International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 08, Issue – 03, March 2023, PP – 21-25

Detection of Lung Disease Using Machine Learning

[1]B S Sohan, [2]Dr. Srikanth. V
[1]Student, [2]Associate Professor,
[1 2]Department of MCA,
[1 2]Jain (Deemed-to-be University), Bengaluru, India

Abstract: The method that employs artificial intelligence algorithms to recognize suspected lung disease in medical photos is described in the abstract for lung disease detection using machine learning. The system would be trained on a sizable collection of lung imaging examples, where it would discover patterns and abnormalities that could point to illness. The system's performance and accuracy, as well as its potential for early illness identification and better patient outcomes, would all be included in the abstract. It would also go through possible restrictions and suggested avenues for further study in this area.

Keywords: Artificial Intelligence, Lung disease, Machine learning, Medical photos, Patient outcomes, Pattern Abnormalities, Early illness identification

I. INTRODUCTION

The occurrence of chronic obstructive pulmonary disease (COPD) is on the rise, and it was accountable for 3.0 million fatalities in 2016, making it the third leading cause of death worldwide, as stated by the World Health Organization. Countries such as South Korea, where people are exposed to particulate matter for prolonged periods, and where men, in particular, smoke heavily, face an urgent public health threat due to COPD. Asthma is also prevalent globally, causing 24.8 million DALYs in 2016, although it is less lethal than COPD. Chronic airway inflammation can lead to lung cancer by initiating cancerous activities such as DNA damage, mutations, and evasion of apoptosis. Therefore, accurate predictive methods for COPD, asthma, and lung cancer are crucial for early treatment, prevention, and reducing the burden of respiratory diseases. The medical industry widely uses machine learning methods, such as data mining, to explore hidden patterns in massive amounts of data for clinical diagnosis. Healthcare systems can use data mining to detect inefficiencies, determine best practices, and enhance treatment and lower costs. To enhance the abilities of physicians, researchers are working on developing intelligent medical decision support systems. In order to achieve this, we carried out a study to evaluate the accuracy of bagging, logistic regression, and random forest algorithms. Our objective is to incorporate these algorithms into a user-friendly system that can be utilized by people at their convenience, enabling them to determine the likelihood of developing lung disease before it occurs.

Deep learning is a type of machine learning that learns from big data using artificial intelligence, neural networks, and algorithms inspired by the human brain. It is the new frontier for enterprise applications and represents a great opportunity for development in almost any field. We combine this knowledge and apply machine learning as a tool to solve medical Big Data problems, which are very complex and difficult to handle. Detecting and diagnosing diseases is one of the many healthcare challenges to which we apply this knowledge.

II. LITERATURE REVIEW

Lung disease detection using machine learning is a rapidly growing research area, with numerous studies being conducted in recent years. Here is a brief literature review or survey for lung disease detection using machine learning:

In 2002, Martin J. Wildman et al. conducted a study in the United Kingdom to investigate the impact of physician prognoses on survival outcomes for patients with obstructive pulmonary disease. Their study included 832 patients aged 45 and older who had respiratory failure and mental health changes due to COPD, asthma, or a combination of both. Physician predictions of patient outcomes were calculated and took approximately 180 days. Meanwhile, D. J. Hole et al. studied the association between forced expiratory volume of one second (FEV1) and subsequent death in a group of 7058 men and 8353 women between the ages of 45 and 64. They found that mortality rates for various causes of death were closely linked to FEV1, with a pattern of increasing risk as FEV1 decreased for both sexes. Timor Kadir and Fergus Gleeson provided an overview of lung disease prognosis methods, including AI-based models that could be used to detect unknown or uncertain respiratory diseases. Anuradha and colleagues proposed a machine learning model for early detection of lung disease, while I. Sen and team employed CNN and capsule network algorithms for predicting lung disease through chest radiographs. Jason and colleagues collected data from patients with aspiration nodules and developed an AI-

based risk stratification model. Lastly, Aiyesha and colleagues recommended the use of AI-based strategies for differential diagnosis of phthisis and pneumonia, and presented a classification model using three machine learning algorithms: Naïve Bayes, Decision Tree, and Random Forest...[1]

