IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Hybrid CNN-Xgboost Framework For Multi-Class Pneumonia Detection And Real-Time Web-Based Diagnosis

[1]Sriraksha Devaraj, [2]Nuthanya V, [3]Usha A, [4]Nandini R, [5]Vivekanand M S

[1]Assistant Professor, [2]UG Student, [3]UG Student, [4]UG Student, [5]UG Student,
Electronics and Communication Engineering,
Sri Siddhartha Institute of Technology SSAHE, Tumkur, India

Abstract: Pneumonia is a leading cause of death worldwide, particularly in those with immune systems that are weak, and lack of trained radiologists frequently makes early diagnosis difficult. To facilitate faster and more accurate detection, this study develops a hybrid system that classifies chest X-rays into Normal, Bacterial and Viral pneumonia using the Extreme Gradient Boosting (XGBoost) and extracts features using the Convolutional Neural Network (CNN). Developing the Real-time web based application for instant diagnosis through the image uploads, resolving class imbalance to increase dependability and robust classification model. A balanced set of data was taken from Kaggle and used for both training and testing. Image enhancement techniques are used to improve the model's performance and provide more precise forecasts with an overall 89.46% accuracy rate the model demonstrated its potential as a valuable tool for rapid and accurate pneumonia detection which can significantly enhance patient outcomes.

Index Terms:**CNN, Diagnosis, Extreme Gradient Boosting, Image Augmentation, Kaggle, Pneumonia, XGBoost .**

I. Introduction

A dangerous lung illness called pneumonia induces the expansion of the alveoli which causes fluid accumulation and difficulty inhaling for small children the elderly and those with impaired immune systems it is one of the main causes of death there aren't enough skilled radiologists to assist in many rural or low-resource locations despite the fact that early and precise diagnosis is crucial in order to address this issue medical imaging is increasingly employing artificial intelligence ai and machine learning permitting clinicians to diagnose pneumonia more quickly and provide patients with better care.

II. MOTIVATION

Pneumonia is a critical health condition that requires immediate and accurate observation of its development. Machine learning offers effective support in the diagnostic process. However, there is an increasing need for automated systems that can efficiently process diverse patient data, allowing for timely and accurate detection. This is essential for improving patient care, especially since many current methods lack the flexibility and consistency needed for wide adoption in the healthcare field.

III. OBJECTIVES

- Developing an automated method that uses chest X-ray dataset to detect pneumonia.
- Improving classification accuracy by using DL based feature extraction and image preprocessing methods.
- Assessing and examining the outcome using multiple algorithms to determine the most correct approach.
- Putting into practice a fully developed online platform that enables a real-time diagnostics.

IV.PROBLEM STATEMENT

Manual diagnosis is often inconsistent and slow with limited access to radiologists. Existing ML models struggle with generalization, class imbalance, and inconsistent accuracy, highlighting the need for a more reliable and efficient diagnostic system.To address diagnostic delays and inconsistent model performance, a ML approach using CNNs and XGBoost is proposed to improve accuracy. The model incorporates class imbalance handling, and a web-based interface enables chest X-ray uploads with instant results, ensuring accessibility across diverse healthcare environments.

V.PROPOSED SYSTEM

This study employs advanced artificial intelligence methodologies, including ml and dl techniques, to create a sophisticated system for pneumonia detection.The primary aim is to transform the diagnostic process, enhance the precision, and eventually increase patient outcomes.The technologies shown below in block diagram can be used to analyze large datasets of chest X-rays in the area of medical imaging, which can help in the more precise and effective detection of anomalies associated to pneumonia. The methods, especially CNNs, have made substantial progress in medical picture categorization using XGBoost and feature extraction.

