

PNEUMONIA DISEASE ASSESSMENT WITH FULL DIAGNOSTICS EVALUATION

Rayala Nithin¹, G Nagarajan², Thammishetti Venkat Sai Prathap³, Srikonda Pradeep⁴, Itharaju Sampreeth⁵

School of Computing, Department of Computer Science and Engineering, Kalasalingam Academy of Research and Education, Anand Nagar, Krishnankoil, Virudhnagar

{rayalanithin2000@gmail.com¹, nagarajanite@gmail.com², saiprataptammisetti@gmail.com³, pradeepsrikonda003@gmail.com⁴, sampreeth3279@gmail.com⁵}

Abstract— Pneumonia remains a critical global health issue, causing significant illness and mortality, especially in vulnerable populations. Early and precise diagnosis is essential for effective management and timely the intervention. This research paper introduces a novel integrated approach for the prediction and comprehensive diagnostics of pneumonia, harnessing cutting-edge advancements in medical imaging, machine learning, and clinical data analysis. The proposed framework merges three fundamental components: (1) Chest X-ray Image Analysis, (2) Clinical Data Integration, and (3) Machine Learning-Based Prediction. The chest X-ray image analysis employs state-of-the-art deep learning techniques to automatically identify and classify pneumonia-related anomalies, offering detailed insights into pulmonary conditions. Simultaneously, clinical data, encompassing patient history, demographics, and laboratory results, are integrated into the predictive model. A hybrid machine learning algorithm, combining Convolutional Neural Networks (CNN) and Gradient Boosting, is utilized to ensure robust and precise disease prediction. The research demonstrates the efficacy of this integrated approach through a thorough evaluation on a substantial dataset of pneumonia cases. Findings indicate a notable enhancement in pneumonia prediction accuracy compared to conventional diagnostic methods. Moreover, the model's interpretability is enhanced, allowing healthcare professionals to scrutinize the rationale behind each prediction, promoting trust and transparency in the decision-making process.

Keywords— Pneumonia, medical diagnostics, artificial intelligence, Convolutional Neural Networks (CNN), X-ray, Artificial Neural Networks (ANNs), Naïve Bayes, Support Vector Machine (SVM).

I. INTRODUCTION

Pneumonia, a potentially life-threatening respiratory infection, remains a prominent global health challenge. Timely and accurate diagnosis of pneumonia is critical for effective treatment and patient well-being. In this era of technological advancement, the integration of medical diagnostics, artificial intelligence, and telemedicine has paved the way for an innovative approach to pneumonia prediction and diagnosis.

The fundamental focus of this research attempt is to develop a comprehensive solution that not only detects pneumonia but also provides a platform for patients to consult with healthcare professionals. This multifaceted system includes features such as appointment scheduling, video consultations with doctors, and the convenient upload of essential medical documents. By combining these elements, we aim to streamline[14] the diagnostic process, promote

early intervention, and enhance the overall quality of patient care.

This research delves into the convergence of traditional diagnostic methods and modern technological advancements to create a holistic approach to pneumonia diagnosis. By allowing patients to interact with healthcare providers remotely and share critical medical information, we seek to improve healthcare accessibility, particularly in times of public health crises. This paper discusses the methodology, results, and implications of our comprehensive diagnostic platform, offering insights into the future of pneumonia prediction and healthcare delivery. As we navigate the intersection of medical science and digital innovation, this research aims to contribute to the ongoing dialogue on improving healthcare outcomes for pneumonia and other medical conditions.

II. LITRATURE REVIEW

Wunderink et al.,[1] proposed that the identification of abnormal chest X rays plays a role, in diagnosing ventilator associated pneumonia. The accuracy of X ray signs related to pneumonia has not been previously evaluated in anteroposterior X rays obtained from ventilated patients. To assess their ability for pneumonia, seven X ray signs were examined individually or in combination with clinical parameters.

