Lungs Disease Prediction Using Deep Learning

Haritha A. H[1], Mrs. Dhivya V.M[2]
[1]*Department of Computer Science and Engineering, Sivaji college of Engineering and Technology, Parasuvaikkal, Tamil Nadu 695504, India*
[2]*Assistant Professor, Department of Computer science and Engineering, Sivaji college of Engineering and Technology, Parasuvaikkal, Tamil Nadu 695504, India*
doi.org/10.64643/IJIRTV12I11-201420-459

Abstract*—lung disease is one of the diseases that can be cured when the disease is spotted in its early stages before getting accumulated. But the most people fail to detect their disease before it comes to chronic. It leads to an increase in the death toll all around the world. Chest X-Ray is one of the most frequently used diagnostic modality in detecting different lung diseases such as pneumonia or tuberculosis. Also, deep learning techniques perform well in medical image classification tasks. This paper proposes a lung disease classification using lung x- ray image in 4 classes- pneumonia, tuberculosis, covid-19 and normal. The dataset is available in the public repository and can be freely download. The classification task is based on the transfer learning concept, it uses pre-trained algorithms for classification. In this implementation, make use of Resnet-50 algorithm for classification. It is one of the deeper neural networks. It performs well in many classification tasks. The algorithm extracts the features present in the input image in the dataset and use these features for learning, thereby, make predictions on new input X-ray images.

Index Terms*—Deep Neural Networks, Resnet-50, Transfer Learning.

I. INTRODUCTION

The novel coronavirus disease (COVID-19) outbreak is caused by a strain of coronavirus known as the severe acute respiratory syndrome coronavirus 2 that originated in Wuhan in the Hubei province in China at the end of 2019. The World Health Organization declared COVID-19 as a pandemic on March 11, 2020, then it had spread across the world. The website of the World Health Organization has listed the total number of reported patients with COVID-19 and the associated deaths.

COVID-19 is diagnosed using real-time polymerase chain reaction (RT-PCR) in many clinical situations. However, RT-PCR sensitivity is not very high in the detection of COVID-19; for example, one study reported that the sensitivity of RT-PCR (71%) was lower than that of chest computed tomography (98%). Owing to the low RT-PCR sensitivity, the effectiveness of chest X-Ray imaging (CXR) and computed tomography in the diagnosis of COVID-19 has been investigated. The combination of CXR and AI, such as deep learning (DL), has been extensively examined for automatic diagnosis of COVID-19. Since CXR is widely available and its cost is relatively low, the combination of CXR and artificial intelligence could be employed for screening purposes of COVID-19 without the need for medical doctors.

Automating the diagnosis of many diseases nowadays has been based on artificial intelligence, which has proven its efficiency and high performance in automatic image classification problems through different machine learning approaches. Moreover, machine learning defines models that have the ability to learn and make decisions by using large amounts of input data examples. Deep learning efficiently generates models that produce more accurate results in predicting and classifying different diseases using images like breast cancer, liver diseases, colon cancer, lung disease, skin cancer, lung cancer, pneumonia, and recently COVID-19 diagnosis, without requiring any human intervention. The main reason for using deep learning is that deep learning techniques learn by creating a more abstract representation of data as the network grows deeper. As a result, the model automatically extracts features and yields higher accuracy results.

II. LITERATURE SURVEY

1. D Santhosh Reddy, R Bharath, and P Rajalakshmi have presented a paper titled “A Novel Computer-

Aided Diagnosis Framework Using Deep Learning for Classification of Fatty Lung Disease in Ultrasound Imaging" at the IEEE 20th International Conference on e-Health Networking, Applications, and Services (Health-com) in 2018. The proposed framework uses deep learning algorithms to analyze ultrasound images and identify patterns that are associated with fatty lung disease. Specifically, the framework uses a convolutional neural network (CNN) to extract features from ultrasound images and a support vector machine (SVM) to classify the images as either healthy or diseased.

2. Jagdeep Singh et al.in 2019 have suggested a computer programmer that determines a patient's risk of developing an illness based on their presentation of symptoms. Using a patient's symptoms to determine the underlying illness is a popular method in medical diagnosis. The creation of a software system to aid in this process has a number of potential advantages. By increasing the precision and speed of diagnosis, for instance, patients may experience better treatment outcomes.

