<div align="center">

Machine Learning Engineering Nanodegree
Capstone Project Report

ChestR
Disease Predictor for Chest
X-Rays

Shishir Horane

February 1, 2019

</div>

1 Project Overview

In medical industry chest x-ray is the most commonly performed diagnostic x-ray examination. It produces images of the heart, lungs, airways, blood vessels and the bones of the spine and chest. An x-ray(radiograph) is a non-invasive medical way for physicians to diagnose and treat medical conditions. Performing such an x-ray involves exposing the the chest to a small dose of ionizing radiation to produce the pictures of the body. X-rays are the most frequently used form of medical imaging.

The chest x-ray is commonly the first imaging test used to help diagnose symptoms such as chest pain, breathing difficulties, bad or consistent cough etc. Physicians will use these imaging to diagnose or monitor treatments for conditions such as pneumonia, heart problems, lung cancer, fluid or air collection in lungs etc. Although this isn't necessarily a conclusive diagnosis for all kind of diseases it is the first tool of diagnosing any possible illnesses for the above said symptoms.[1]

¡Student provides a high-level overview of the project in layman's terms. Background information such as the problem domain, the project origin, and related data sets or input data is given.¿

2 Problem Statement

Chest X-ray exams are one of the most frequent and cost-effective medical imaging examinations available. However, clinical diagnosis of a chest X-ray can be challenging and sometimes more difficult than diagnosis via chest CT.[2]

Reading and diagnosing chest x-ray images may be a relatively simple task for radiologists but, in fact, it is a complex reasoning problem which often requires careful observation and knowledge of anatomical principles, physiology and pathology. Such factors increase the difficulty of developing a consistent and automated technique for reading chest X-ray images while simultaneously considering all common thoracic diseases.[3]

There is a need to develop an automated technique that could analyze a given chest x-ray and which could predict the possible common thoracic diseases with some certainty. Also, it should aid radiologists or physicians by highlighting the suspected common areas of disease in order to pay close attention to certain areas of a chest x-ray, so as to not misdiagnose these diseases. This automated technique could benefit patients in developing countries who do not have access to the radiologists to read these x-rays. ¡Metrics used to measure performance of a model or result are clearly defined. Metrics are justified based on the characteristics of the problem.¿

3 Data Exploration

If a dataset is present, features and calculated statistics relevant to the problem have been reported and discussed, along with a sampling of the data. In lieu of a dataset, a thorough description of the input space or input data has been made. Abnormalities or characteristics about the data or input that need to be addressed have been identified.

4 Exploratory Visualization

A visualization has been provided that summarizes or extracts a relevant characteristic or feature about the dataset or input data with thorough discussion. Visual cues are clearly defined.

5 Algorithms and Techniques

Algorithms and techniques used in the project are thoroughly discussed and properly justified based on the characteristics of the problem.

6 Benchmark

Student clearly defines a benchmark result or threshold for comparing performances of solutions obtained.

7 Data Preprocessing

All preprocessing steps have been clearly documented. Abnormalities or characteristics about the data or input that needed to be addressed have been corrected. If no data preprocessing is necessary, it has been clearly justified.

8 Implementation

The process for which metrics, algorithms, and techniques were implemented with the given datasets or input data has been thoroughly documented. Complications that occurred during the coding process are discussed.

9 Refinement

The process of improving upon the algorithms and techniques used is clearly documented. Both the initial and final solutions are reported, along with intermediate solutions, if necessary.

10 Model Evaluation and Validation

The final model's qualities — such as parameters — are evaluated in detail. Some type of analysis is used to validate the robustness of the model's solution.

11 Justification

The final results are compared to the benchmark result or threshold with some type of statistical analysis. Justification is made as to whether the final model and solution is significant enough to have adequately solved the problem.

12 Free-Form Visualization

A visualization has been provided that emphasizes an important quality about the project with thorough discussion. Visual cues are clearly defined.

13 Reflection

Student adequately summarizes the end-to-end problem solution and discusses one or two particular aspects of the project they found interesting or difficult.

14 Improvement

Discussion is made as to how one aspect of the implementation could be improved. Potential solutions resulting from these improvements are considered and compared/contrasted to the current solution.

References

[1] https://www.radiologyinfo.org/en/info.cfm?pg=chestrad

[2] https://www.kaggle.com/nih-chest-xrays/data/home

[3] https://www.nih.gov/news-events/news-releases/nih-clinical-center-provides-one-largest-publicly-avail

[4] https://www.kaggle.com/nih-chest-xrays/data/downloads/data.zip/3

[5] https://arxiv.org/pdf/1711.05225.pdf

[6] https://www.kaggle.com/kmader/train-simple-xray-cnn