To improve pneumonia detection accuracy by combining the strengths of Convolutional Neural Networks (CNN) and Random Forest classifiers.

Amarjeet Kumar
Dept of CSE
Chandigarh University
Mohali, India
ak7462514@gmail.com

Ujjwal Rai
Dept of CSE
Chandigarh University
Mohali, India
ujjwalrai6666@gmail.com

Manav Agrawal
Dept of CSE
Chandigarh University
Mohali, India
aggarwalm363@gmail.com

Ajay Pal Singh
Dept of CSE
Chandigarh University
Mohali, India
apsingh3289@gmail.com

Abstract-- **Pneumonia poses a critical global health challenge, necessitating accurate and prompt diagnosis to facilitate effective treatment. This research paper introduces a Python-based predictive model employing Convolutional Neural Networks (CNNs) to detect and classify pneumonia from chest X-ray images. Leveraging advanced deep learning techniques and implemented using TensorFlow and Keras, the proposed model automates pneumonia detection with notable precision. Evaluated on a comprehensive dataset, the CNN model achievedan impressive accuracy rate of 98.54%, demonstrating its ability to effectively differentiate between pneumonia-positive and normal images. The model's performance, supported by additional metrics such as precision, recall, and AUCROC, underscores its potential to rival established diagnostic methods and experienced radiologists. This study highlights the transformative potential of CNNs in enhancing healthcare diagnostics, paving the way for rapid, consistent, and reliable pneumonia detection systems. As medical imaging and deep learning technologies evolve, this research contributes significantly to improving patient care and optimizing healthcare system efficiency.**

Keywords- Convolutional Neural Networks(CNN), Random Forest classifier, TensorFlow, Keras, SVM

I. INTRODUCTION

A frequent respiratory illness that has the potential to be fatal, pneumonia continues to be a major worldwide health concern and influences public health. A timely and precise diagnosis of pneumonia is essential for both patient concerns and successful treatment. Traditionally, the view of pneumonia has mainly reckoned with clinical and radiological interpretations by medical practitioners. These approaches can be time-consuming and are susceptible to interobserver variability, despite their shown effectiveness.

This research paper embarks on a journey to explore the application of Python in predictive modeling for pneumonia diagnosis, with a novel approach that combines the strengths of CNNs and Random Forest classifiers. Significant progress has been made in analyzing medical images in recent years with the integration of deep learning. In particular, convolutional neural Networks (CNNs) show great promise in the analysis of medical images, including chest X-rays, foridentifying and classifying various diseases, such as pneumonia. CNNs excel in capturing intricate patterns and features within images, which is essential for the accurate identification of pathological changes in radiological data. However, combining CNNs with machine learning techniques like Random Forests can further enhance prediction accuracy by leveraging the power of ensemble learning.

The main objectives of this study are as follows:

- To develop a robust and highly accurate pneumonia detection system by integrating CNNs with Random Forest classifiers.
- Evaluating the performance of the integrated model on large and variable datasets of chest X-ray images. To compare the model's diagnostic accuracy with that of experienced radiologists and existing diagnostic methods.
- To demonstrate the potential of combining CNNs and Random Forests in aiding healthcare professionals by providing rapid, consistent, and reliable diagnoses.

This research paper is organized as follows: Section II reviews the existing literature in medical image analysis, emphasizing pneumonia detection. Section III outlines the methodology, including the dataset, the architecture of CNN and Random Forest models, the training process, and the metrics used for evaluation. In Section IV, we present experimental results and compare them with current approaches. Section V covers the conclusions and suggests future research directions. Lastly, Section VI offers a summary of the key findings and provides insights into the potential advancements in automated pneumonia detection using CNNs and Random Forest classifiers.

With the growing accessibility of medical data and the swift progress in machine learning and deep learning methods, this study seeks to add to the ongoing efforts to enhance pneumonia diagnosis, thereby improving patient outcomes and optimizing healthcare system efficiency.

II. LITERATURE REVIEW

Lee, J.P, Dean, S.U., Khan, A., and Saboor, A. (2020) [1]: The article presents an advanced heart disease diagnosis system using machine learning and novel feature selection methods. Through algorithms like SVM and FCMIM, it enhance accuracy and reduces execution time, showing promise for efficient implementation in healthcare for timely disease identification.

