CHAPTER 1

INTRODUCTION

Pneumonia is an epidemic with high mortality rate. Its seriousness ranges from mild to life-threatening. It is most serious for infants and young children, people older than age 65, and people with health problems or weakened immune systems [1]. Pneumonia is caused by either bacteria, viruses or fungi. Luckily, pneumonia can be well treated with allopathic medicines.

The risk of pneumonia is found to be high in developing nations. Lack of medical resources and personnel aggravates this risk. Early detection and treatment are required to prevent disease spread and mortality rate of pneumonia. Accurate and fast diagnosis methods can guarantee timely access to treatment and save time and costs.

Chest scan X-ray images are one of the best modalities for diagnosing pneumonia. However, visual analysis and examination of the images are time consuming and less efficient. CNN-motivated deep learning algorithms can serve the purpose of medical image classifications efficiently. A spatial pyramid pooling module can enhance the capability of CNNs to predict accurately with varying image sizes and fewer number of parameters. Their combined effect can be applied here to classify different types of pneumonia for various severity levels.

A lightweight model based on CNNs with a modified SPP is proposed for fast diagnosis of chest X-ray images to detect different types of pneumonia for various severity levels. It is cost-efficient and can be deployed on various platforms including mobile phones, tablets, and normal computers. The model can help doctors diagnose and predict disease risk accurately and quickly.

1.1 Problem Statement

Pneumonia is a highly contagious disease and has high mortality rates. To cure pneumonia, early detection and treatment are required. However, lack of accurate and fast diagnosis tools for pneumonia delays the start of medication and treatment, which also leads to potential disease spread. At present, disease diagnosis relies on the expertise of the doctor who performs visual examination of chest X-rays of the patient. This method is time consuming, less accurate and costly.

A diagnostic software that can classify different types of pneumonia for various severity levels quickly and accurately is much needed. Additionally, it should be cheap and capable of deployment in all common application platforms to provide ease of access and affordability to all people.

1.2 Motivation and Objectives

Recent research and development in artificial intelligence and machine learning have led to an outburst of innovations in technology that finds its applications in a plethora of fields. As responsible citizens and future graduates in technology, we find it our bounden duty to exploit technology to the best use for humanity. There's nothing more precious than human lives, but sadly, we have lost many due to technical or/and economic incapabilities. This realization has led us to the conclusion that the health sector should be the primary beneficiary of technological research.

Pneumonia is a highly contagious disease and has a high mortality rate. As per WHO statistics [2], pneumonia accounts for 15% of all deaths of children under 5-years old, killing 8,08,694 children in 2017. The same survey reveals that the cost of antibiotic treatment and diagnosis for all children with pneumonia is around US$ 109 million per year. Early detection and treatment are necessary to cure pneumonia patients and prevent the death rate. However, at present, technical ineptness delays fast diagnosis and treatment of the disease. A chest X-ray is the most common and cheapest way for diagnosing pneumonia. But since diagnosis relies on a doctor who performs a visual examination of the chest X-ray, the process is time-consuming and inaccurate.

CNNs are an efficient tool for image analysis and detection. Spatial Pyramid Pooling (SPP) enhances the diagnostic capability of CNNs and enables predictions with a comparatively lesser number of parameters. With the aid of CNNs and SPP, we aim to develop a lightweight deep learning model for accurate and fast diagnosis of different types of pneumonia for various severity levels using chest X-ray images.

It is possible to deploy a lightweight model on all application platforms like mobile phones, tablets and personal computers, without worrying about memory and processing capabilities. Moreover, it reduces the cost and storage space required by the model. Hence, being lightweight is one of the main objectives of this project.

The proposed model can help in alleviating the economic burden of diagnostic tests for pneumonia. It offers a fast, accurate and affordable solution.

In most situations, fast and affordable diagnosis can help in the early start of treatment. We hope the proposed model can help to save the lives of many pneumonia patients.

1.3 Area of Project

Development of the proposed model involves the following major research areas:

1. **Convolutional Neural Networks (CNNs)**
2. **Spatial Pyramid Pooling (SPP)**

1.3.1 Convolutional Neural Networks (CNNs)

A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. The role of the CNN is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction [3]. CNNs are regularized versions of multilayer perceptrons. A multilayer perceptron usually means fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, kernels.

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The activation function is commonly a ReLU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution.

When programming a CNN, the input is a tensor with shape (number of images) x (image height) x (image width) x (image depth). Then after passing through a convolutional layer, the image becomes abstracted to a feature map, with shape (number of images) x (feature map height) x (feature map width) x (feature map channels) [4]. Following are the layers **[INPUT-CONV-RELU-POOL-FC]** that are usually used to construct Convolutional neural networks [5].

- **INPUT**− As the name implies, this layer holds the raw pixel values. Raw pixel values mean the data of the image as it is.
- **CONV**− This layer is one of the building blocks of CNNs as most of the computation is done in this layer.

- **RELU**−Also called rectified linear unit layer, that applies an activation function to the output of previous layer. In other manner, a non-linearity would be added to the network by RELU.
- **POOL**− This layer, i.e., Pooling layer is one other building block of CNNs. The main task of this layer is down-sampling, which means it operates independently on every slice of the input and resizes it spatially.
- **FC**− It is called Fully Connected layer or more specifically the output layer. It is used to compute output class score and the resulting output is volume of the size 1*1*L where L is the number corresponding to class score.

