AI-Powered Chest Disease Detection and Classification using Deep Learning

P.R.N.Srivalli[1], S.Sai Navya Sri[2], B.Hemanth[3], T.Aasrith Varma, R.Rajender[5]

[1-4]B.Tech Student , [5] Professor & HOD of CSSE Dept ,LIET

[1,2,3,4,5] Computer Science and System Engineering ,Lendi Institute of Engineering and Technology, Vizianagaram

---***---

ABSTRACT

AI-Powered Chest Disease Detection and Classification using Deep Learning Method presents an innovative approach to revolutionize chest disease diagnosis through the integration of advanced deep learning techniques. This initiative focuses on the development of an intelligent system that can effectively detect and classify a variety of chest diseases by analysing medical images, such as chest X-rays and CT scans. At its core, the project involves the compilation and preparation of a diverse dataset containing both normal and disease-affected chest images. Leveraging the capabilities of deep learning, particularly convolutional neural networks (CNNs), the project then designs and trains models to learn intricate patterns and features indicative of different chest diseases. The models are optimized for accuracy and robustness. This efficiency translates to quicker diagnoses, enabling prompt medical interventions. The AI system serves as a complementary tool for healthcare professionals, aiding in decision-making by providing valuable insights and augmenting their expertise.

1.INTRODUCTION

The pressing demand to improve chest disease diagnostics arises from the significant impact of respiratory illnesses on global mortality. WHO data for 2020 reveals that in India alone, lung disease deaths reached 879,732, constituting 10.38% of total deaths and ranking the country third globally with an age-adjusted death rate of 87.90 per 100,000. Tuberculosis (TB), a formidable global health challenge, claimed 1.3 million lives in 2022, making it the second leading infectious killer after COVID-19, surpassing even HIV and AIDS.

This paper introduces an innovative initiative, "AI-Powered Chest Disease Detection and Classification using Deep Learning Method," aimed at revolutionizing the diagnosis of chest diseases through advanced deep learning techniques. The foundation of this project lies in the compilation and preparation of a comprehensive dataset, meticulously curated to include both normal and disease-affected chest images. The dataset, sourced from various repositories and meticulously refined to eliminate redundancy, is augmented to facilitate robust model training. This augmentation process results in a dataset comprising 10,000 images, encompassing four types of lung diseases—Covid-19, Tuberculosis, Viral Pneumonia, and Bacterial Pneumonia—alongside a folder representing normal lung conditions.

Central to the project's methodology is the utilization of deep learning techniques, particularly convolutional neural networks (CNNs). These networks are designed and trained to discern intricate patterns and features within the chest X-ray images indicative of different chest diseases. The models are meticulously optimized for both accuracy and robustness, ensuring reliable performance in disease detection.

The significance of this approach lies in its potential to expedite diagnoses, allowing for prompt medical interventions. By leveraging the efficiency of AI-powered disease detection, this system serves as a valuable complementary tool for healthcare professionals. It aids in decision-making by providing insights derived from the analysis of medical images, thereby augmenting the expertise of healthcare practitioners.

The dataset, a critical component of this project, not only enables the training of the deep learning models but also serves as a representative collection of chest images capturing diverse disease manifestations. The diseases of focus—Covid-19, Tuberculosis, Viral Pneumonia, and Bacterial Pneumonia—hold particular relevance in the current medical landscape, with each contributing to the global burden of respiratory illnesses.

In subsequent sections, we delve into the technical aspects of the project, including the development of a CNN model with multiple convolutional layers. This model is seamlessly integrated with a user-friendly front end, allowing users to input chest X-rays. The backend of the system processes the images, detects diseases, and presents the results on the front end. Additionally, we explore the impact of these diseases on mortality rates and highlight the advantages of utilizing X-rays over CT-scans in the context of disease diagnosis. This paper aims to contribute to the growing body of knowledge in AI-driven healthcare solutions and underscores the potential of such innovations in improving diagnostic accuracy and healthcare outcomes.