Nagavelli, U., Samanta, D., & Chakraborty, P. (2022) [2]: This paper addresses heart failure disease detection using various machine-learning approaches. It explores Naïve Bayes for predicting heart disease, analyzes ischemic heart disease localization with SVM and XGBoost, introduces an improved SVM for heart failure identification, and presents a comprehensive heart disease prediction model utilizing DBSCAN, SMOTE-ENN, and XGBoost. The study aims to provide clinicians with an effective tool for early heart problem diagnosis.

O. E., Ezekiel, P. S., & F. B. (2019) [3]: present a heart disease detection model utilizing machine learning algorithms within the framework of Agile Methodology. The research utilized a Heart Dataset to train four different classifiers, with the Decision Tree Classifier attaining the highest accuracy of 98.83%. Implemented using Python and Flask, the model allows for user input-based predictions, showing significant potential for early heart disease detection.

Miao, K. H., Miao, J. H., & Miao, G. J. (2016) [4]: This study aims to improve coronary heart disease diagnosis using an advanced ensemble machine learning approach, specifically by implementing an adaptive Boosting algorithm. The models, evaluated on datasets from diverse sources, achieved superior accuracy compared to prior studies. These ensemble learning models provide dependable and clinically significant diagnoses, especially advantageous for patients in developing areas with limited access to heart disease specialists, potentially saving many lives worldwide.

A.H., Thabtah, F., Md, R.M.A., and Singh. (2019) [5]: This paper outlines the creation of eight models using machine learning to predict mortality among hospital patients with pneumonia during their initial assessment. The model was constructed using data from 9847 patient cases and tests from 4352 additional patients. The first statistical test evaluates the error of each model in predicting survival given different patient survivals(between 0.1 and 0.6). The results show that all models have similar errors, especially when predicting survival rates around 30%. The difference between the models is related to their complexity rather than efficiency, indicating potential for future use in clinical procedures.

Ed-Daoudy, A., & Maalmi, K. (2019, April) [6]: This paper addresses the pressing need for early heart disease detection by proposing a real-time prediction system using Apache Spark. Leveraging streaming big data analytics and machine learning, the system combines Spark MLlib and Apache Cassandra for efficient data processing, classification, storage, and visualization, offering a powerful and cost-effective solution.

Lutimath, N. M., Chethan, C., & Pol, B. S. (2019) [7]: This paper explores the application of machine learning, particularly Naïve Bayes Classification and support vector machines, in detecting heart diseases using the UCI machine learning repository dataset. Focusing on coronary heart disorder, the study utilizes R language for implementation and aims to predict the classification accuracy of patients suffering from heart disease, showcasing the significance of machine learning in healthcare.

Kukar, M., Kononenko, I., Grošelj, C., Kralj, K., & Fettich, J. (1999) [8]: This study addresses the importance of improving diagnostic procedures for Ischaemic heart disease. By applying machine learning techniques, experiments with different algorithms achieved performance similar to that of clinicians, showing improved sensitivity and specificity as evidenced by ROC analysis. The research highlights the potential of machine learning in increasing diagnostic accuracy for heart disease.

Choudhary, G., & Singh, S. N. (2020, October) [9]: This study addresses the challenging task of heart disease diagnosis using machine learning algorithms. Analyzing a vast dataset, the proposed work focuses on identifying crucial features for an effective diagnostic system. Employing decision trees and Ada-Boost algorithms, the research aims to assist doctors in diagnosing heart patients accurately, emphasizing the importance of feature reduction for enhanced classifier performance.

Metersky, M. L., Ma, A., Bratzler, D. W., & Houck, P. M. (2004) [10]: Finding bacteremia predictors was the goal of a study including 13,043 Medicare participants who had pneumonia. Based on the probability of bacteremia, a decision support tool was developed that suggests how many blood cultures to perform. Applying this technique to a validation cohort of 12,771 patients resulted in a 38% reduction in blood cultures while identifying 88-89% of patients with bacteremia. This allows blood cultures to be used more specifically and economically in cases of pneumonia.

