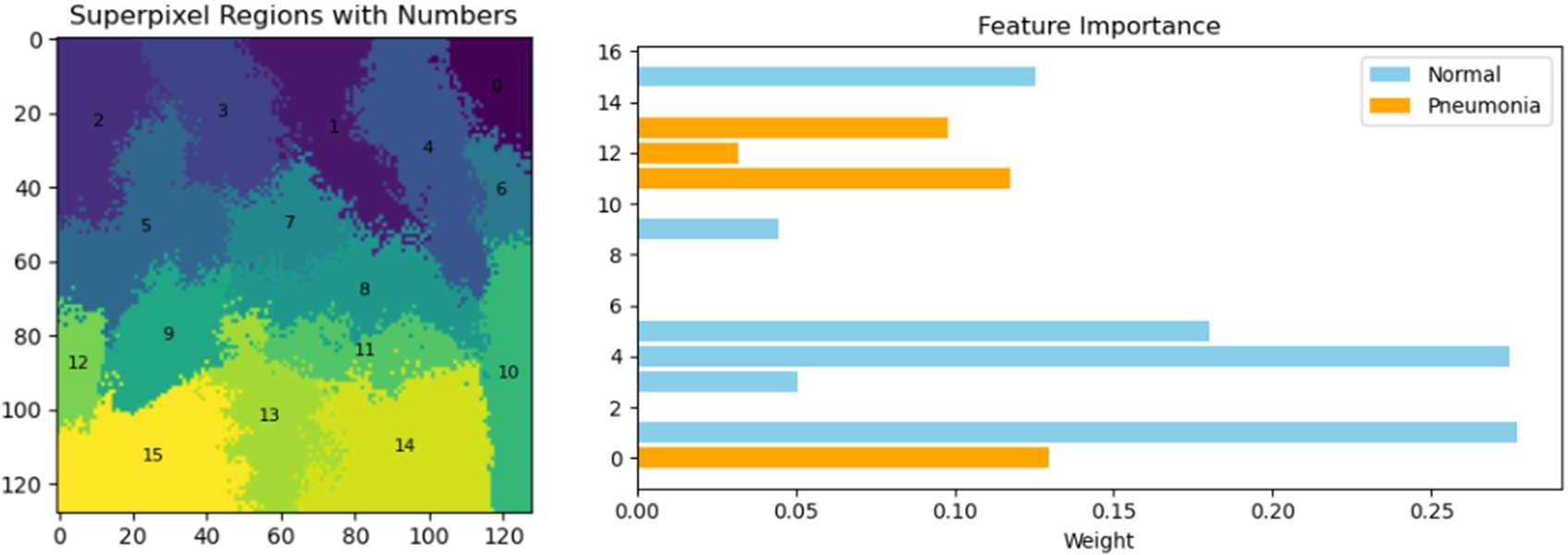
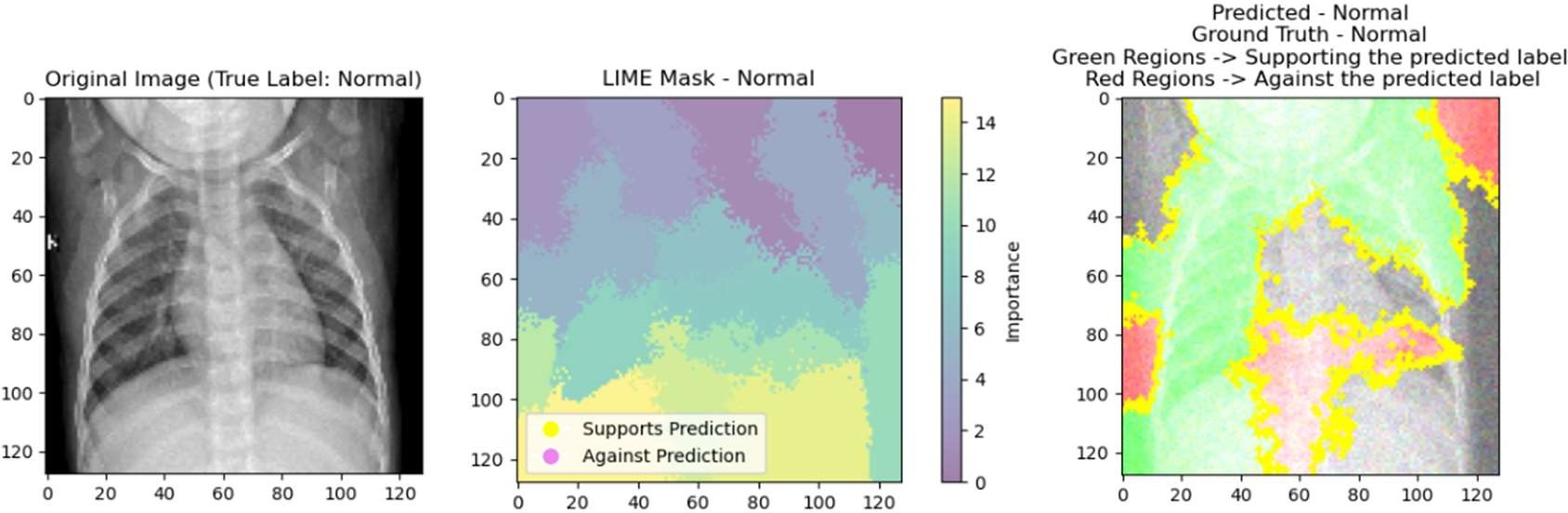