2. PROPOSED METHOD

Our methodology revolves around leveraging this augmented and diverse dataset to construct a robust model for efficient disease detection from chest X-ray images. We initiate the process with meticulous preprocessing to enhance image quality. Subsequently, advanced deep learning techniques are applied to extract intricate patterns and features. The model is then trained and rigorously evaluated, considering metrics like accuracy, precision, recall, and F1 score to ensure effectiveness.

Our plan for detecting lung diseases involves using advanced computer techniques to analyze medical images accurately. Here's how we'll do it:

Dataset Acquisition:
We have a good mix of different chest diseases like pneumonia, tuberculosis, and COVID-19 in our pictures.

Data Preprocessing:
Preprocessing steps are performed to enhance the quality of the data. This includes resizing the images to a standard size (64x64 pixels in this case), and possibly other preprocessing steps like normalization, removing green pixels, converting RGB images to grayscale.

Feature Extraction:
Features are extracted from the preprocessed images to represent the characteristics of different lung diseases. The code extracts various types of features, including GLCM (Grey Level Co-occurrence Matrix) features, texture features, and histogram features.

Model Development:
We selected Support Vector Machine (SVM) and a Convolutional Neural Network (CNN) which works best for image classification.

Model Training:
We use the features extracted from lung disease images as input data for training both models. To prevent overfitting and ensure that our models generalize well to new data, we split our dataset into two parts: a training set and a validation set.

Model Evaluation:
Once our SVM and CNN models are trained, we test them using a different dataset. We measure accuracy, precision, recall, and F1-score to see how well they classify lung diseases. Comparing these measures helps us pick the best model for our task. We might also use confusion matrices to visualize how accurately our models distinguish between different lung diseases. This thorough evaluation ensures we choose the most effective model, improving our ability to diagnose lung diseases accurately.

2.1 . ALGORITHM

In our project, two algorithms are utilized for lung disease classification: Support Vector Machine (SVM) and Convolutional Neural Network (CNN).

Support Vector Machine (SVM):
In this project, the SVM model is implemented using the scikit-learn library, a popular tool for machine learning tasks in Python. Unlike neural networks, SVM does not involve layers; instead, it utilizes a hyperplane to delineate classes within the feature space. By identifying the optimal hyperplane, SVM aims to maximize the margin between distinct classes, enhancing classification accuracy. The model is trained using features extracted from lung disease images along with their corresponding labels, enabling it to discern between different lung diseases effectively. **Convolutional Neural Network (CNN):**
Convolutional Neural Network (CNN) is a smart tool specifically made to understand and classify images. Using the Keras library, we design the structure of the CNN model. This model has different layers like the ones that help in spotting patterns in images, others that shrink the information, and ones that connect all the learned patterns to make decisions. By training the CNN

with features we extract from lung disease pictures and their labels, it learns to tell apart different lung diseases, which helps doctors diagnose them better.

Input Layer:
The input layer receives feature vectors representing lung disease images with a size of 216 pixels in height and 1 pixel in width.:

Convolutional Layer:
The convolutional layer applies 32 filters with a size of 1x1, using ReLU activation to extract hierarchical features from the input.

MaxPooling Layer :
The max-pooling layer downsamples the feature maps obtained from the convolutional layer, reducing dimensionality while preserving essential information.

Convolutional Layer :
Another convolutional layer further extracts features from the downsampled feature maps using 32 filters with a size of 1x1 and ReLU activation.

MaxPooling Layer:
The subsequent max-pooling layer further reduces dimensionality in the feature maps obtained from the second convolutional layer.

Flattening Layer:
The flattening layer reshapes the output from the previous layers into a one-dimensional vector, preparing it for input into the fully connected layers.

Dense Layer:
The dense layer consists of 256 units with ReLU activation, combining features learned by the convolutional layers to make classification decisions.