Ning, W., Lei, S., Yang, J., Cao, Y., Jiang, P., Yang, Q., ... & Wang, Z. (2020) [11]: This resource contains information on 1,521 patients with pneumonia. Chest CT scans, clinical characteristics, and SARSCoV-2 status are all included. The usefulness of this method for diagnosis and patient management was demonstrated when it was developed using this data to create a deep learning algorithm that could predict the number of cases and fatalities in an independent system validation cohort with accuracy.

Jakhar, K., & Hooda, N. (2018) [12]: Deep Learning, especially DCNN is widely applied in the field of medicine to forecast diseases from extensive and intricate datasets, including X- ray images. For pneumonia, DCNN has demonstrated considerablepotential as a predictive tool. Key characteristics from high- quality X-ray pictures have been uncovered by researchers, whohave also achieved impressive AUC values and an 84% prediction accuracy. In comparison to conventional classifiers like SVM and random forest, DCNN surpasses them across various evaluation metrics. As pneumonia and other illnesses continue to rise, the strategic implementation of deep learning could greatly improve disease prediction in the future.

Huang, J. S., Chen, Y. F., & Hsu, J. C. (2014, June) [13]: This study aims to predict 30-day pneumonia readmissions using data from 520 patients in a Taiwanese hospital. They identified six significant predictors (age, gender, medications, length of admission, comorbidities, and admission cost) and designed a predictive model using RBF-SVM. The model achieved an accuracy of 83.85%, showing promise in identifying high-risk pneumonia readmission cases.

K. R., M., M. P., Y. M. (2021) [14]: The research emphasizes the use of big data and advanced ML, particularly Convolutional Neural Networks (CNNs), to predict pneumonia, a life-threatening lung disease, using chest X-rays. Automating this process is seen as a valuable approach to improving healthcare. Pre-trained CNN models and efficient feature extraction techniques are used to achieve highly accurate results in pneumonia prediction.

Cohen, J. P., Dao, L., Roth, K., Morrison, P., Bengio, Y., Abbasi, A. F., ... & Duong, T. (2020) [15]: The study introduces a model for predicting the severity of COVID-19 pneumonia using frontal chest X-ray images. This tool aims to assist in managing COVID-19 patients, guiding care decisions, and monitoring treatment effectiveness,especially in ICU settings.

Satici, C., Demirkol, M. A., Altunok, E. S., Gursoy, B., Alkan, M., Kamat, S., ... & Esatoglu, S. N. (2020) [16]: Amid the COVID-19 pandemic, early diagnosis is vital due to limited healthcare resources. Chest X-rays are a low-radiation tool for detecting diseases. Deep learning, particularly Transfer Learning with the ResNet50V2 model, successfully predicts pneumonia from 5216 images with 99.69% accuracy. This approach holds promise for early detection of conditions like lung cancer, COVID-19, and heart failure, potentially saving lives.

Harrigan, T., and Miyasaka. (2005) [17]: This study assessed the risk of pneumocystis pneumonia (PCP) in patients with connective tissue diseases who were on medium to high doses of corticosteroids. They determined that the use of immunosuppressants, low lymphocyte counts, and initial steroid dosage were risk factors. In high-risk patients, prophylactic trimethoprim-sulfamethoxazole (TMP/SMX) proved successful in preventing PCP; its usage is advised in these circumstances.

Bodapati, J. D., Rohith, V. N., & Dondeti, V. (2022) [18]: A novel deep neural network model is introduced for pediatric pneumonia detection in chest radiographs. It combines multiple candidate networks, capturing both high-level and low-level features, with an accuracy of 94.84%. This model outperforms existing methods and aids clinicians in pneumonia diagnosis for children.

Singal, B. M., Hedges, J. R., & Radack, K. L. (1989) [19]: This study aimed to establish low-yield criteria (LYC) for reducing unnecessary chest X-ray orders in patients with a low probability of pneumonia. However, the research found that LYC did not significantly improve upon the physician's clinical judgment. The study highlights the challenges in developing guidelines to limit chest X-ray orders based on decision rules.