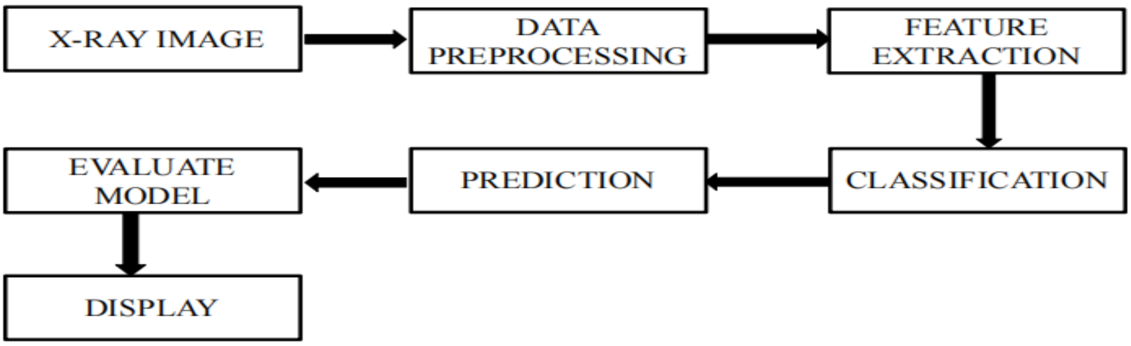

Fig1: Proposed system.

VI.METHODOLOGY

The dataset contains chest X-ray images divided into three categories: Normal, Bacterial Pneumonia, and Viral Pneumonia. To maintain data integrity, all duplicate and corrupted images were excluded. Each image was resized to a consistent format of 224x224x1 pixels for standardization. To enhance the model's accuracy and generalization, data augmentation methods such as image rotation, horizontal flipping, contrast adjustment, and the injection of random noise were utilized. Additionally, pixel values were normalized to the [0, 1] range to accelerate and optimize the training process.
For feature extraction, a Convolutional Neural Network (CNN) was employed, enabling the automatic identification of key visual patterns within the X-ray images. The CNN architecture includes convolutional layers to capture features, ReLU activation functions to introduce non-linearity, and pooling layers to reduce the spatial dimensions of the feature maps. The final CNN output is flattened into a one-dimensional vector that encapsulates the critical features of the image. This vector is then passed to an XGBoost classifier, chosen for its effectiveness with imbalanced data and its robustness against overfitting. The classifier assigns each image to one of the three defined classes.
To make the system user-friendly and accessible, a web interface was built using Flask. This application enables users to upload chest X-ray images and receive immediate diagnostic feedback, supporting faster and more convenient pneumonia detection.

VII. SYSTEM REQUIREMENTS

- **Kaggle Datasets (Chest X-rays):** online library filled with pictures of people's chests. Some pictures how healthy lungs, and others show lungs with pneumonia. These pictures are like practice material for a computer to learn what pneumonia looks like.
- **Python 3.8:** Think of this as a powerful and user-friendly computer language. It's got lots of built-in tools that make it good at learning from data and building smart programs.
- **Anaconda Prompt:**This is like a special command center on your computer.It helps you manage different sets of tools and keep them organized, especially for working with Python.
- **Flask 2.2.5:** This is a simple set of tools for building websites. It helps connect the smart computer program you built to a webpage so people can actually use it.
- **HTML, CSS, and JavaScript:** These are the basic building blocks for any website you see. HTMLcreates the structure, CSS makes it look good with colors and styles, and JavaScript makes it interactive, like buttons that do things when you click them.

VIII.SYSTEM IMPLEMENTATION

Data Collection and Preprocessing:
Three distinct classifications-Normal, Bacterial pneumonia, and Viral pneumonia have been created with the dataset of chest X-ray images to make sure adequate representation sourced from Kaggle, the initial preprocessing stage involved the removal of corrupted and redundant images.All images were subsequently resized to 224x224x1 pixels of uniformity. To boost the data diversity and model robustness, data augmentation techniques such image rotation, horizontal flipping, contrast enhancement, and noise insertion were applied. Furthermore, pixel values were standardized, typically to the range of [0, 1], to enhance training efficacy. These preprocessing procedures improve the dataset's overall quality and its suitability for CNN analysis.