3. Sanjay Kumar, Sarthak Katyal, "Effective Analysis and Diagnosis of Lung Disorder by Data Mining", International Conference on Inventive Research in Computing Applications (ICIRCA), 2019The use of data mining techniques for the analysis and diagnosis of Lung problems is most likely the paper's primary topic. In order to find pertinent patterns in medical data for efficient diagnosis and treatment, data mining is a process for extracting patterns and insights from massive datasets.

4. K. Thirunavukkarasu, Ajay S. Singh, Md Irfan, Abhishek Chowdhury, "Prediction of Lung Disease using Classification Algorithms", 4th International Conference on Computing Communication and Automation (ICCCA), 2019.The main topic of the paper is the classification algorithms-based prediction of Lung disease. Lung diseases a major global health concern, and early detection and diagnosis are essential for successful management and treatment. The authors suggest using classification algorithms to create a model that can predict lung disease.

III. SYSTEM ANALYSIS

Lung disease causes is a serious issue that leads to death. Thus, it is important to detect as early as possible. The chest X-ray images can be used for tumor detection. This kind of imaging is often used by scientists in detecting lung disease. There are different methods developed by researchers for detecting and classifying lung diseases using chest X-ray images.

These approaches range from traditional medical image processing to advanced machine learning techniques.
Also, deep learning techniques can perform prediction without direct human intervention. Thus, in this implementation, I use ResNet Neural Network 50 (ResNet 50) to make a prediction model for lung disease classification. Most of the existing work apply image processing techniques for the disease classification. Most of the methods based on machine learning or deep learning technique make predictions for binary class that is, diseased or not. Thus, it is important to create an efficient system that can make predictions on lung X-ray images for Different disease categories.

3.1. EXSITING SYSTEM

In most of the existing lung X-Ray image classification papers, the authors consider COVID-19 disease only. Few researches include pneumonia chest X-ray classification. existing works does not include multi class classification for lung X-ray classification.

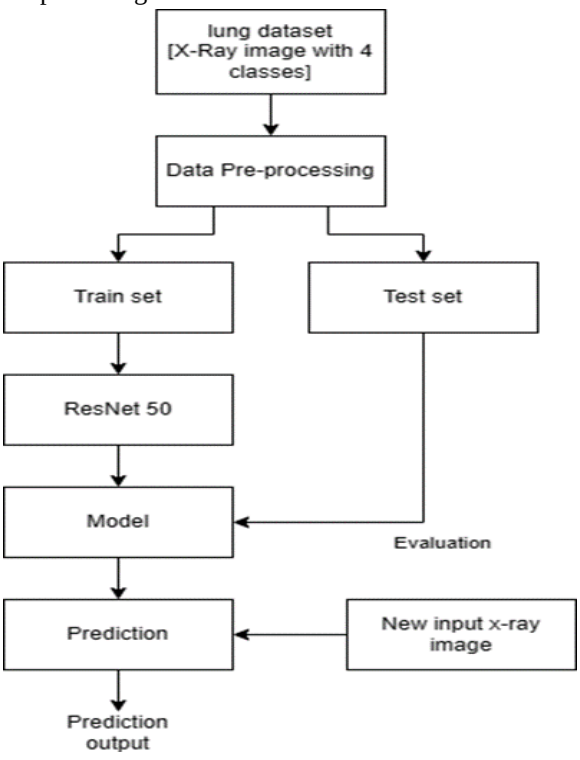
3.2. PROPOSED SYSTEM

The proposed work is based on deep learning technique for the classification of lung diseases using lung X-Ray images. The lung disease classes for the proposed work are Tuberculosis, Covid 19, Pneumonia, Healthy. The dataset for the proposed work is available in the public repository. The deep learning algorithm used is Residual Network 50 (ResNet 50). For the proposed implementation, technology selected is python. For the implementation, the IDE is Visual Studio Code etc.

IV. SYSTEM ARCHITECTURE

The implementation starts with the collection of chest X-Ray image data for the lung disease classification

system. The dataset is available in the Kaggle repository, which contains 4 classes of Lung X-ray Images. Then, pre-process the data to remove outliers to improve the data for further processing. Then, partition the data as train set and test set. Then, choose the algorithm for training. Here, I select Residual Network 50 (ResNet 50) for classification, since Resnet 50 can work better for image classification. Design the neural network by setting up the network layers and network parameters like weight, input shape, activation, optimization technique, learning rate, decay value etc. compile the designed network and then move to the training phase by applying train set and test set. After training completes evaluate the model for performance evaluation. If the model has acceptable performance, then, save the model for future prediction. For prediction, apply a pre-processed input chest X-Ray image and then the model predict the result. For the implementation of the neural network, keras can be use, which is a python API for deep learning.



