ABSTRACT

Tuberculosis is one of the most ancient diseases and still it is one of the top 10 causes of death across the world. Most people who get infected with tuberculosis can be saved with proper treatment and their Life can be saved but due to lack of medical support to detect tuberculosis in most parts of World still mortality rate due to tuberculosis is high. This project helps to detect tuberculosis by using image processing techniques over chest x-rays.

Our objective is to prepare a model which classifies the chest x-ray. This model contains two classes normal and abnormal (infected with TB) we need to classify between these two classes and also to achieve high accuracy while classifying.

In our proposed approach we will be improving the accuracy of model by using deep neural networks to train the model and that model helps us in classifying new chest x-ray given as input to the model thus meeting our objective.

To achieve good accuracy, we need to pre-process the images first we pre-processed the images we took images from both datasets Shenzhen and Montgomery and together there are 800 chest x rays we did augmentation over these images and normalized them followed by giving these pre-processed images as inputs to our models.

In this project we used two models baseline CNN model and pretrained VGG16 model and gave pre-processed images as inputs to these both models and evaluated the models to see which performed better comparing using different performance metrics like accuracy, specificity, sensitivity, precision and f1-score and depicted them using graphs and tables From the above results we made a classification reports for both models.

CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER-1

INTRODUCTIO

N

1.1 OVERVIEW

According to WHO TB (tuberculosis) [1] is one of the top 10 causes of death across the world in 2018 around 10.4 million people fell ill from TB i.e. around 28,500 people per day out of them 1.8 million people have died i.e. around 4500 per day. With proper treatment TB can be cured but most of the people have no access for treatment there aren't many specialists to check whether people are infected or not. Manually checking the chest x-rays and detecting the infection needs radiology experts and there are not many experts compared to number of people getting infected.

Image processing is that the wide used modern field in automating utterly totally different techniques and procedures. Image processing has nice impact in medical sector

[2] as a result of as patients unit increasing day by day so treating them manually is very powerful, so this approach is utilized for quick automation. we tend to tend to automatize utterly totally different medical procedures previously and it helped at intervals the designation and treatment of various diseases. Variety of the designed automatic systems unit CT scan processed axial imaging digital analysis associate disease that's transmitted through a medium that's a bacterium referred to as mycobacteria. The chest x rays samples of T.B patients unit taken as a medium for screening. The paper collectively includes discussion and analyzation of the techniques that unit previously designed for

T.B detection mistreatment analysis photos and chest x rays.

Major issues involved in tuberculosis control programs [2]are the management of pulmonary tuberculosis and Case-finding. The countries which are affected by epidemics of human immunodeficiency virus infection are facing tuberculosis as a more serious problem. The diagnosis of tuberculosis using accurate methods is one of the crucial steps involved to control the occurrence and prevalence of TB. However, the manual diagnosis of tuberculosis is quite complex these days, so there is no standard method at present.

Hence, we need to develop a solution detects whether the person is infected or not with high accuracy using image processing.

1.2 Importance of Project and objectives

Objective of this project is to detect whether the person is infected with TB or not using the chest x-ray of the person.

- To classify across various chest x-rays with high accuracy.
- With this it is easy to detect whether the person is infected or not without involvement of a specialist.
- Thus, it helps many people to know infection at an early stage and take necessary steps to cure it.
- Deaths caused by this epidemic will be drastically reduced as detection becomes easy.

1.3 Scope of Project

- This project has a lot of scope in medical sector as there are many TB infected patients compared to radiology specialists.
- Requirement to achieve this project is dataset containing CXR of normal and infected images.

1.4 Motivation

- Many people are dying across the world because of TB and most of these can be saved by just giving the proper treatment at right time i.e. at early stages.
- First step to do that would be detecting whether a person is infected or not.
- Thus, many lives across the world can be saved and most of the developing countries are facing the problem of TB thus helps them.

1.5 Organisation of Project Report

Organisation of project is done in the following way, initially we started with introduction part where we explained the overview of project including objective, scope of project and motivation.

In the preceding chapters we explained all the research papers we used as a reference to this project. we went through all the papers and studied various advantages and limitations of these papers and explained the summary of all papers in the below sections. In the next chapters we explained the technologies being used, proposed methodology, results and performance evaluation along with conclusion and future scope followed by reference of the project.

SciKit Learn, Pandas library, NumPy, Keras, Kaggle kernel are the technologies being used in this project.

In the proposed methodology we explained about the datasets being used along with pre- processing techniques and baseline CNN architecture and pretrained VGG16 model architecture and followed by the steps involved in training the models.

In chapter 4 evaluation of project is explained and performance results are depicted using tables and graphs followed by conclusion and future scope of project in chapter 5.

2.1 OVERVIEW

- In a paper [3] published in 2017 at “IEEE International Conference on Signal and Image method Applications (ICSIPA)” detailed study is done and written about the potential technique for communicable disease classification victimization CNN is used which classifies the images into 2 categories, throughout this paper they have used a CNN style with 7 convolutional layers and 3 all connected layers and performance of assorted optimizers are a validation accuracy of 82.09% has been obtained.
- In a paper [4] Deep Learning at Chest radiography published by Paras Lakhani and Baskaran Sundaram both AlexNet pre-trained, AlexNet untrained and GoogleNet pre-trained and GoogleNet untrained CNN networks were used and compared pretrained models have provided a better accuracy compared to untrained models and image augmentation has also improved the accuracy of the models. Untrained models have shown an AUC of 0.88 and with augmentation it has improved slightly. Advantage with deep learning is it provides a better accuracy with higher dimensional datasets like images and pre-trained models have performed better compared to untrained model as starting layers of neural networks are the same for all images like edges and blobs.
- In a paper [5] published in 2018 in NCBI (National centre for biotechnology information) focusing on different technologies and methods that can be used for pre-processing segmentation and processing of images paper reviews on common methods of computer aided detection of chest radiographs based on AI. Usually rib structure is not removed but removing rib structure and clavicle the accuracy can be improved. Traditional algorithms like SVM (support vector machine) and random forest may be better, but deep learning methods are providing better performance compared thus they deep learning methods are becoming mainstream in terms of image processing.
- In paper [6] “Comparing deep learning models for population screening using chest radiography” compared across many models and published the results in this