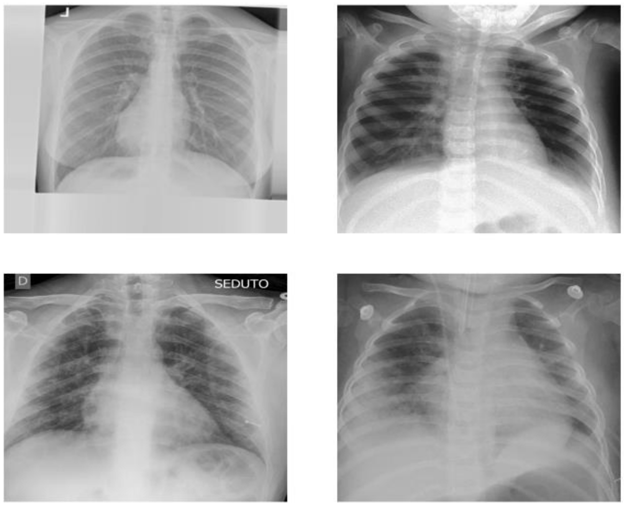
Output Layer (Dense):
The output layer, with 5 units corresponding to lung disease categories, employs a softmax activation function to produce classification outputs, assigning probabilities to each category based on the extracted features.

ReLU activation function:
ReLU activation function in our project sets negative values to zero and leaves positive values unchanged. It helps in introducing non-linearity to the model, allowing it to learn complex patterns in the data effectively.

Softmax activation:
Softmax activation in our project normalizes output values into probabilities. It ensures that the sum of probabilities for all classes equals 1, enabling the model to predict the likelihood of each disease category accurately based on extracted features.

2.2 . DATASET STUDY :

Our dataset comprises chest X-ray images, meticulously compiled from various sources with a thoughtful consolidation process. The collection is diverse, incorporating images from different datasets, seamlessly combined to enhance representativity. It encompasses four primary lung diseases—Covid-19, Tuberculosis, Viral Pneumonia, and Bacterial Pneumonia—alongside a dedicated folder portraying normal lung conditions. To amplify the dataset's richness and ensure effective model training, we augmented it sixfold, resulting in a substantial dataset of approximately 10,000 images.

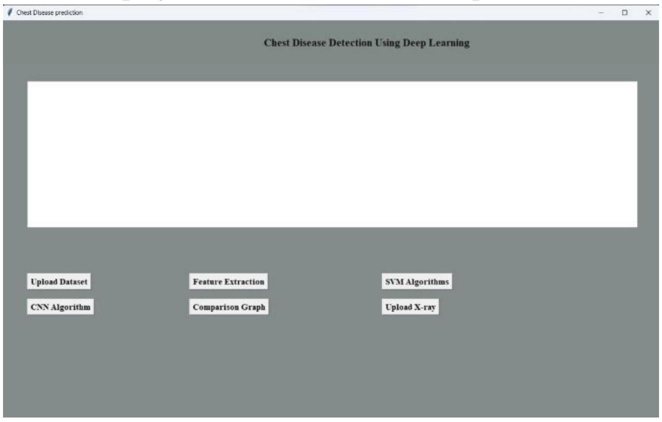


The following images shows different types of lung diseases:
1. Bacterial Pneumonia
2. Corona Virus Disease
3.Tuberculosis
4.Viral Pneumonia

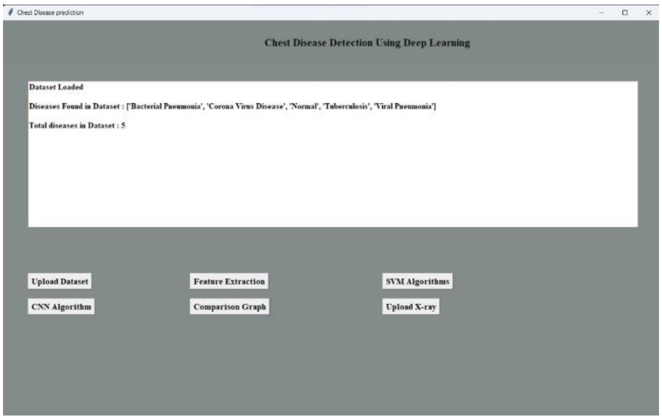
2.3 .RESULTS:

