International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

AN AUTOMATED FRAMEWORK FOR DIAGNOSING LUNGS RELATED ISSUES USING ML AND DATA ANALYTICS

Mohammed Maaz Ahmed Khan[1], Mohammed Siddiq S[2], Mythri J L[3], Naveen A[4], Dr. Rajesh T M[5]

[1,2,3,4,5] *Department of Computer Science and Engineering, Dayananda Sagar University, Bangalore, Karnataka, India*

Abstract - *Compared to most other tissues, lungs are directly exposed to oxygen concentrations. Lung diseases are one of the leading causes of death. There are many different lung diseases, some of which are caused by viral, bacterial, or fungal infections. Other lung diseases are associated with environmental factors, including COVID19, tuberculosis, bronchitis, pneumonia, etc. Deep learning has shown great potential when applied to medical images for disease detection including lung disease. We build and compare two pre-trained models, MobileNet and VGG16 architectures using the Transfer learning approach. We have also used Supervised Machine Learning algorithms like Random forest, Decision Trees, Support Vector Machines, and Logistic Regression. This paper provides the analysis which we have performed on the different algorithms.*

Key Words: **Fungal infections, Pneumonia, MobileNet, VGG16, Transfer Learning, Supervised Machine Learning algorithms**

1. INTRODUCTION

Lung disorders often called respiratory diseases, are illnesses that affect the lungs' airways and other components. Pneumonia, TB, and Coronavirus Disease are all examples of lung diseases (COVID-19). According to the Forum of International Respiratory Societies, around 334 million people have asthma, tuberculosis kills 1.4 million people each year, lung cancer kills 1.6 million people, and pneumonia kills millions. COVID-19 was a global pandemic that infected millions of individuals and put a strain on healthcare services. Lung illnesses are without a doubt one of the world's leading causes of death and disability. Early identification is crucial for boosting long-term survival rates and enhancing the possibilities of recovery. Lung disease is usually discovered by a physical exam & skin tests, blood tests, sputum sample tests, chest X-ray exams, and computed tomography (CT) scan exams are all used to identify cancer. Deep learning has recently shown considerable promise in disease identification using medical pictures, particularly lung disease.

The threat of lung illnesses is enormous, in particular in growing and low-middle-income countries, wherein tens of thousands and thousands of humans are dealing with poverty and air pollutants. According to the estimation of WHO, over four million untimely deaths arise yearly from household air pollutants-associated illnesses, together with asthma, and pneumonia. Hence, it's far more important to take important steps to lessen air pollutants and carbon emissions. It is likewise vital to enforce green diagnostic structures that can help in detecting lung illnesses. Since December 2019, a singular coronavirus sickness 2019 (COVID-19) has been inflicting critical lung harm and respiration problems. In addition, pneumonia, a form of lung sickness, may be because of the causative virus of COVID-19 or can be because of different viral or bacterial infections. Hence, early detection of lung illnesses has turned out to be more essential than ever. Recently, the virtual era has turned out to be a greater essential worldwide. This challenge can offer medical doctors and different researchers a course for detecting lung sickness with the assistance of a deep studying methodology. A huge quantity of lung X-ray pics are used as a dataset

1.1 Problem Definition

Machine Learning performs an essential position in clinical systems. Lung illnesses are one of the main reasons for death. The early identity and prediction of lung illnesses have emerged as a need withinside the research, as it may facilitate the subsequent scientific control of patients. Machine Learning-primarily based totally selection help systems offers the contribution to the medical doctors of their analysis decisions. The mission taken into consideration is the class of lung illnesses like Pneumonia, Tuberculosis, Lung cancer, and Covid 19. Machine Learning and Deep Learning are used to manage information in addition to creating fashions for diagnosing patients. Combining the processing of affected person statistics with information from chest X-rays and CT scans, using CNN with the famous pre-skilled model, These Neural networks for information of this shape are the techniques used for this mission to pick out lung illnesses.

2. RELATED WORKS

A fully CNN has been proposed in Ref. [6] to reduce the false-positive rate in classifying the lung nodules. This method can only analyze the nature of the CT scan images in order to reduce the probability of a wrong diagnosis.

In Ref. [7], a framework for deep learning is proposed to predict lung cancer and pneumonia offering two deep learning methods. Initially, they used modified AlexNet for the diagnosis of chest X-rays. Moreover, in the modified AlexNet, SVM is implemented for the purpose of classification.

Deep learning methods are also proposed in Ref. [8] where several transfer learning methods such as DenseNet121, AlexNet, Inception V3, etc., are used for pneumonia diagnosis. However, the parameter tuning for their implemented methods is very complex.

