



**INFORMATICS
INSTITUTE OF
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Hot Spring

**Machine Learning for Disease Prediction Using Symptoms and X-ray
Images**

A dissertation by

Ravija Gunawardana

2019340

Supervised by

Mr. Banuka Athuraliya

Submitted in partial fulfillment of the requirements for the BEng in Software
Engineering degree at the University of Westminster.

April 2023

ABSTRACT

People of all ages, ethnicities, and nations are impacted by diseases, which are a worldwide problem. The common cold, Influenza, Tuberculosis (TB), HIV/AIDS, and more recently COVID-19 and Pneumonia are some of the most well-known illnesses. Diseases vary in how quickly they spread, from those that are extremely infectious and spread quickly, like the flu, to others that spread more slowly, like TB. The way a disease spreads, such as by direct contact, airborne droplets, or insect bites, can have a significant influence on how quickly it spreads. A person's age, immune system, and general health can also have an impact on how quickly a disease spreads. Pneumonia diagnoses continue to increase annually by millions of individuals. In order to accurately diagnose and treat diseases and get the greatest results, it is crucial. Medical professionals go to the subsequent stage, which involves reviewing the patient's medical records, after determining the presence of a disease based on symptoms. This is a drawn-out procedure since there are increasingly more patients than there are physicians and medical equipment. This work must be carried out by qualified radiologists in serious cases; else, illness identification may not be successful. To fulfill the task of disease prediction efficiently and precisely, a system is needed.

The study utilized clear pneumonia images as input for the model. Following necessary pre-processing techniques, the dataset was trained on a Mask R-CNN, which achieved an overall accuracy of 85% on the dataset. The best model weight was loaded to make the final prediction using the testing dataset.

Support Vector Machine Classifier (SVM), K-Nearest Neighbor Classifier (KNN), and Random Forest Classifier (RF) techniques are utilized for the Machine Learning (ML) aspect (disease prediction using symptoms) since it would be simple to acquire the optimum answer by categorizing data into branches and then getting predictions from each branch. The trained ensemble model achieved an accuracy of 97% on the dataset.

Keywords:

MASK R-CNN, Random Forest, K-Nearest Neighbor, Support Vector Machine, X-ray, Chest, Disease, Symptoms, Prediction, Convolution Neural Networks