The results of our project include the evaluation metrics obtained from testing the SVM and CNN models. These metrics, such as accuracy, precision, recall, and F1-score, provide insights into the performance of each model in classifying lung diseases. Additionally, visualizations like confusion matrices are presented to illustrate the model's ability to distinguish between different disease categories. Comparisons between the SVM and CNN models can be made to determine which one performs better for our specific classification task. The results highlight the effectiveness of our models in accurately predicting lung diseases based on extracted features from medical images.

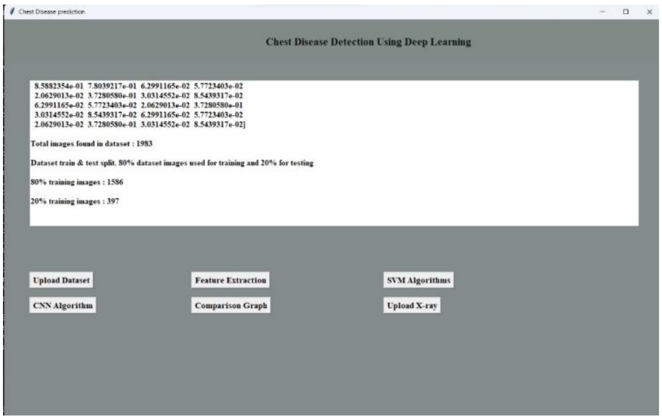
First it displays the user interface with specified buttons.



Now its time to upload the dataset. By clicking on the Upload dataset button we can select the dataset to be entered. The it displays the details of the dataset on the user interface.

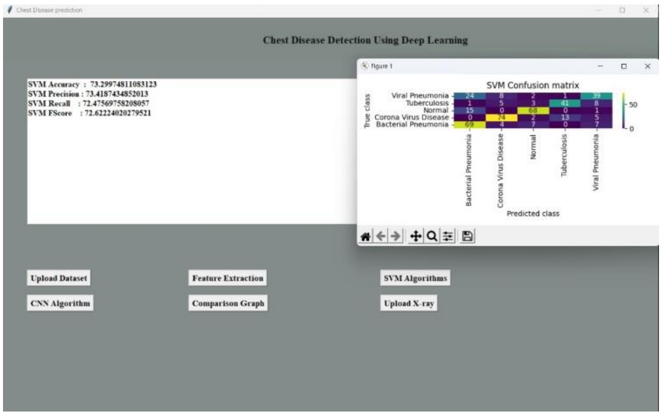


By clicking on the Feature Extraction button we can see what features are extracted and how the dataset is divided into training and testing parts.

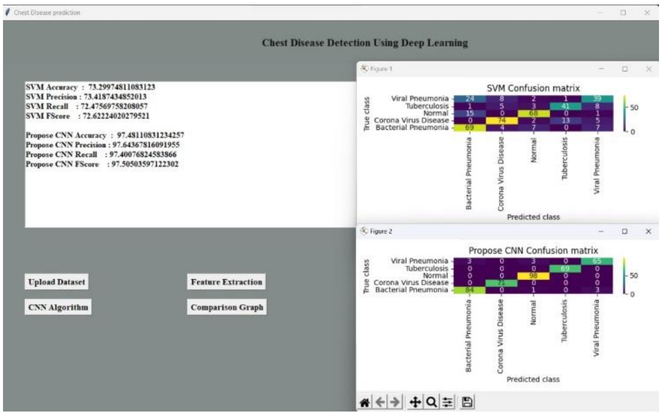


Now it's time for algorithms. We used SVM and CNNs in our project. Confusion matrix shows the model's ability to distinguish between different disease categories.