Müller et al.,[2] stated this study is designed to assess how effective clinical signs and symptoms, as well as specific laboratory tests, are in accurately diagnosing and predicting the outcomes of individuals with CAP. The goal is to improve our understanding of how to identify and manage this condition more effectively. Community-acquired pneumonia is a common and serious infection that can lead to fatalities. It's the leading cause of death related to infections. To identify and diagnose CAP, doctors typically rely on certain criteria. These include the presence of a new infiltrate, which is an abnormal area on a chest X-ray, along with recent onset of respiratory symptoms.

Hashmi et al.,[3] This research addresses the global health burden of pneumonia, which claims many children's lives and affects a substantial portion of the population. It puts forward an effective model that recognize pneumonia in chest X-ray photographs, which can help radiologists make more accurate choices. A novel weighted classifier combines predictions from advanced deep learning models. Supervised learning and transfer learning are employed to enhance model accuracy. The study incorporates partial data augmentation to balance the training dataset. The proposed

weighted classifier achieves impressive results, with a 98.43% test accuracy and a 99.76 AUC score on unseen data, showcasing its potential for quick and accurate pneumonia diagnosis.

Moujahid et al.,[4] proposed that the context of diagnosing lung diseases, particularly during critical periods like the COVID-19 pandemic, the analysis and classification of X-ray images serve as vital initial steps in pneumonia diagnosis. With the growing number of cases, there is an increasing demand for highly accurate automated methods for lung disease classification. Convolutional Neural Networks (CNN) have gained widespread popularity due to their rapid processing and precision in image classification tasks. This article presents an approach utilizing CNN-based classification models incorporating transfer learning for diagnosing pneumonia. The research compares these models to determine the most effective one based on specific parameters, considering architectural, layer, and evaluation criteria. The literature review explores traditional and deep learning methods, assessing performance based on accuracy and loss functions, and conducts a critical analysis to identify areas for improvement.

Swetha et al.,[5] stated that the traditional pneumonia diagnosis relies on chest X-rays and expert interpretation. The pressing need for automated prediction systems, harnessing big data and deep learning, is evident. Convolutional Neural Networks (CNNs) have emerged as prominent players in this field, and pre-training them on extensive healthcare datasets holds the potential for precise classification. Combining a pre-trained CNN model with effective feature extraction techniques and diverse classifiers offers the prospect of achieving highly accurate results. This literature review delves into the prediction of pneumonia through the union of big data, dL, and ML techniques, providing valuable insights into the latest advancements in this crucial area of healthcare research.

Ning et al.,[6] by collecting vital components from chest X-ray visuals, Deep Learning Neural Network (DNN) images, issued with different transfer learning methods. This study involves some machine learning (ML) classifiers. The algorithm was trained and tested using a collection of X-rays of the chest and CT pictures. The system's efficiency and stability have been assessed employing several performance gauges. The overarching objective of the suggested procedure is to be pleasant and supportive.

According to Yi et al.,[7], an excellent radiologist needs to be experienced and be an expert in the area being dealt with so as to enable them effectively carry out analysis for chest X-ray images. For instance those human-assisted approaches that are currently in-use do not have diagnostic tools hence limiting their ability to handle complex diseases like pneumonia which requires specialized treatment hence expensive or unaffordable by many patients due absence of specialists for such diseases at every local health facility level. The study proposes an interpretable scalable DCNN that identifies pneumonia from chest X-ray pictures. To start with, the suggested updated DCNN model relies on the image's useful properties segregating it into normal or

pneumonia through classes. As the outcome, the advocated system underwent training and testing using a dataset of chest X-ray illustrations.

Saul et al.,[8] has always said that the flow of fluid in the lungs that leads to drowning, pneumonia calls for attention since it was a historical killer disease and still has grave implications within no time. If not treated with drugs at the right time, pneumonia can be fatal. Therefore, early diagnosis is important as far as the progression of this disease is concerned. This study provides an overview of past studies conducted in order to improve diagnostics levels; it also looks at the biological stage so pneumonia and how they are detected through x-ray imaging. Furthermore, this paper discusses the methodology and results of automating x-ray images using multi-parameters to detect early-stage diseases.