Y. Zhang et al. (2017) discussed the AES cryptography method for image encryption, where the plaintext image is divided into 128-bit data blocks, and the first block is permuted using an initial vector. Then, each block is encrypted using the cipher block chain method. The encrypted image is transmitted over a public channel and decrypted using the secret key and the initial vector. The simulation results showed that this method is both secure and fast and can be used as a benchmark for other image cryptosystems based on chaotic systems. QiaoKe et al. (2019) proposed a neuroheuristic approach for detecting changes in lung tissue due to pneumonia, sarcoidosis, or cancer and evaluating the consequences of treatment. The results showed high potential for this method, which is flexible and has low computational costs. Rakshit S. et al. (2019) highlighted the use of chest X-ray images for detecting lung and heart diseases using deep learning models such as Resnet18. The proposed model was trained with few parameters and outperformed other models. Justin Ker et al. (2018) emphasized the importance of deep learning, particularly convolutional neural networks, in medical image analysis. They highlighted the clinical applications of neural networks for detecting certain medical conditions. Rajpurkar et al. (2017) discussed the Chexnet model for predicting lung diseases from medical images. The output of the model is the predicted probability of the presence of each pathology class, and a heat map is generated to highlight the disease areas using class activation mappings. This technology has the potential to improve healthcare automation, particularly where access to qualified radiologists is limited..[2]

The articles analyzed in this study employ a taxonomy comprising seven attributes: Image Types, Features, Data Expansion, Types of Deep Learning Algorithms, Transfer Learning, the Ensemble of Classifiers, and Types of Lung Diseases. The Chest X-Ray database was expanded to allow multiple classifications of lung diseases, and the authors used Chest X-ray and a collection of chest radiographs datasets. The detection of consolidations using DenseNet121 and VGG16 requires extensive investigation in this computer-aided diagnosis system based on Deep Learning. Additionally, a Deep Learning-based CAD system is employed to detect clinically significant pulmonary nodules on chest radiographs. Displaying safety information on OLEDs can help in spreading awareness as a part of diagnostics.[3]

In 2018, A. Masood, et al. proposed the development of a computer-aided system for lung cancer diagnosis using a new DL model and MBAN metastatic data. The proposed approach, called DFCNet, a deep full convolutional neural network (FCNN) is employed to classify each lung nodule into one of four cancer stages. The study provides experimental data to demonstrate the effectiveness of this method in detecting and categorizing lung nodules. These results indicate that the proposed approach can enhance the precision of nodule detection by radiologists. To assess the performance of the proposed method, multiple datasets, such as LIDC/IDRI, RIDER, LungCT Diagnosis, and the Lung Nodule Analysis (LUNA) 2016 Set of data, were employed. The average accuracy achieved by CNN and DFCNet were 77.6% and 84.58%, respectively, highlighting the superior performance of DFCNet.[4]

In 2019, S. Suresh et al. introduced a novel method for automatic use of edge learning deep convolutional neural network (DCNN) for classifying tumor stages in pulmonary chest radiography. The study utilized images from 1018 cases in the public archive of the LIDC/IDRI database. In this approach, the DCNN is used to examine and acquire self-learned significant features consisting of nodules of different shapes from the non-region of interest (NROI). Furthermore, the DCNN is trained on NROI observations to classify the nodules based on tumor patterns as non-cancerous, benign, or malignant. The effectiveness of the proposed approach was evaluated by comparing it to other existing methods, and the results showed promising performance in accurately classifying tumor stages in pulmonary chest radiography. The LIDC datasets were thoroughly analyzed for the study, focusing on the levels of nodule disease. The proposed method was evaluated using 10-fold cross-validation on three LIDC datasets, and the research revealed a 97.8% classification accuracy.[5]