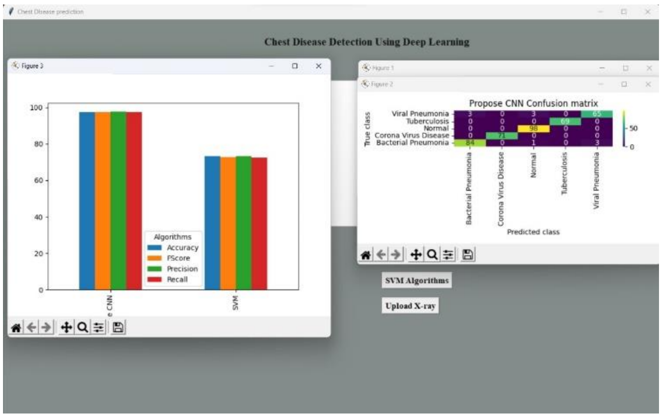
To visualize the SVM performance we have to click on the SVM Algorithm button.



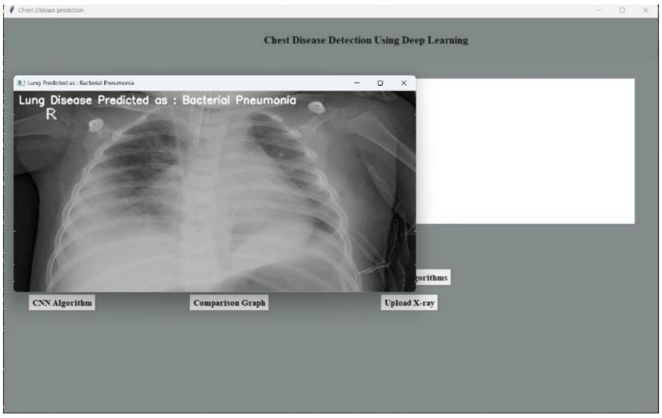
Now to see the details of CNN we have to click on the CNN Algorithm button. Then the confusion matrix along with the details is displayed on the screen.



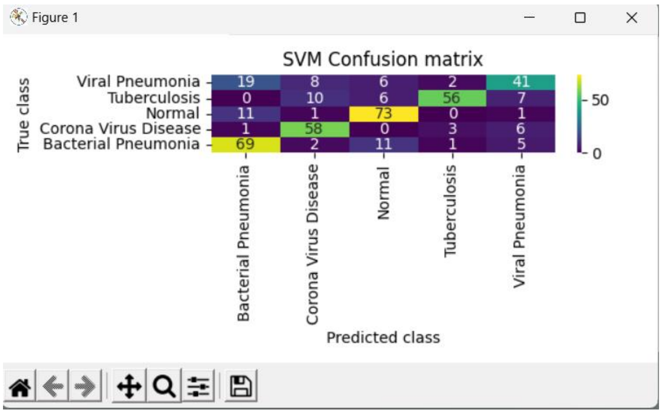
Comparison Graphs of SVM and CNN models can be made to determine which one performs better for our specific classification task.



Its time for our final result. By clicking on the Upload X-ray button you are allowed to select the X-ray that you want to test. By analysing the X-ray the system shows the predicted lung disease name along with the X-ray.