Fig.2. 224x224x1 Dataset resolution

Feature Extraction Using CNN: The system employs CNNs for feature extraction from chest X-ray images. The Convolutional layers within the CNN detect the essential patterns, and ReLU activation introduces non-linearity. Pooling layers reduce dimensionality, focusing on the key image features. The CNN output is transformed into a flattened feature vector, which encapsulates the image's key characteristics.This extracted feature vector is then passed to the XGBoost classifier. Essentially, CNN learns and extracts useful data from the chest X-ray images on it's own, which renders it simpler for XGBoost to classify the images into Normal, Bacterial, and Viral Pneumonia.

Classification using XGBoost: The XGBoost is employed as the classifier to identify the final class of the chest radiograph image. Its ability to effectively manage the class imbalance and avoid data overfitting, which are prevalent in the medical datasets, led to its selection. The XGBoost receives the feature vector extracted by the CNN as input. This vector represents the image features that are analyzed by the XGBoost to predict the image’s category. The classifier provides an output corresponding to one of three categories: Normal, Bacterial pneumonia, or Viral pneumonia.The XGBoost provides an accurate diagnosis by interpreting the CNN’s feature analysis.

Evaluate the model: The model's performance is evaluated using metrics such as the confusion matrix, accuracy, precision, recall, and F1-score. These metrics evaluate the overall accuracy and efficiency of the model in identifying cases of normal, bacterial, and viral pneumonia.

Web Application: A user-friendly web application, built with usage of flask, provides an interface for pneumonia detection. The user interface is built with HTML, CSS, and JavaScript, enabling users to upload X-ray images.The backend then processes these images, performing preprocessing, feature extraction with a CNN and classification using XGBoost, The system rapidly predicts the output and displays the diagnosisNormal, Bacterial or Viral pneumonia within the web browser. This application streamlines the diagnostic process, making pneumonia detection more accessible and efficient for both medical professionals and patients.

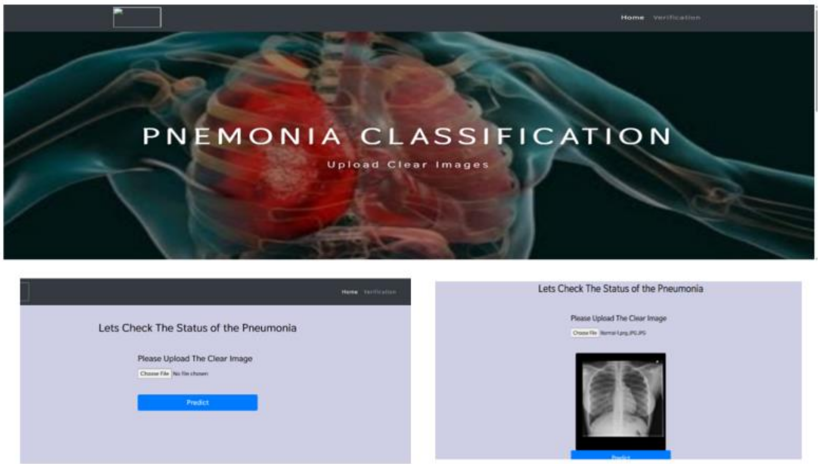

Fig.3. User interface

IX.RESULTS AND DISCUSSIONS

The table.1 shows the developed system reached an overall accuracy of 89.4% with strong precision, recall and F1-scores across Normal, Bacterial and Viral pneumonia classifications which demonstrates the robust performance.While many existing studies focus on the binary classification, this research addresses the more complex multi-class problem. Earlier models frequently exhibited shortcomings like limited generalization capabilities, inadequate data augmentation, and lower accuracy levels.In the contrast, this system utilizes CNNs for feature extraction and XGBoost for classification, emphasizes on the real world applicability by including a web based interface for the immediate diagnosis.