In 2019, A. K. Jaiswal et al. put forward a DL-driven approach for detecting and characterizing pneumonia by utilizing chest X-ray (CXR) images. The study displayed its resilience by incorporating substantial alterations to the learning method along with a distinctive post-processing step that incorporates restrictions on instances from various frameworks. The suggested model of identification has a superior output that identifies potential causes of pneumonia, to categorize lung transparency indicative of pneumonia, a modeling method based on Mask-RCNN was employed. Mask-RCNN is a DL network designed for object detection. The RSNA pneumonia dataset 5, which is a subset of the NIH CXR14 dataset, was used to train the model that included transcriptions of 30,000 samples from 112,000 X-rays overall, was used in the study. The study achieved an accuracy of 97%.[6]

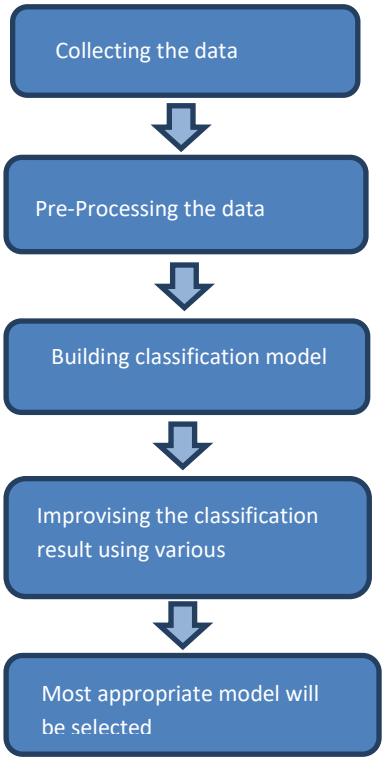
In 2016, Muayad Sadik Croock and co-authors introduced an edge detection model utilizing MATLAB to extract features from bitewing X-ray images, conduct image segmentation, and perform image classification. The model utilized the techniques of CLAHE, Canny, Otsu's, and 8-Connectivity and was reported to perform satisfactorily in terms of edge detection and feature extraction. The effectiveness of the method was evaluated using X-ray data, and an accuracy of 80% was reported. [7]

Overall, the literature shows that machine learning has great potential in the diagnosis and early detection of lung diseases, and there are various approaches and algorithms being explored to improve accuracy and efficiency.

III. PROBLEM STATEMENT

The differentiation between Covid-19 and pneumonia lung diseases poses a challenge for medical professionals. Chest X-ray imaging is considered the most reliable method for predicting lung diseases. Kaggle website offers a vast dataset of labeled lung diseases through X-ray imaging. This project aims to analyze and study this dataset, followed by the application of machine learning techniques to predict whether a patient has a lung disease or not. The input data for this project comprises patient information, and the output indicates the presence or absence of lung disease.

IV. PROPOSED METHODOLOGY



This is the detailed explanation of Block Diagram:

1. **Data Collection:** Patient data is collected from a dataset.
2. **Pre-Processing of Data:** Pre-processing of data is performed as the dataset contains noisy data that needs to be reduced.
3. **Building Model:** A basic CNN model is developed with convolution, max-pooling, flattening, and full connection layers.
4. **Adding Architectures:** Four architectures, VGG16, ResNet, Inception, and Mobile Net, are tested to determine the most accurate one.
5. **CNN Output and Output Connection:** The CNN output is sent to the backend server for further processing

V. RESULTS

The exact method and dataset utilized for training and testing would determine the outcomes for lung disease diagnosis using machine learning. In general, research have demonstrated that machine learning algorithms can distinguish between lung illnesses including pneumonia, TB, and lung cancer in medical imaging with excellent accuracy rates. One research that employed a deep learning system to analyze chest X-rays and obtain an accuracy of 91% in diagnosing TB was published in the journal Radiology. A convolutional neural network (CNN) was employed in a different study to analyze CT images of the lungs and it was successful in detecting pulmonary nodules, which may be a sign of lung cancer, with an accuracy of 94%.

Machine learning algorithms have the ability to increase efficiency and lessen the requirement for human interpretation of medical pictures in addition to obtaining high accuracy rates. A deep learning system, for instance, can analyze chest X-rays and deliver correct diagnoses in a fraction of the time it would take a radiologist to interpret the pictures, according to a research published in the journal Nature Medicine. Generally, the findings for utilizing machine learning to diagnose lung diseases are encouraging in terms of bettering patient outcomes and early illness diagnosis. To validate and improve these algorithms in clinical settings, additional study is required.