3. REQUIREMENTS

The Software requirements used for performing the experiments are as follows, Transfer learning, React framework, Tensorflow, Python3, OpenCV, Keras, Numpy, Matplotlib, Scikit-learn, and Tkinter

In order to successfully run the project the following hardware requirements are needed CPU Intel i3+, RAM 2GB+, Storage Space At least 2GB, Mobile phone (Android & iPhone)

3.1 System Features

Constructing dataset for Tuberculosis, pneumonia, lung cancer, and COVID 19, Model portable on IOS and Android platform, Recognition of lung-related diseases

4. METHODOLOGY

Fig -1: Proposed Design

4.1 Image Acquisition Phase

The first step is to acquire images. To produce a classification model, the computer needs to learn by example. The computer needs to view many images to recognize an object. Other types of data, such as time-series data and voice data, can also be used to train deep learning models. In the context of the work used in this project, the relevant data required to detect lung disease will be images. Images that could be used include chest X-ray, and CT scan. The output of this step is images that will later be used to train the model.

4.2 Preprocessing Phase

The second step is preprocessing. Here, the image could be enhanced or modified to improve image quality. Image modification such as lung segmentation and bone elimination could be used to identify the region of interest (ROI), whereby the detection of the lung disease can then be performed on the ROI. Edge detection could also be used to provide an alternate data representation. Data augmentation could be applied to the images to increase the amount of available data. Feature extraction could also be conducted so that the deep learning model could identify important features to identify a certain object or class. The output of this step is a set of images whereby the quality of the images is enhanced, or unwanted objects have been removed. The output of this step is images that were enhanced or modified that will later be used in training.

4.3 Training Phase

In the third step, namely training, three aspects could be considered. These aspects are the selection of deep learning algorithms, usage of transfer learning, and usage of an ensemble. There are numerous deep learning algorithms, for example, multilayer perceptron neural network (MPNN), recurrent neural network (RNN), and the aforementioned CNN. Different algorithms have different learning styles. CNN works particularly well with images. A deep learning algorithm should be chosen based on the nature of the data at hand. Transfer learning refers to the transfer of knowledge from one model to another. Ensemble refers to the usage of more than one model during classification. Transfer learning and ensemble are techniques used to reduce training time, improve classification accuracy and reduce overfitting. The output of this step is models generated from the data learned.

4.4 Classification Phase

In the fourth and final step, which is classification, the trained model will predict which class an image belongs to. For example, if a model was trained to differentiate X-ray images of healthy lungs and tuberculosis-infected lungs, it should be able to correctly classify new images (images that are never seen by the model before) into healthy lungs or tuberculosis-infected lungs. The model will give a probability score for the image. The probability score represents how likely an image belongs to a certain class. At the end of this step, the image will be classified based on the probability score given to it by model

5. TESTS AND RESULTS

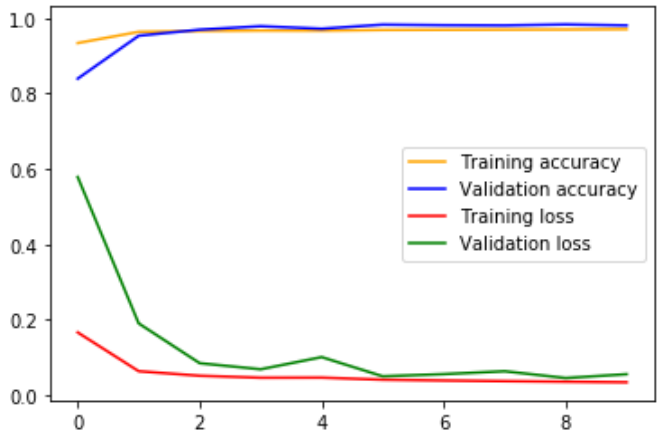

Chart -1: Summary of training evaluation & loss of Image Segmentation

Table -1: Summary of Dataset

Summary of Dataset		
Category	**Training dataset**	**Testing Dataset**
Pneumonia	6119	2623
Normal	4965	2128

Table -1: Summary of Supervised and Transfer learning experimentation results

Summary of Supervised and Transfer learning experimentation results		
Algorithm	Training Accuracy	Testing Accuracy
Random forest	86.3%	84.1%
Logistic Regression	80.2%	75.1%
Decision Trees	80.1%	79.1%
SVM	79.22%	67.88%
Mobile Net	99.83%	99.83%
VGG16	99%	98.24%