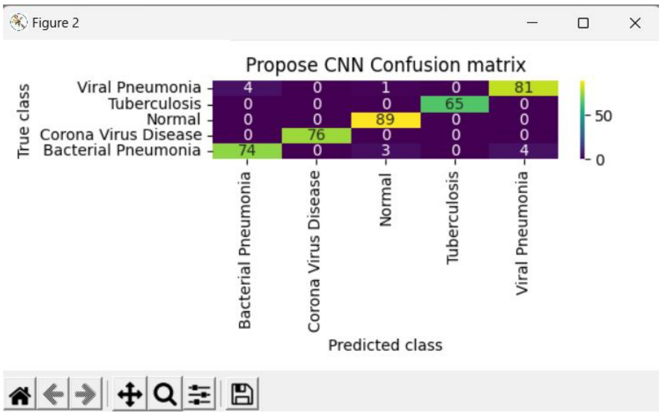


SVM confusion matrix**:**

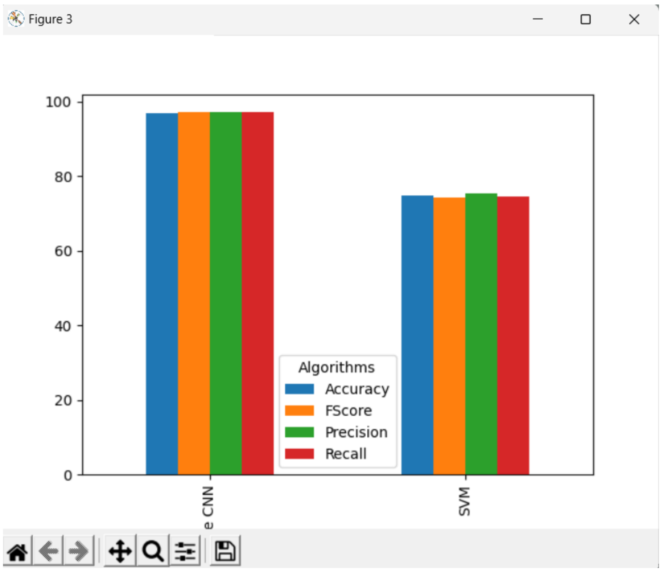


This confusion matrix shows how well a machine learning model classified chest X-rays into four categories: viral pneumonia, tuberculosis, bacterial pneumonia, and normal. Most cases were classified correctly, but some were misclassified. For example, 58 cases of COVID-19 were classified as normal.

CNN confusion matrix :



The image shows a proposed CNN confusion matrix, which is used to compare the frequency of certain diseases. It compares four classes: viral pneumonia, tuberculosis, normal, and corona virus disease. Most cases were classified correctly. Comparison graph of SVM and CNN:



This accuracy graph shows how efficient is CNN compared to SVM.

CONCLUSIONS

In conclusion, our project presents a highly accurate and versatile system for lung disease detection and classification. Achieving a remarkable accuracy of 96-97%, it proficiently categorizes chest images into bacterial pneumonia, viral pneumonia, tuberculosis, COVID-19, and normal lung cases. Featuring a user-friendly interface, the system ensures ease of use for healthcare professionals, facilitating streamlined diagnostics. With its comprehensive approach and intuitive design, our project promises to significantly augment clinical decision-making and patient care in respiratory healthcare, showcasing the potential for enhanced diagnostic accuracy and improved outcomes in lung disease management.

REFERENCES

[1] Abobaker Mohammed Qasem Farhan & Shangming Yang (2023), Automatic lung disease classification from the chest X-ray images using hybrid deep learning algorithm.

[2] J. Juditha Mercina, J. Madhumathi, V. Priyanga, M. Deva Priya, (2019), A Deep Convolutional Neural Network Based Lung Disorder Diagnosis, 5,2.

[3] Omkar Gaikwad, Divyanshu Tripathi, Madhuri Dange, Harshada Mohite, Prof. Pallavi Shimpi, (2021), Detection of Lungs Infection Using Convolutional Neural Network, 8.

[4] Saravanan. M, Manoj Kumar, Dilsha Vijay, Gayathri. K and Jenifar. A, (2022), Chest X-Ray based Lung Disease Detection using Convolutional Neural Network.

[5] Chandrasekar KS (2022) Exploring the deep-learning techniques in detecting the presence of coronavirus in the chest X-ray images: a comprehensive review. Arch Comput Methods Eng 29:5381–5395.