Thoma, & Antani, S. (2019) [20]: This study focuses on improving pediatric pneumonia diagnosis using CNNs. Emphasizing transparency, it advocates for visualizing and explaining CNN predictions in chest radiographs to enhance reliability and reduce diagnostic errors.

III. METHODOLOGY

A. Data Source:

The dataset comprises 5,856 X-ray images classified as either Pneumonia or Normal. Out of the total images, 1,583 are normal, and 4,273 depict Pneumonia cases. The data is split into 624 images for testing, 5,216 for training, and 16 for validation. The distribution across the validation, testing,and training sets is as follows: (8 normal, 8 pneumonia in validation), (234 normal, 390 pneumonia in testing), and (1,341 normal, 3,875 pneumonia in training).

Fig 1: Normal X-ray pictures of the chest

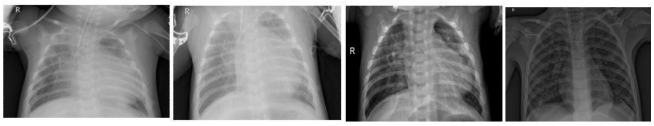

Fig 2: Pneumonia X-ray pictures of the chest

B. Evaluation Metrics

For the research paper, the equations necessary for calculating various performance metrics and operations of the Convolutional Neural Network (CNN) model can be expressed as follows:

Cross-Entropy Loss Function:

The failure rate, often called the cost function, is utilized in machine learning and statistics to assess the discrepancy between predicted outcomes and actual results. This failure is crucial in training a CNN (convolutional neural network) to identify X-ray pictures as either showing pneumonia or being normal. The cross-entropy loss function is typically used for this kind of binary classification to assess the discrepancy between the real label and the anticipated likelihood. By minimizing this loss, the model enhances its accuracy, leading to better classification results. Throughout the training process, this function was fine-tuned to help the CNN differentiate between pneumonia and normal images with an accuracy of 98.54%.

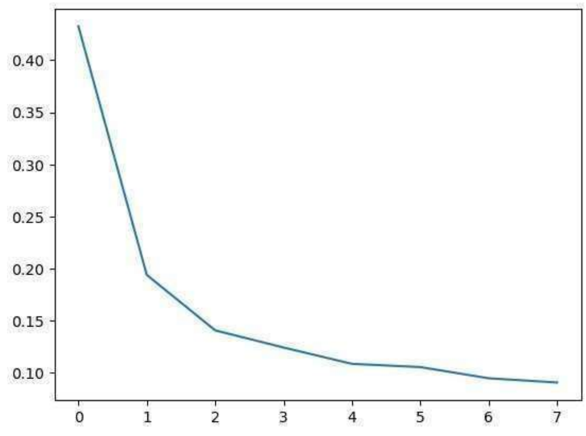

Fig 3: Loss Function Graph

The loss function for binary classification using the cross-entropy can be written as:

$$L = -\frac{1}{N}\sum_{i=1}^{N}[y_i \cdot \log(\hat{y}_i) + (1-y_i)\cdot \log(1-\hat{y}_i)]$$

Where:

- N is the total number of samples,
- y is the actual label (0 for normal, 1 for pneumonia).
- y is the predicted probability,
- L is the loss.

Accuracy:

The proportion of accurately recognized X-ray pictures to all images evaluated is known as accuracy. In this investigation, the convolutional neural network (CNN) demonstrated a 98.54% accuracy rate in differentiating between pneumonia and normal radiographs. This remarkable accuracy demonstrates how well the model can distinguish between the two groups, which makes it a useful tool for lung disease diagnosis.

When paired with additional evaluation measures like accuracy, recall, and AUC-ROC, it offers valuable insights into the dependability and uniformity of the model in diagnostic assessments.