FIGURE 5: XAI Output Using Quick Shift with noise Superpixel Method

The introduction of noise in the Quick Shift superpixel method significantly enhances the interpretability and performance of our pneumonia detection system. The XAI output demonstrates superior results in Jaccard Coefficient compared to both Quick Shift and SLIC. The original X-ray, LIME mask, and model prediction images maintain their roles, while the superpixel image, enriched with noise, offers a more refined and nuanced understanding of the model's decision boundaries. The incorporation of noise proves to be a pivotal enhancement, providing richer insights and elevating the interpretability of the CNN models in pneumonia detection.

# Conclusion and Future Work

XAI techniques significantly improve the interpretability of pneumonia prediction models, making their decision-making process more transparent.

The superpixel method, integrated into XAI, excels in pinpointing critical image regions relevant to pneumonia prediction, aiding in the identification of key features.

The combined use of XAI techniques and the superpixel method contributes to building trust in pneumonia prediction models by providing a clearer understanding of how predictions are derived.

The interpretability achieved through XAI, especially with superpixel localization, enhances the clinical relevance of the predictions, allowing healthcare professionals to make more informed decisions.

XAI techniques, including the superpixel method, offer a valuable tool for validating model outputs, ensuring that predictions align with clinically meaningful patterns and contributing to the robustness of the predictive model.

## Future Work:

In future work for our pneumonia detection project, we aim to further enhance the robustness and user-friendliness of our system. Firstly, we plan to integrate additional superpixel methods to expand the interpretability of our Convolutional Neural Network (CNN) models. Exploring **methods beyond Quick Shift, SLIC, and Quick Shift with noise** could provide richer insights into the model's decision-making process. Additionally, we envision the development of a **user-friendly interface (UI)** to facilitate seamless interaction for healthcare professionals and researchers. A well-designed UI will not only streamline the process of inputting and analyzing chest X-ray images but will also offer a more intuitive platform for interpreting the XAI outputs. These advancements will contribute to the continual improvement and practical applicability of our pneumonia detection system.

# REFERENCES

1. Brima, Y., Atemkeng, M., Tankio Djiokap, S., Ebiele, J., & Tchakounté, F. (2021). Transfer learning for the detection and diagnosis of types of pneumonia including pneumonia induced by COVID-19 from chest X-ray images. Diagnostics, 11(8), 1480.
2. Habib, N. (2020). Fusion of deep convolutional neural network with PCA and logistic regression for diagnosis of pediatric pneumonia on chest X-rays. Network Biology, 10(3), 62.
3. Kareem, A., Liu, H., & Sant, P. (2022). Review on pneumonia image detection: A machine learning approach. Human-Centric Intelligent Systems, 2(1-2), 31-43.
4. Kundu, R., Das, R., Geem, Z. W., Han, G. T., & Sarkar, R. (2021). Pneumonia detection in chest X-ray images using an ensemble of deep learning models. PloS one, 16(9), e0256630.
5. Sourab, S. Y., & Kabir, M. A. (2022). A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms. Sensors International, 3, 100167.
6. https://[www.geeksforgeeks.org/introduction-to-explainable-aixai-using-lime/](http://www.geeksforgeeks.org/introduction-to-explainable-aixai-using-lime/)