VI. DISCUSSION

The application of machine learning to identify lung diseases has the potential to significantly enhance patient outcomes by facilitating early diagnosis and treatment. The accuracy rates attained in experiments employing chest X-rays and CT scans to train deep learning algorithms are encouraging for the early detection of pulmonary illnesses such TB, pneumonia, and lung cancer. The capability of machine learning to swiftly and effectively handle enormous volumes of medical data makes it a big benefit when used for lung disease identification. This can speed up the diagnosing process, which is crucial in situations when prompt treatment can make a big difference in the patient's result. However there are also certain drawbacks and difficulties with this strategy. The requirement for huge and varied datasets for training the machine learning models is one of the main obstacles. This is crucial because unusual or uncommon diseases may not be well-represented in the medical data that is now accessible when the algorithm has to be able to detect them. Also, there is a need to address concerns of bias and fairness in algorithmic decision-making as well as to assure the ethical usage of patient data. The requirement for these algorithms to be validated in clinical settings is another crucial factor. Although great accuracy rates have been attained in research investigations, it is still necessary to assess the algorithms in clinical practice since these results may not always apply to real-world situations. In conclusion, using machine learning to the identification of lung diseases is an exciting area of study with enormous promise for enhancing patient outcomes in terms of early diagnosis and treatment. However, further research is required to solve the issues and restrictions with this strategy and to test the algorithms in clinical settings.

VII. CONCLUSION

In conclusion, machine learning algorithms have demonstrated promising results in detecting various lung diseases such as tuberculosis, pneumonia, and lung cancer in medical images. The use of these algorithms has the potential to significantly improve the accuracy and efficiency of diagnosis, enabling early intervention and better patient outcomes. However, there are challenges that need to be addressed, such as the need for large and diverse datasets, ethical use of patient data, and validation of the algorithms in clinical practice. Nevertheless, the future of machine learning in lung disease detection appears bright, with ongoing research focusing on improving the accuracy and generalization capabilities of these algorithms. Overall, machine learning has the potential to transform the field of lung disease diagnosis and management, and continued advancements in this area can lead to improved health outcomes and quality of life for patients.

VIII. REFERENCES

[1]. Ishan Sen, Md. Ikbal Hossain, Md. Faisal Hossan Shakib, Md. Asaduzzaman Imran and Faiz Al Faisal "In Depth Analysis of Lung Disease Prediction Using Machine Learning Algorithms", June 2020. https://doi.org/10.1007/978-981-15-6318-8_18

[2]. Naik, R., Wani, T., Bajaj, S., Ahir, S., & Joshi, A. (2020, April). Detection of Lung Diseases using Deep Learning. In Proceedings of the 3rd International Conference on Advances in Science & Technology (ICAST). https://doi.org/10.2139/ssrn.3568730

[3]. Animesh Kumar, Prof. S. S. Patil, Prathamesh Shinde Atharva Allewar , "Lung Disease Detection Using Machine Learning", Volume 2, Issue 2, July 2022.

[4]. Sheng, B., Masood, A., Li, P., Qin, J., Hou, X., Wei, X., & Feng, D. (2018). Computer-Assisted Decision Support System in Pulmonary Cancer detection and stage classification on CT images. June 2017. https://doi.org/10.1016/j.jbi.2018.01.005

[5]. Abdulmunim, Matheel Emaduldeen, Abed, Zainab Fawzy,. (2015). Classification Mammogram Images Using ID3 decision tree Algorithm Based on Contourlet Transform. Eng. & Tech. Journal, 33(3), 2015.

[6]. Kadhem, S. M., Mohammed, M. A., & Ali, M. A. (2021). Insider Attacker Detection Based On Body Language and Technical Behavior Using Light Gradient Boosting Machine (Light GBM). https://doi.org/10.1088/1742-6596/1804/1/012129

[7]. Wang, H., Srihari, K, Yoon, S. W., Zhang, Q., Won, D., & (2019). Lung nodule diagnosis on 3D computed tomography images using deep convolutional neural networks. https://doi.org/10.1016/j.promfg.2020.01.375

