**PNEUMOCONIOSIS DETECTION**

**USING DEEP LEARNING**

**ABSTRACT**

Pneumonia is a disease which occurs in the lungs caused by a bacterial infection. Early diagnosis is an important factor in terms of the successful treatment process. Pneumonia impacts all the elderly and young people’s families and children everywhere but is most prevalent in Sub-Saharan Africa and South Asia. In December 2019 Wuhan, a city of China was affected by deadly, gruesome Pneumonia which was declared a pandemic by World Health Organisation. But the reason for the outbreak was not clear to everyone. Later, the doctors identified the disease as a new species of coronavirus, also currently known as COVID-19. Generally, the disease can be diagnosed from chest X-ray images by an expert radiologist. The diagnoses can be subjective for some reasons such as the appearance of disease which can be unclear in chest X-ray images or can be confused with other diseases. Therefore, computer-aided diagnosis systems are needed to guide the clinicians. In our project **Pneumoconiosis**, we used two well-known convolutional neural network models Xception and Vgg16 for diagnosing of pneumonia. We used transfer learning and fine-tuning in our training stage. The test results showed that Vgg16 network exceed Xception network at the accuracy with 0.87%, 0.82% respectively. However, the Xception network achieved a more successful result in detecting pneumonia cases. As a result, we realized that every network has own special capabilities on the same dataset.

**CHAPTER 1**

**INTRODUCTION**

* 1. General Introduction

Pneumonia is inflammation of the tissues in one or both lungs that usually caused by a bacterial infection. In the USA annually more than 1 million people are hospitalized with the gripe of pneumonia. Unfortunately, 50.000 of these people die from this illness. Fortunately, pneumonia can be a manageable disease by using drugs like antibiotics and antivirals. However, early diagnosis and treatment of pneumonia is important to prevent some complications that lead to death. Chest X-ray images are the best-known and the common clinical method for diagnosing of pneumonia. However, diagnosing pneumonia from chest X-ray images is a challenging task for even expert radiologists. The appearance of pneumonia in X-ray images is often unclear, can confuse with other diseases and can behave like many other benign abnormalities. These inconsistencies caused considerable subjective decisions and varieties among radiologists in the diagnosis of pneumonia. Therefore, there is a need for computerized support systems to help radiologists for diagnosing pneumonia from chest X-ray images. Recent developments in deep learning field, especially convolutional neural networks (CNNs) showed great success in image classification. The main idea behind the CNNs is creating an artificial model like a human brain visual cortex. The main advantage of CNNs, it has the capability to extract more significant features from the entire image rather than handcrafted features. Researchers developed different CNN based deep networks and these networks achieved state of results in classification, segmentation, object detection and localization in computer vision. Besides the natural computer vision problems, CNNs achieved very successful results in solving medical problems such as breast cancer detection, brain tumor segmentatio, Alzheimer disease diagnosing, skin lesion classification etc.

In **Pneumoconiosis**, we modified and trained two well-known networks for classifying pneumonia from chest X-ray images. Our first network is based on the Xception model. The second one is Vgg16 based model. Besides, we utilized transfer learning, fine-tuning and data augmentation methods. For an objective comparison between them, we used same parameters when training both networks. Also, we compared the performance of two networks on the test data with different metrics. The results show that the Xception model outperforms Vgg16 model in diagnosing pneumonia. On the other hand, Vgg16 model showed better performance in diagnosing normal cases.