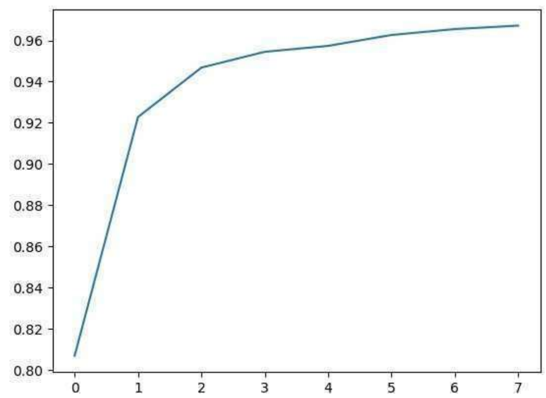

Fig 4: Accuracy Model Graph

The accuracy measure is the ratio of correctly anticipated events to all instances.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Where:

- True Positives, or TP,
- True Negatives is TN.
- FN is False Negatives.
- False Negatives, or FN

Precision:

The rate of directly prognosticated positive compliance to all anticipated cons is known as perfection.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall:

Recall is defined as the proportion of correctly predicted positive observations to all observations made in the actual class.

$$\text{Recall} = \frac{TP}{TP + FN}$$

F1 Score:

The harmonic mean of recall and accuracy is the F1 score:

$$\text{F1 Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

The Curve of Receiver Operating Characteristics(ROC):

The following formula is used to get the area under the ROC curve (AUC):

$$\text{AUC} = \int_0^1 TPR(FPR)\, d(FPR)$$

Where:

- The True Positive Rate is known as TPR (Recall)
- The False Positive Rate (FPR)

Activation Functions:

Rectified Linear Units, or ReLUs, are a popular activation function for each neuron in the CNN.

$F(x) = \max(0, x)$

Where:

An input to a neuron is x.

The output layer employs the sigmoid activation function for binary classification

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Where:

The neuron's weighted sum of inputs is represented by x.

Weight Update (Gradient Descent):

The weights are updated using the gradient descent algorithm:

$$W_{new} = W_{old} - \eta \cdot \nabla W$$

Where:

- W is the weight vector
- η is the learning rate
- ∇W is the gradient of the loss function concerning the weights

Convolutional Layer Operation:

In the convolutional layers, the convolution operation is defined as:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n)$$

Where:

- The input picture is I.
- I is the input picture, while K is the kernel or filter.
- S(i, j) is the output of the convolution at position (i, j).

C. Convolutional Neural Networks (CNNs):

Convolutional Neural Networks (CNNs) are advanced deep learning models tailored for analyzing and interpreting visual information. They perform exceptionally well on a variety of tasks, such as video analysis, pattern recognition, and picture categorization. CNNs can automatically recognize visual features, allowing them to interpret raw pixel data with minimal preparation. CNNs are designed with multiple layers and are modeled after the human visual cortex:

Convolutional Layers: These layers utilize filters to extract features such as edges, textures, and intricate patterns, resulting in feature maps that capture different details at various levels.

Activation Functions: Functions, like Rectified Linear Unit (ReLU), are applied following convolution to introduce non-linearity, enabling the network to recognize complex features and transmit important information to subsequent layers.

Pooling Layers: Pooling layers, commonly employing Max Pooling, downsample the feature maps to lessen computational demands and mitigate overfitting by preserving crucial information while simplifying the data.

Fully Connected Layers: Once feature extraction is complete, the data is flattened and processed through dense layers, whereby every neuron in the following layer is linked to every other neuron, finally resulting in the final categorization.

Output Layer: In cases of binary classification, such as pneumonia detection, a sigmoid activation function produces a probability score ranging from 0 to 1, reflecting the likelihood of each class.

Training Process: CNNs are trained using labeled image datasets, with the model adjusting filter weights based on prediction errors. By employing optimization methods like Stochastic Gradient Descent or Adam, the network minimizes the loss function, enhancing classification accuracy through backpropagation.

D. Naive Bayes:

Naive Bayes is a type of probabilistic classifier that relies on Bayes' theorem, operating under the assumption that all features are conditionally independent when the class label is known. Even with this simplified premise, Naive Bayes proves to be very effective in various real-world scenarios. The algorithm determines the class label by assessing the probability of each class based on the provided feature values and choosing the one that has the highest probability. Its computational efficiency makes Naive Bayes especially advantageous for datasets that contain a large number of features.