Jaiswal et al.,[9] argued that diagnostic imaging studies by researchers typically employ CXRs. Interpreting chest X-rays becomes more complicated due to varying appearances which depend on multiple factors such as patient position or inspiration depth. The implied identification model relies on a Mask-RCNN neural network, which synthesises global and local attributes for pixel-level segmentation. When tested against a dataset containing probable causes of pneumonia depicted in chest radiographs, the recommended identification model performed better.

Sourab et al.,[10] explained that the most frequent technique X-rays of the lung serve for the detection of pneumonia. However, detecting it is not easy or even accurate process done by radiologists especially where there are few experienced radiologists available in some countries. Preciseness of identification must be enhanced. Hence, this method is proposed to help radiologists and simplify the diagnostic complexity. In this recommended method, a 22-layer CNN model was constructed, and three different ML methods were applied to Retrieve and organize. learnt features of CNN model. SVM, K-Nearest Neighbor, and Random Forest Classifier. Several data augmentation approaches were used in order to introduce diversity into the existing dataset.

Kumar A et al., [11] examined the use of machine learning algorithms for forecasting pneumonia illness and making eventual medical consultation suggestions. The study reinforces the significance of precisely predicting pneumonia, especially in the setting of respiratory health, when early detection may drastically enhance patient outcomes. We used machine learning methods such as Naïve Bayes, Random Forest, Logistic Regression, and KNN to predict pneumonia based on patient symptoms. The study found the most exact approach for anticipating pneumonia through examining the reliability of several approaches.

Thakur R et al., [12] examine the use of computer-based technology in the healthcare sector, such as exploiting electronic medical databases for accurate analysis and illness prediction. The research focuses the need of using machine learning gets closer to early illness prediction,

patient care enhancement, and community health services. The study is divided into two sections, one on illness prediction and one on other topics. The first section extensively covers relevant work and gives details about the dataset used for analysis. The paper's second section concerns the set-up and efficiency assessment of machine learning algorithms for illnesses diagnosis.

Mehta A et al., [13] Address the grave problem of emotional well-being education and guidance, which rose to prominence during the COVID-19 epidemic. Considering the stigma tied to mental diseases, the article points out chatbots' ability as an unbiased and easily accessed resource for people facing mental health issues. The study is divided into two sections, which include a brief summary of existing research and an unconventional approach to mental health care using chatbots.

III. PROBLEM STATEMENT

Difficulties with pneumonia detection stem from inadequate access to healthcare, a late diagnosis, and possible errors in automated systems. Although dependable, traditional procedures frequently have delays that impede the efficacy of treatment. Barriers based on geography restrict access to specialized care, especially in isolated locations. Although promising, automated detection methods might not have the sophisticated knowledge of medical professionals, which could result in diagnostic mistakes. The therapeutic landscape for pneumonia is further complicated by the need to coordinate prescribed treatments and ensure medication adherence.

This research suggests an integrated healthcare approach as a way to address these issues. It acknowledges the necessity of prompt doctor visits made possible by online scheduling tools, guaranteeing patient accessibility. By overcoming regional barriers, video calling capabilities democratize access to healthcare. By fusing the advantages of technology and human knowledge, real-time medical diagnosis improves diagnostic precision. Medication is streamlined by integrating medical stores. Maximizes treatment adherence by streamlining the coordination of medication. In today's tech-infused healthcare environment, ethical considerations— Client confidentiality and the safety of data are vital.. To guarantee patient trust and wellbeing, a strong ethical framework and open communication are necessary. The portions of the study that follow will go into great length, offering suggestions and in-depth studies on how to redefine pneumonia detection using this combined strategy.

IV. PROPOSED SYSTEM

The proposed technique's main function is to diagnose pneumonia in its early stages using X-ray images and prevent further damage. Using the most recent DCNN architectures based on optimised versions, relevant descriptors are derived from the images as done in this paper. Various preprocessing techniques normalize the data before it is fed into classification models. Many machine learning models were employed in this study including KNN. In order to determine how well each utilised ML

model performs over time several performance assessment metrics are calculated. For development and training of our proposed system we use TensorFlow as backend of Keras deep learning framework. The work was executed using a range of libraries and packages such as TensorFlow, Keras, Sklearn, Matplotlib Pyplot Seaborn and Numpy libraries. We applied Jupyter Notebook with Anaconda integrated development environment (IDE) for all of the studies.