* 1. Goal of the Project

The goal of the new system is to develop a model that will identify whether a patient is having Pneumonia or not by passing the chest X-Ray images through the Deep Learning model. The model should be highly precise as people's lives are at stake. As doctors must do a lot of certain tests to identify if the patient has Pneumonia or not. In the past, it has been observed that the doctors undergo testing or get an X-Ray and give a false positive or a false negative result, which resulted in bad medical conditions of the patients. New coronavirus get severe pneumonia in both lungs. Covid-19 pneumonia is a serious illness than can be deadly. In Covid-19 situations, corona virus spread over the world. Corona patients are increased day by day. It is difficult to find when a patient having corona or not by doctors also. Corona can detected by the person conducting the test will insert a long stick with a very soft brush on the end — kind of like a pipe cleaner — up your nose and twirl it around for a few seconds. The soft bristles will collect a sample of secretions there for analysis. The swab has to go pretty far back, because cells and fluids must be collected from along the entire passageway that connects the base of the nose to the back of the throat to get a really good specimen. It takes one week for the result. Within in this period, if the person is suffering corona in a critical condition his life will be at risk. So, Pneumoconiosis can help in the medical fields in order to get the result fast and to reduce these types of errors and save lives. To solve the cumbersome problem, an ensemble of two deep learning models is developed, to make the work of the doctors simpler. In Pneumoconiosis, the input image is compared with the images in the dataset, then each image in the dataset is compared with these input image and check whether the patient is infected or not. The main motivation behind this project was to identify Covid-19 pneumonia just by using the X-Ray images of the patients.

**CHAPTER 2**

**LITERATURE STUDY**

**2.1 Study of Similar Works**

Dimpy Varshn explained the development of an automatic system for detection of pneumonia through various deep learning models. The authors analyzed medical images and developed a Convolutional Neural network for disease classification and scaling of data. The architecture consists of a DenseNet-169 layer architecture for feature extraction. The architecture was combined with a SVM Model for binary classification. The results of the model are analyzed with visualization curves and a summary is provided for the same. Okeke Stephen. provides a similar insight on classifying numerous x-rays to detect pneumonia based on Convolutional Neural Networks. The accuracy obtained through their research helps us evaluate our model in comparison, depending on the loss and accuracy of the neural network. Their network is provided with a (200x200x3) dimension input shape, while the images are focused and it uses (64x64x3) dimensions, to decrease the computations.

Over the last decade, several machine learning based automated methods for identifying different types of pneumonia have been widely studied. Fiszman used a natural language processing (NLP) tool to identify acute bacterial pneumonia-related disease in chest X-ray. Performance of this type of resource intensive application is very much comparable to that of the human expert. Chapman demonstrated three computerized methods using a rule base, a probabilistic Bayesian network, and a decision tree to diagnose the chest X-ray report associated with acute bacterial pneumonia. The study of feasibility of an NLP-based monitoring system is done to identify healthcare-associated pneumonia in neonates. However, practical clinical applications of these types of methods are limited due to the dependency on the information extracted from the narrative reports of the patients. Parveen reports an unsupervised fuzzy c-means classification learning algorithm to detect pneumonia infected X-ray images. This approach improves classification accuracy as fuzzy c-means allocate weights to all the pixels of the input X-ray images. Rajpurkar demonstrated ChexNet, a 121-layer deep Convolutional Neural Network (CNN), that provides the probability of detecting or identifying pneumonia using a heatmap to localize the area of the infection. Kermany introduced a transfer learning based DL framework to diagnose pediatric pneumonia using chest X-ray images. However, none of the methods are exploited to classify X-ray images with pneumonia for the CS framework to meet the need of remote end analysis.

**2.1.1 Existing System**

Tedros and other experts agree that mass testing for COVID-19, the disease caused by the coronavirus, would allow positive cases to be isolated and help identify those who came in contact with them, helping to curb further transmission. One method to detect corona through the person conducting the test will insert a long stick with a very soft brush on the end — kind of like a pipe cleaner — up your nose and twirl it around for a few seconds. The soft bristles will collect a sample of secretions there for analysis. The swab has to go pretty far back, because cells and fluids must be collected from along the entire passageway that connects the base of the nose to the back of the throat to get a really good specimen. The body is not used to having an object in that area, though, so it creates a lot of very odd sensations. For one thing, it activates the lachrymal reflex, which means it’ll bring tears to your eyes if it’s done correctly. I wouldn't go so far as to say it hurt, but it is uncomfortable. Since the swab will also touch the back of the throat, it may also trigger a gag reflex.