E. TensorFlow:

Google developed TensorFlow, a free and open- source machine literacy frame, with the thing of making the creation and training of deep literacy models easier. Its adaptable design makes it possible to produce and apply a variety of neural network models, which makes it applicable for a broad range of uses, similar to speech recognition, picture identification, and natural language appreciation. TensorFlow is a well-liked option for AI experimenters and inventors likewise because to its scalability and robust community support. It's extensively used in both academia and assiduity.

F. Keras:

Keras is an open-source API based on Python, tailored for the development of high-level neural networks. Keras works seamlessly with leading deep learning frameworks like TensorFlow, Theano, and Microsoft Cognitive Toolkit. Renowned for its simplicity and straightforward syntax, Keras empowers developers to create complex neural networks with little coding, making it a perfect choice for both beginners and experienced experts in machine learning and AI.

G. Decision Tree Classifier:

A decision tree classifier is a model that resembles a flowchart, where internal nodes denote decisions based on specific features, and leaf nodes represent the final class labels. This method recursively divides the data according to various features to form subsets that best distinguish between different classes. Splits are determined using measures such as entropy or Gini impurity to reduce uncertainty and enhance information gain. Decision trees are valued for their clarity and interpretability and can process both categorical and numerical data effectively.

H. Support Vector Machines (SVM):

Support Vector Machines (SVM) is an effective classification technique that identifies the most suitable hyperplane to distinguish between different classes in a feature space by maximizing the distance between the hyperplane acts as a separating boundary between different classes of data, with the closest points from each class being called support vectors. By using kernel functions to translate the data into a higher-dimensional space, SVM is able to handle both linear and nonlinear classification problems. This approach is especially useful for datasets that exhibit clear margin boundaries and is effective with high-dimensional data. Furthermore, SVM excels at handling outliers because of its emphasis on maximizing the margin. Choosing the best algorithm usually involves testing various options and evaluating their effectiveness based on the particular dataset and problem at hand.

I. Logistic regression:

Logistic Regression is an instance falling into a particular class, such as Normal or Pneumonia, based on its input features. Applying the logistic function converts the output values into probabilities ranging from 0 to 1, making it ideal for tasks involving two distinct categories.

J. Random classifier:

A Random Classifier, also known as a Random Baseline, is the simplest form of classification model in machine learning. It randomly assigns class labels to instances without any learning or pattern recognition. It serves as a baseline to compare the performance of more sophisticated models.

K. AdaBoost classifier:

In machine learning, AdaBoost (Adaptive Boosting) is a well-liked joint learning method. It builds a powerful classification system by combining a number of weak classifiers, such decision trees. It gives more weight to misclassified events during iteration, allowing the weaker classifiers to focus on correcting these errors. The final model combines the weighted predictions for the actual classification.

IV. RESULTS

S. No	Model	Accuracy
1	Convolutional Neural Networks	98.54%
2	TensorFlow	90%
3	Naive Bayes	94.5%
4	Decision Tree Classifier	95.8%
5	Support Vector Machines (SVM)	92%
6	Logistic regression	96.5%
7	Random classifier	92%
8	AdaBoost classifier	95%