Our proposed system, which primarily focuses on pneumonia detection and has an intuitively integrated online consultation platform, is a ground-breaking step toward modernizing healthcare. We provide an advanced method that not only guarantees precise diagnosis but also revolutionizes the ensuing medical experience. Once pneumonia has been predicted, consumers will move smoothly into an online setting, combining the initial visit to the doctor, the diagnosis, and the link to an online pharmacy into a unified, seamless experience.

MODULE 1: DISEASE DETECTION

1.1 DATASET

The original dataset is divided into two subfolders with normal (N) and pneumonia (P) chest X-ray pictures, respectively, and three main folders. A total of 5,863 anterior-posterior chest X-ray pictures were selected with care from paediatric retrospective patients. Here are the images of chest X-rays with and without pneumonia.

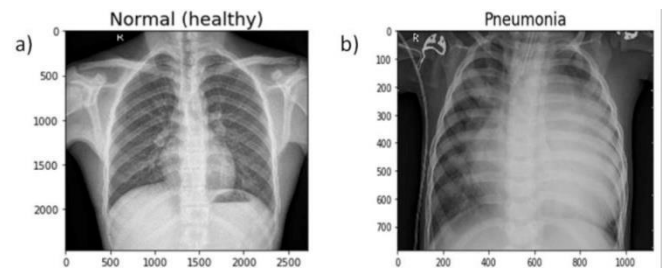


Fig-1: X-ray images of Normal Lungs and Pneumonia lungs

1.2 DATA PREPROCESSING

Data preparation is one of the most important techniques for providing well-formatted data to classification models— those that use normalized data for training and testing. Noise removal, contrast enhancement, high or low frequency filtering (and other methods) are applied to all images to improve their visuals before they are used by classifiers. This research has examined Min-Max normalization, and intensity normalization as pre-processing methods. Min-Max normal distribution, CLAHE, and intensity normalization provide intriguing preprocessing approaches for image processing applications. These because CLAHE was suitable for image processing applications as well as Min-Max normal distribution and intensity normalization were interesting pre-processing steps used here.

The dataset is unbalanced because, when comparing the two classes it represents—pneumonia and normal photos—nearly 75 percentage of the photographs show pneumonia, while the remaining 25 percentage show normal images. Several augmentation strategies have been employed to address the imbalanced dataset and overfitting issues, as well as to improve the accuracy of the models. Geometric transformations including rotations, zooms, rescales, shifts, flips, and shears are among the data augmentation techniques that are used.

1.3 CLASSIFICATION ALGORITHMS

1.3.1 BACK PROPAGATION NEURAL NETWORK

The effective technology known as BPNN, which is based on Artificial Neural Networks (ANNs), is used to detect intrusion activity. A neuron is the fundamental unit of the BPNN, processing and storing data. BPNN is a supervised algorithm that uses backpropagation to spread errors between the calculated and desired outputs. In order to reduce error by varying the weights through error propagation backward, the process is repeated. Weight changes cause hidden units to set their weights to correspond with significant task domain attributes.

There are three layers in BPNN: 1) Input Layer, 2) Hidden Layer, and 3) Output Layer. The problem complexity determines the hidden layers and the number of hidden units in each hidden layer.

1.3.2 NAÏVE BAYES CLASSIFIER

There is a fairly straightforward ML method called Naive Bayes. This is based on probability models and the Bayesian theorem. A Set of classifiers known as Naïve Bayes are a collection of classification methods for other Bayes-based classification methods. It is not one algorithm, but a family of algorithms that are characterized by the assumption that each pair of features to be classified are independent. Essentially, it means that every attribute's value given its class is conditionally independent of all others. This assumption is referred to as class-conditional independence in naïve Bayesian classifiers.