V. SYSTEM IMPLEMENTATION PHASES

Data Collection

Data collection is the process of gathering and measuring information from countless different sources. In order to use the data, we collect to develop practical artificial intelligence (AI) and machine learning solutions, it must be collected and stored in a way that makes sense for the business problem at hand.

- Here, the data used is the chest X-Ray images.
- X-Ray is basically used in the biomedical to detect and visualize finer details in the internal structure of the body.
- The dataset is available in the public repository, and can be directly downloaded.
- The input is the chest X-Ray image with 4 classes- Tuberculosis, Covid 19, Pneumonia, and Healthy.
- The file extension of chest X-ray image is png.

Data Pre-processing

A real-world data generally contains noises, missing values, and maybe in an unusable format which cannot be directly used for machine learning models. Data preprocessing is required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

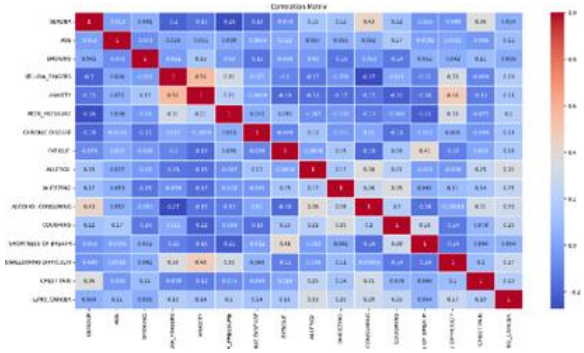
- Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model.
- Pre-processing is the data cleaning phase that enhances and improves the input data for further tasks.
- The primary task of medical imaging analysis is to clean the chest X-ray images and to enhance the contrast.
- Thus, different machine leaning and image processing algorithms were deployed to enhance the contrast of chest R-ray images.
- Image resizing is performed to resize the image having any size to a fixed size. This process is required as the neural network process every data in a uniform manner.
- Then, data normalization is performed on the pixel values of chest X-ray images, to convert the range of input pixel value ranges from (0,255) to (0,1) for faster processing.
- The process of transforming the columns in a dataset to the same scale is referred to as normalization.
- Then, convert the 4 class labels (Tuberculosis, Covid 19, Pneumonia, and Healthy) to categorical value, since the neural network handle numerical value.

Data Partitioning

Data partitioning in data mining is the division of the whole data available into two or three non-overlapping sets: the training set, the validation set, and the test set. If the data set is very large, often only a portion of it is selected for the partitions.

- At the time of training the models, partition all the provided data into two sets: the training set, which will actually be used to adjust the parameters of the models; and the test set, which will be used to measure the accuracy of the chosen model.
- The proportions of these partitions are normally 80/20.
- For partitioning the data into training set and testing set is performed by using the train_test_split () method of sklearn python package by providing the dataset and split ratio as parameters to t

VI. RESULTS



VII. CONCLUSION

Lung disease is a leading cause of death worldwide. In this paper, proposes an efficient method for automatic lung disease classification using X-ray images. The proposed method is based on deep learning technique and implemented a ResNet 50 architecture for lung disease classification. For the proposed implementation the dataset is available in the public repository. It consists of 4 classes of lung X-ray images- Tuberculosis, Covid 19, Pneumonia and Healthy. All images are in .png format. ResNet 50 is proposed to classify the pre-processed input image.

For evaluation, different performance evaluation matrices can calculate to evaluate the performance of the trained model.

REFERENCES

[1] S. Raschka and V. Mirjalili, *Python Machine Learning*. Packt Publishing.

[2] Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.

[3] Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press.

[4] M. Bishop, *Pattern Recognition and Machine Learning*. Springer.

[5] W. McKinney, *Python for Data Analysis*. O'Reilly Media.

[6] "Scikit-learn Official Documentation," Scikit-learn.

[7] "TensorFlow Official Documentation," TensorFlow.