Table 1 – Evaluation Table of the classification model on test data

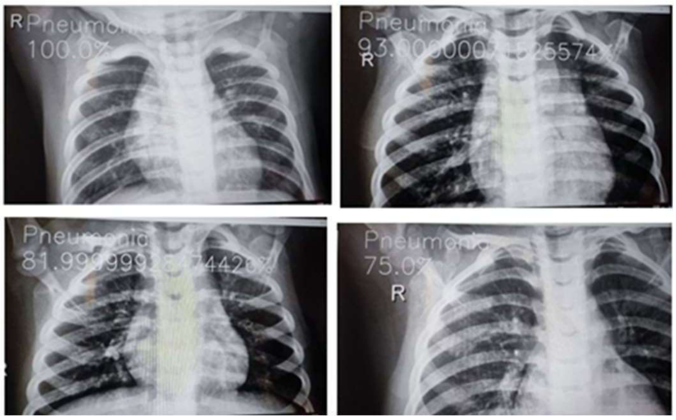

Fig 5: Model Predicts the Image Shows Pneumonia Disease

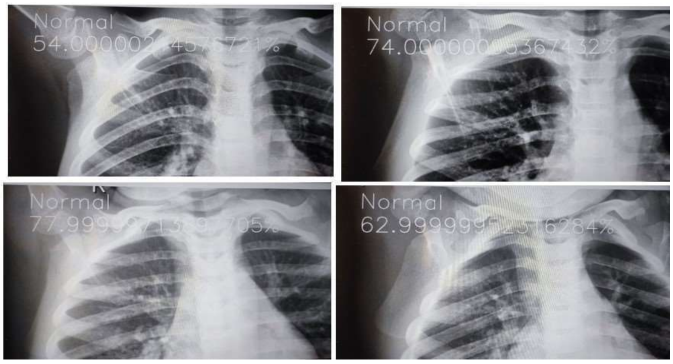

Fig 6: Model Predicts the Image Shows Normal

The CNN developed for detecting pneumonia from chest X-rays demonstrated a remarkable accuracy of 98.54%. This high accuracy highlights CNN's strong capability in effectively distinguishing between X-rays showing pneumonia and those considered normal. The model's impressive performance underscores the successful integration and application of advanced tools and technologies, including TensorFlow, Keras, and Python, which were pivotal in building this automated diagnostic system.

The CNN model was thoroughly evaluated using other measures in addition to accuracy to achieve this level of precision. In order to assess how successfully the model detected cases while lowering false positives and false negatives, precision and recall were used. The model's capacity to differentiate between classes at different thresholds was further evaluated in detail by looking at the area under the receiver operating characteristic curve (AUC-ROC).

In comparison to other classification methods, the CNN outperformed several algorithms in accuracy. For instance, TensorFlow, which was instrumental in developing the CNN, achieved an accuracy of 90%, while other classifiers like Naive Bayes and Decision Tree Classifier both reached 95.8%. Logistic Regression and Random Classifier showed slightly lower accuracy at 96.5%, and the AdaBoost Classifier also attained 95%. However, Support Vector Machines (SVM) significantly with an accuracy of only 92%, indicating it was less effective for this specific task

Overall, CNN's superior accuracy, coupled with its thorough performance evaluation using precision, recall, and AUC-ROC, validate its effectiveness and reliability as a diagnostic tool. This comprehensive assessment supports its potential for implementation in clinical settings, where accurate and efficient pneumonia detection is crucial.

V. CONCLUSION AND FUTURE WORK

In summary, this study highlights the potential of Convolutional Neural Networks (CNNs) for predicting pneumonia using Python, TensorFlow, and Keras. The developed model, trained on comprehensive chest X-ray datasets, demonstrated impressive accuracy in identifying pneumonia cases, suggesting its potential as a valuable tool for early disease detection. The implementation of CNNs effectively captures intricate patterns within medical images, making it a promising approach for enhancing diagnostic accuracy.

Moving forward, future efforts will focus on refining the model to ensure even greater precision and reliability. This involves the ongoing optimization of hyperparameters, expanding training datasets, and incorporating more sophisticated evaluation metrics to address various clinical scenarios. Collaboration with healthcare professionals is essential to validate the model's effectiveness in real-world settings and to adapt the system based on their insights.

Additionally, the model's performance must be assessed for its balance between sensitivity and specificity, ensuring that it minimizes false positives and negatives to reduce diagnostic errors. Ethical considerations, patient privacy, and adherence to regulatory standards will remain top priorities throughout this journey, ensuring that the model aligns with medical guidelines and practices.

Continuous advancements in machine learning and medical imaging will play a crucial role in evolving the model's capabilities. Therefore, regular monitoring, adaptation to new developments, and integration of emerging technologies will be necessary to enhance its pneumonia detection performance further. Ultimately, the successful application of CNNs in this domain has the potential to revolutionize pneumonia diagnosis, providing healthcare professionals with a robust, AI-driven tool that can significantly contribute to early intervention and improved patient outcomes. This approach signifies a meaningful step toward integrating artificial intelligence into healthcare, offering promising prospects for more accurate and timely pneumonia detection in the future.