1.3.3 SUPPORT VECTOR MACHINE (SVM)

In both types of classification problems, Support Vector Machines (SVM) a supervised machine learning algorithm is applied. Usually applied to segregation problems. Every data point in this process is plotted as a dot in n-dimensional space which is the number of features you have. Each feature value represents a specific location. Simply put, support vectors are the coordinates of each single observation. In relation to this study, training data will mostly be built around support vector machines whereas testing data will largely contain decision values.

The steps involved in this method are as follows: Load Dataset; following this, the dataset will be classified into features (attributes) based on class labels; next, the Candidate Support Value will be estimated; for example, the

condition is $While (instances \neq null);$ if $Support Value = Similarity$ between each attribute instance, then the Total Error Value will be found. Assume that the projected decision value = $Support Value / Total Error$ for every occurrence < 0 . This process is repeated for all points until the result is empty. Consequently, we have primarily computed the gini index and entropy.

MODULE 2: VIRTUAL DOCTOR CONSULTATION

Let me introduce you to the virtual doctor's office, a place where medical experts and patients can communicate through technology. Users can participate in meaningful consultations using interactive interfaces and video chatting, guaranteeing individualized care catered to their individual health needs.

This module addresses the issue of delayed pneumonia diagnosis by integrating online scheduling tools to enable timely doctor appointments. Patients can see medical professionals more quickly by optimizing the appointment procedure, which guarantees prompt intervention and lowering the possibility of a symptom flare-up. This part highlights how crucial accessibility is to the early detection of pneumonia.

MODULE 3: MEDICAL VIDEO CALLING

Acknowledging geographic obstacles to healthcare accessibility, the suggested solution integrates video conferencing features. By enabling patients in rural or underserved locations to have remote consultations with medical specialists, this feature seeks to democratize healthcare.

Incorporating video calling into healthcare not only improves accessibility but also promotes a patient-centered approach by providing patients with expert medical views regardless of their location.

MODULE 4: REAL-TIME DIAGNOSIS

The combination of automated detection techniques with human expertise through real-time physician diagnosis forms the basis of the proposed system. This part improves diagnostic accuracy by verifying and honing the results of automated algorithms. By means of case studies, Using repeated feedback loops, the suggested method aims to reduce the possibility of mistakes and foster trust between patients and healthcare professionals.

MODULE 5: COORDINATION WITH MEDICAL STORES FOR TREATMENT OPTIMIZATION

The suggested solution incorporates medicinal stores into the healthcare continuum to expedite the treatment procedure. This maximizes treatment adherence by guaranteeing effective communication between pharmacies and healthcare professionals. By making quick access to recommended prescription drugs, the system seeks to improve overall patient outcomes and provide a more integrated and successful treatment strategy.

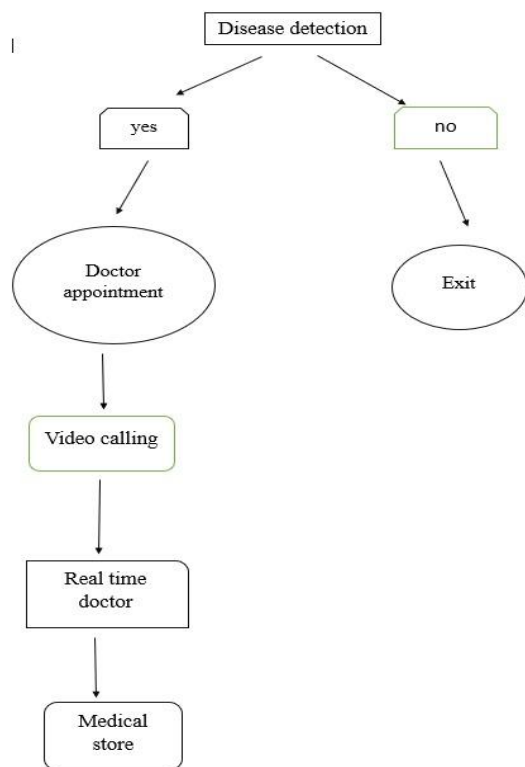


Fig-2: Flow Chart for automated decision modules.