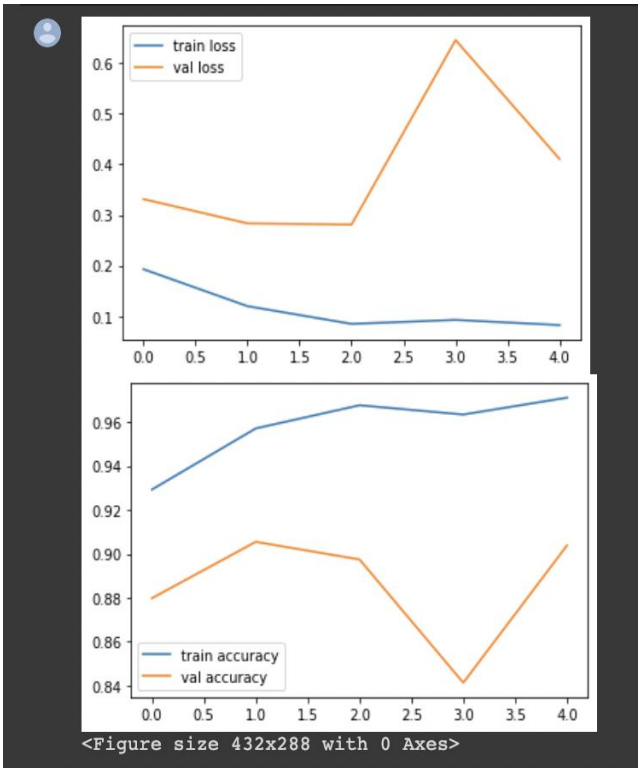

Chart -2: Summary of training evaluation & loss of VGG Classifier

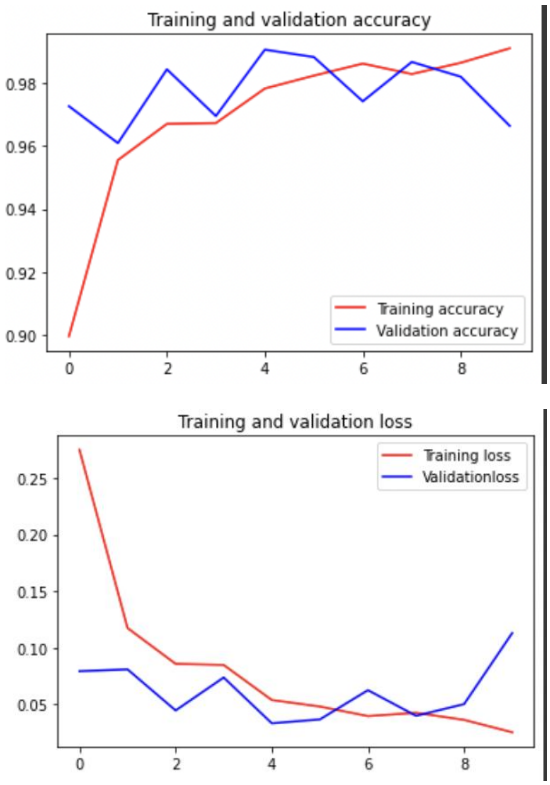

Chart -3: Summary of training evaluation & loss of MobileNet

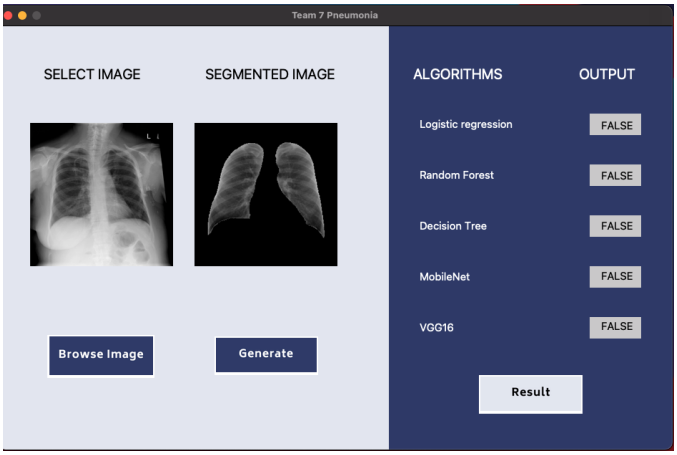

Fig -2: Integrated Application displaying negative for Pneumonia Disease

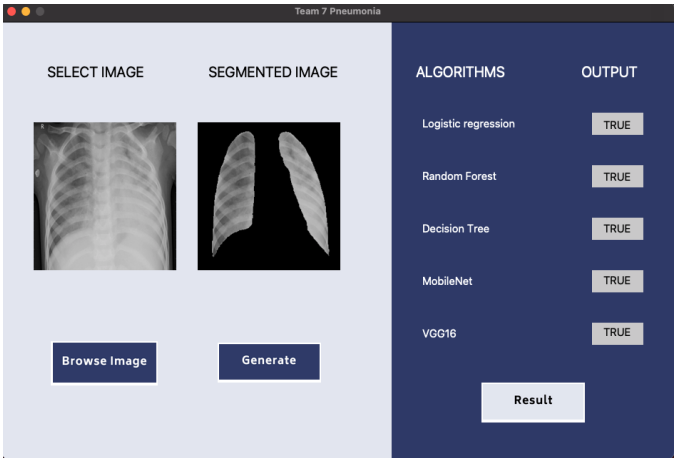

Fig -3: Integrated Application displaying Positive for Pneumonia Disease

6. DELIVERABLES

The final product is a fully integrated desktop application that can classify various lung-based diseases through X-Rays and CT-Scans, these include diseases such as Pneumonia, Lung cancer, Covid-19, and Tuberculosis. This product is designed in such a way that any person with basic knowledge can upload their X-Ray or CT-Scans to get a detailed analysis.