REFERENCES

[1] Sharma, Natasha, and Priya. "Big Data Disease Prediction System Using Vanilla LSTM: A Deep Learning Breakthrough." Proceedings of Emerging Trends and Technologies on Intelligent Systems: ETTIS 2022. Singapore: Springer Nature Singapore, 2022. 167-176.

[2] Jakhar, Karan, and Nishtha Hooda. "Big data deep learning framework using Keras: A case study of pneumonia prediction." *2018 4th International Conference on computing communication and automation (ICCCA)*. IEEE, 2018.

[3] Singh, Utkarsh, Aditi Totla, and Prakash Kumar. "Deep Learning Model to Predict Pneumonia Disease based on Observed Patterns in Lung X-rays." *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 2020.

[4] Ji, Ruijun, et al. "Novel risk score to predict pneumonia after acute ischemic stroke." Stroke 44.5 (2013): 1303-1309.

[5] Cooper, Gregory F., et al. "An evaluation of machine-learning methods for predicting pneumonia mortality." Artificial intelligence in medicine 9.2 (1997): 107-138.

[6] Caruana, Rich, et al. "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission." Proceedings of the 21st ACM SIGKDD international conference on knowledge discovery and data mining. 2015.

[7] Özger, Hasan Selçuk, et al. "The factors predicting pneumonia in COVID-19 patients: preliminary results from a university hospital in Turkey." Turkish journal of medical sciences 50.8 (2020): 1810-1816.

[8] Zubair, S. H. A. H. "An efficient method to predict pneumonia from chest X-rays using deep learning approach." The Importance Of Health Informatics In Public Health During A Pandemic 272 (2020): 457.

[9] Kolditz, Martin, Santiago Ewig, and Gert Höffken. "Management-based risk prediction in community-acquired pneumonia by scores and biomarkers." European Respiratory Journal 41.4 (2013): 974-984.

[10] Metersky, Mark L., et al. "Predicting bacteremia in patients with community-acquired pneumonia." American journal of respiratory and critical care medicine 169.3 (2004): 342-347.

[11] Ning, Wanshan, et al. "Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning." Nature Biomedical Engineering 4.12 (2020): 1197-1207.

[12] Feng, Zhichao, et al. "Early prediction of disease progression in COVID-19 pneumonia patients with chest CT and clinical characteristics." Nature Communications 11.1 (2020): 4968

[13] Jakhar, Karan, and Nishtha Hooda. "Big data deep learning framework using Keras: A case study of pneumonia prediction." 2018 4th International Conference on computing communication and automation (ICCCA). IEEE, 2018.

[14] Huang, Jhih Siou, Yung Fu Chen, and Jiin Chyr Hsu. "Design of a clinical decision support model for predicting pneumonia readmission." 2014 International Symposium on Computer, Consumer and Control. IEEE, 2014.

[15] Swetha, K. R., et al. "Prediction of pneumonia using big data, deep learning, and machine learning techniques." 2021 6th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2021.

[16] Cohen, Joseph Paul, et al. "Predicting COVID-19 pneumonia severity on chest x-ray with deep learning." Cureus 12.7 (2020).

[17] Satici, Celal, et al. "Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19." International Journal of Infectious Diseases 98 (2020): 84-89.

[18] Ogawa, Jun, et al. "Prediction of and prophylaxis against Pneumocystis pneumonia in patients with connective tissue diseases undergoing medium-or high-dose corticosteroid therapy." Modern rheumatology 15.2 (2005): 91-96.

[19] Bodapati, Jyostna Devi, V. N. Rohith, and Venkatesulu Dondeti. "Ensemble of deep capsule neural networks: an application to pediatric pneumonia prediction." Physical and Engineering Sciences in Medicine 45.3 (2022): 949-959.

[19] Singal, Bonita M., Jerris R. Hedges, and Kenneth L. Radack. "Decision rules and clinical prediction of pneumonia: evaluation of low-yield criteria." Annals of Emergency Medicine 18.1 (1989): 13-20.

[20] Rajaraman, Sivaramakrishnan, et al. "Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs." Medical Imaging 2019: Computer-Aided Diagnosis. Vol. 10950SPIE,2019.