Table.1.Performance Metrices Of XGBoost :

sCLASS	PRECISION	RECALL	F1-SCORE	SUPPORT
Normal(0)	0.90	0.91	0.91	398
Bacterial(1)	0.86	0.85	0.86	399
Viral(2)	0.92	0.91	0.92	403
Accuracy			**0.894**	1200
Macro avg	0.89	0.89	0.89	1200
Weightned avg	0.89	0.89	0.89	1200

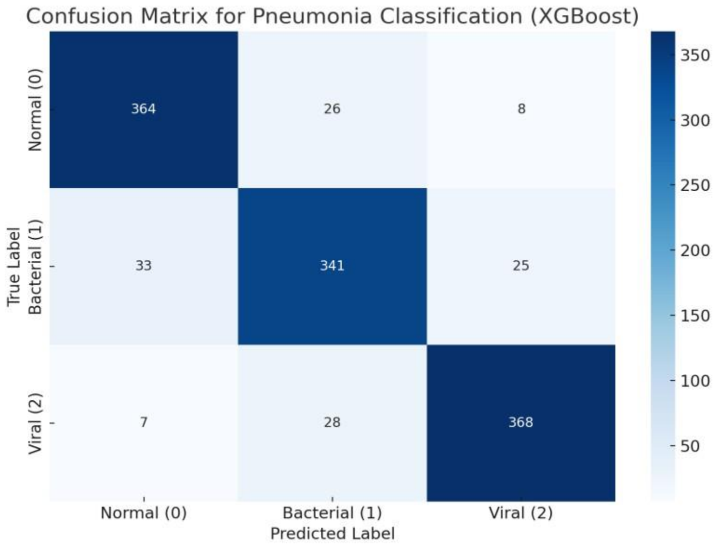

Fig.4.Confusion matrix

A popular tool for evaluating a machine learning model's efficiency is a confusion matrix. It presents a tabular overview of the prediction findings, with each column denoting the actual class and each row representing the predicted class.This structure helps in analysing how well the model distinguishes between the various classes

Fig.5.Accuracy plot

The plot shows training and validation accuracy over the 10 epochs. The training accuracy gradually improves from 0.70 to 0.90, indicating the learning of the model. validation accuracy also rises and reaches around 0.89 at the 10^{th} epoch.

X.MERITS

- Automates pneumonia detection, reducing doctor workload and diagnostic time.
- Flask-based web interface allows user-friendly access for testing the model.
- Reduction of Human Error

XI.DEMERITS

- Overfitting
- Data insufficiency

XII.CONCLUSION

The research is successfully implemented a ml model for detecting pneumonia using chest X-ray images. The system leverages CNN for efficient feature extraction and XGBoost for precise classification into Normal, Bacterial, and Viral pneumonia categories. The results showcase the model's proficiency in attaining high accuracy in pneumonia detection, effectively tackling issues such as class imbalance and stablishing a resilient classification framework. A key achievement is the creation of a Flask-based web application.To improve on the system, further study could focus on on expanding the dataset with a wider range of chest X-ray images, which could enhance the adaptability and generalization of the model.Additionally exploring the integration of other advanced deep learning architectures or techniques could lead to even higher accuracy and more refined diagnostic capabilities.

REFERENCES

[1].V. Cheekuri, M. Veeramachaneni, V. M. Manikandan and R. K. Kumar, "A Machine Learning-Based Pneumonia Detection System," 2024 5th International Conference for Emerging Technology (INCET), Belgaum, India, 2024.

[2].Das, S. Kumar Pradhan, S. Mishra, S. Pradhan ands P. K. Pattnaik, "A Machine Learning based Approach for Detection of Pneumonia by Analyzing Chest X-Ray Images," 2022 9th (INDIACom), New Delhi, India, 2023.

[3].G. Zhang, X. Li, D. Zhang and T. T. Toe, "Pneumonia Detection Method based on Convolutional Neural Network," 2023 ICAIBD, Chengdu, China, 2023.

[4].S. D. Mazumdar, A. Kumar and R. Sethuraman, "Pneumonia Detection using Chest X-ray Images using CNN Algorithm," 2023 ICICCS, Madurai, India, 2023.

[5].Kadali, B. S. Pudi, K. A. Shaik, A. Janjam and J. Javvadi,"Pneumonia Detection of Chest X-Ray Images by using Resnet-50 Machine Learning Algorithm", (ICAIS), Coimbatore, India, 2022.

[6].Shah, H. Mehta and P. Sonawane, "Pneumonia Detection Using Convolutional Neural Networks," 2020 and Inventive Technology (ICSSIT),Tirunelveli, India, 2020.