The Application will be used not only as a second opinion but also as a way of initial screening and Cross Verification among radiologists, as detection of diseases from a Chest X-ray and CT-Scan is complex and may go unnoticed.

This is a real-time application that uses machine learning and artificial intelligence (AI) based algorithms to assess the condition of the lungs, return the health of the lungs and the percentage of infection (if affected by the diseases mentioned above). The final application is projected to be implemented once the development and initial testing are done.

This is a project under the collaboration of Dayananda Sagar University, CDSIMER, and Sagar Group of Hospitals.

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and the support of many individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to the School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor's degree in this institution.

We would like to thank Dr. A Srinivas. Dean, School of Engineering & Technology, Dayananda Sagar University for his constant encouragement and expert advice. It is a matter of immense pleasure to express our sincere thanks to Dr. Girisha G S, Department Chairman, Computer Science, and Engineering, Dayananda Sagar University, for providing the right academic guidance that made our task possible.

We would like to thank our guide Dr. Rajesh T M, Associate Professor, Dept. of Computer Science and Engineering, Dayananda Sagar University, for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and the fruitful culmination of the project.

We would like to thank our Project Coordinator Dr. Meenakshi Malhotra and all the staff members of Computer Science and Engineering for their support.

We are also grateful to our family and friends who provided us with every requirement throughout the course. We would like to thank one and all who directly or indirectly helped us in the Project work.

REFERENCES

[1] [1] Qin, Z. Z., Ahmed, S., Sarker, M. S., Paul, K., Adel, A. S. S., Naheyan, T., ... Creswell, J. (2021). Tuberculosis detection from chest x-rays for triaging in a high tuberculosis-burden setting: an evaluation of five artificial intelligence algorithms. The Lancet Digital Health, 3(9), e543–e554. doi:10.1016/s2589-7500(21)00116-3

[2] [2] Pasa, F., Golkov, V., Pfeiffer, F. et al. Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. Sci Rep 9, 6268 (2019). https://doi.org/10.1038/s41598-019-42557-4

[3] [3] Yoo, Seung Hoon; Geng, Hui; Chiu, Tin Lok; Yu, Siu Ki; Cho, Dae Chul; Heo, Jin; Choi, Min Sung; Choi, Il Hyun; Cung Van, Cong; Nhung, Nguen Viet; Min, Byung Jun; Lee, Ho (2020). Deep Learning-Based Decision-Tree Classifier for COVID-19 Diagnosis From Chest X-ray Imaging. Frontiers in Medicine, 7(), 427–. doi:10.3389/fmed.2020.00427

[4] [4] Chandra T.B., Verma K. (2020) Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In: Chaudhuri B., Nakagawa M., Khanna P., Kumar S. (eds) Proceedings of 3rd International Conference on Computer Vision and Image Processing. Advances in Intelligent Systems and Computing, vol 1022. Springer, Singapore. https://doi.org/10.1007/978-981-32-9088-4_3

[5] [5] van Cleeff, M., Kivihya-Ndugga, L., Meme, H. et al. The role and performance of chest X-ray for the diagnosis of tuberculosis: A cost-effectiveness analysis in Nairobi, Kenya. BMC Infect Dis 5, 111 (2005). https://doi.org/10.1186/1471-2334-5-111

[6] [6] Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal 2017;42: 1–13.

[7] [7] Bhandary Abhir, et al. Deep-learning framework to detect lung abnormality – a study with chest X-Ray and lung CT scan images. Pattern Recogn Lett January 2020;129:271–8 https://doi.org/10.1016/j.patrec.2019.11.013.

[8] [8] Chouhan V, et al. A novel transfer learning-based approach for pneumonia detection in chest X-ray images. Appl Sci 2020;10(2):559. https://doi.org/10.3390/app10020559.

BIOGRAPHIES

Mr. Mohammed Maaz Ahmed Khan is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in his final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Mohammed Siddiq S is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in his final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Miss. Mythri J L is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. She is currently in her final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Mr. Naveen A is pursuing a Bachelor of Technology degree in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India. He is currently in his final year of Engineering and will be graduating from Dayananda Sagar University in the year 2022.

Dr. Rajesh T M is an Associate Professor in the Department of Computer Science and Engineering from Dayananda Sagar University, located in Bangalore, Karnataka, India.