Another method of detecting pulmonary tuberculosis following the architecture of two different DCNNs AlexNet and GoogleNet. Lung nodule classification mainly for diagnosing lung cancer proposed by Huang also adopted deep learning techniques. Performance of different variants of Convolutional Neural Networks (CNNs) for abnormality detection in chest X-Rays was proposed by Islam using the publicly available OpenI dataset. For the better exploration of machine learning in chest screening, Wang released a larger dataset of frontal chest X-Rays.

**2.1.2 Drawbacks of Existing system**

During our project we went through the different system development life cycle. First of all we started with system study which helped us understand scope of the system. During this phase, we are able to understand the limitations of the existing system and it also helped us in realizing the requirements from the client’s perspective.

* The test result for detecting corona will take one week. Within in this period, if the person is suffering corona in a critical condition his life will be at risk.
* Researchers have reported that testing of oropharyngeal secretions — secretions from the part of the throat at the back of the mouth — may reduce the number of false negative results.
* False negatives — when a person carrying a pathogen tests negative — have been reported several times during the [Covid-19](https://indianexpress.com/about/coronavirus/) [pandemic](https://indianexpress.com/article/explained/pandemic-explained-who-novel-coronavirus-covid19-what-is-a-pandemic-6309727/).
* These results have come up during nasal swab testing of patients who have seemingly recovered from the disease — but have later been found to be still carrying the virus.
* A small number of patients who had tested negative through nasal swabs were found to be positive through the testing of oropharyngeal secretions.
* The study included 75 ready-for-discharge Covid-19 patients who tested negative using nasal swabs. Because of detection of potential false-negatives in that cohort, a second study paired oropharyngeal and nasopharyngeal samples collected from 50 additional Covid-19 recruits during their recovery stage.
* The famous [CNN](https://www.baeldung.com/cs/ml-relu-dropout-layers) AlexNet for image classification, for example, required six days to train on two GPUs. A problem then arises when either the dataset or the size of the neural network becomes too large. Neural networks do not generally scale well as the amount of data or the number of their layers and neurons increases.
* In GoogleNet, further changes in the architecture are proposed, which make the model perform better. One such change is termed as an Xception Network, in which the limit of divergence of inception module are increased.

**CHAPTER 3**

**OVERALL DESCRIPTION**

3.1 Proposed System

3.2 Features of Proposed System

3.3 Functions of Proposed System

3.4 Requirement Specification

System analyst tasks to a variety of persons to gather details about the business process and their opinions of why things happen as they do and their ideas for changing the process. These can be done through questionnaires, details investigation, observation, collection of samples etc. As the details are collected, the analyst study the requirements data to identify the features the new system should have, including both the information the system produce and operational features such as processing controls, response times, and input output methods.

Requirement specification simply means, “Figuring out what to make before you make it”. It determines what people need before you start developing a product for them. Requirement definition is the activity of translating the information gathered in to a document that defines a set of requirements. These should accurately reflect what consumer wants. It is an abstract description of the services that the system should provide and the constraints under the system must operate. This document must be written for that the end user and the stake holder can understand it.

The notations used for requirements definition should be based on natural languages, forms and simple intuitive diagrams. The requirements fall into two categories: functional requirements and non-functional requirements.

The requirements of specification of the proposed system are as follows:

Minimum time needed for various processing

Better Service

Faster response time

User Friendly

**3.5 Feasibility Analysis**

Feasibility analysis is the test of the system proposal made to identify whether the user needs may be satisfied using the current software and hardware technologies, whether the system will be cost effective from a business point of view and whether it can be developed with the given budgetary constraints.

When a new project is proposed, it normally goes through feasibility assessment. Feasibility study is carried out to determine whether the proposed system is possible to develop with available resources and what should be the cost consideration. Facts considered in the feasibility analysis were:

**3.5.1 Technical Feasibility**

Technical Feasibility deals with the hardware as well as software requirements. Technology is not a constraint to type system development. We have to find out whether the necessary technology, the proposed equipment have the capacity to hold the data, which is used in the project, should be checked to carry out this technical feasibility.