V. CONCLUSION

Patients can schedule a visit with the physician. Establishing a virtual chat room with audio and video calling capabilities for patients and physicians. If the physician is able, they should physically consult the facility. We are including functions such as virtual counseling and tangible form for our undertaking. To get the most accurate picture of the patient's health, you must upload any prior medical documents that we may have to our website. By establishing a connection with the online diagnostic center, the doctor can examine prior patient records or provide new ones.

VI. REFERENCES

- [1] Wunderink RG, Woldenberg LS, Zeiss J, Day CM, Ciemins J, Lacher DA. The radiologic diagnosis of autopsyproven ventilator-associated pneumonia. *Chest*. 1992 Feb 1;101(2):458-63.
- [2] Müller B, Harbarth S, Stolz D, Bingisser R, Mueller C, Leuppi J, Nusbaumer C, Tamm M, Christ-Crain M. Diagnostic and prognostic accuracy of clinical and laboratory parameters in community-acquired pneumonia. *BMC infectious diseases*. 2007 Dec;7:1-0.
- [3] Hashmi MF, Katiyar S, Keskar AG, Bokde ND, Geem ZW. Efficient pneumonia detection in chest xray images using deep transfer learning. *Diagnostics*. 2020 Jun 19;10(6):417.
- [4] Moujahid H, Cherradi B, Gannour OE, Bahatti L, Terrada O, Hamida S. Convolutional neural network based classification of patients with pneumonia using X-ray lung images. *Advances in Science, Technology and Engineering Systems Journal*. 2020 Sep;5(5):167-75.
- [5] Swetha KR, Niranjanamurthy M, Amulya MP, Manu YM. Prediction of pneumonia using big data, deep learning and machine learning techniques. In 2021 6th International Conference on Communication and Electronics Systems (ICCES) 2021 Jul 8 (pp. 1697-1700). IEEE.
- [6] Ning W, Lei S, Yang J, Cao Y, Jiang P, Yang Q, Zhang J, Wang X, Chen F, Geng Z, Xiong L. Open resource of clinical data from patients with pneumonia for the prediction of COVID-19 outcomes via deep learning. *Nature biomedical engineering*. 2020 Dec;4(12):1197-207.
- [7] Yi R, Tang L, Tian Y, Liu J, Wu Z. Identification and classification of pneumonia disease using a deep learning-based intelligent computational framework. *Neural Computing and Applications*. 2023 Jul;35(20):14473-86.
- [8] Saul CJ, Urey DY, Taktakoglu CD. Early diagnosis of pneumonia with deep learning. *arXiv preprint arXiv:1904.00937*. 2019 Apr 1.
- [9] Jaiswal AK, Tiwari P, Kumar S, Gupta D, Khanna A, Rodrigues JJ. Identifying pneumonia in chest X-rays: A deep learning approach. *Measurement*. 2019 Oct 1;145:511-8.
- [10] Sourab SY, Kabir MA. A comparison of hybrid deep learning models for pneumonia diagnosis from chest radiograms. *Sensors International*. 2022 Jan 1;3:100167.
- [11] Kumar A, Sharma GK, Prakash UM. Disease prediction and doctor recommendation system using machine learning approaches. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*. 2021 Jul;9:34-44.
- [12] Thakur R, Khatri J, Virkar S, Mehta A, Dalvi A. Performance Measurement of different ML models for Disease Prediction.
- [13] Mehta A, Virkar S, Khatri J, Thakur R, Dalvi A. Artificial Intelligence Powered Chatbot for Mental Healthcare based on Sentiment Analysis. In 2022 5th International Conference on Advances in Science and Technology (ICAST) 2022 Dec 2 (pp. 185-189). IEEE.
- [14] Nagarajan, G., Singh, K., Poongodi, T., & Yadav, S. A. (2023). Adopting Streaming Analytics for Healthcare and Retail Domains. In *Machine Intelligence* (pp. 287-310). Auerbach Publications.