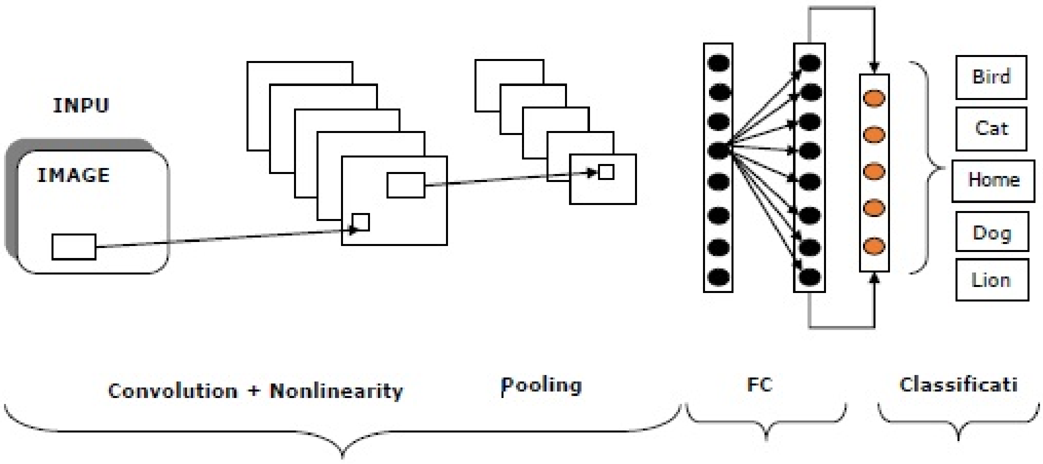

Figure 1.1 Architecture of CNNs.

1.3.2 Spatial Pyramid Pooling (SPP)

Spatial pyramid pooling (SPP) adds a new layer between the convolutional layers and the fully-connected layers. Its job is to map any size input down to a fixed size output. The ideal of spatial pyramid pooling, also known as spatial pyramid matching or just ‘multi-level pooling’ pre-existed in computer vision, but had not been applied in the context of CNNs.

SPP works by dividing the feature maps output by the last convolutional layer into a number of spatial bins with sizes proportional to the image size, so the number of bins is fixed regardless of the image size. Bins are captured at different levels of granularity – for example, one layer of 16 bins dividing the image into a 4×4 grid, another layer of 4 bins dividing the image into a 2×2 grid, and a final layer comprising the whole image. In each spatial bin, the responses of each filter are simply pooled using max pooling [6].

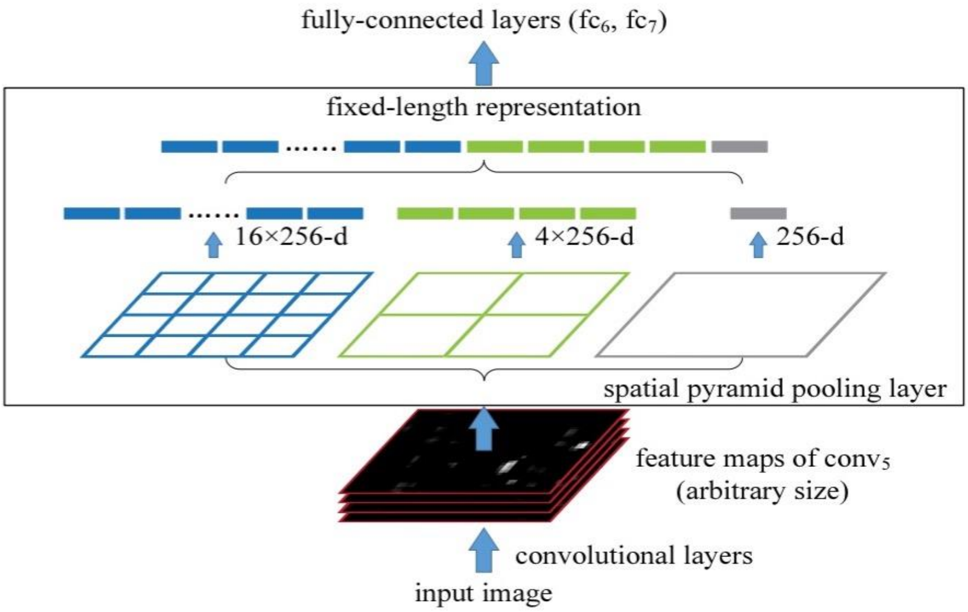

Figure 1.2. Network structure with an SPP layer. Here, 256 is filter number for $conv_5$ layer and it is the last layer.

1.4 Major Features

1.4.1 Lightweight:

The proposed model is suitable for mobile phones, tablets, and normal computers without worrying about the memory storage capacity. It fastens the screening process of the pneumonia disease so that better-targeted diagnoses can be performed to optimize the test time and cost. It comprises of ~0.86 million total number of parameters (estimated), which uses less than 4 MB memory storage.

1.4.2 Multiscale-ability:

The multiscale ability of the proposed network allows the model to identify different types of pneumonia disease for various severity levels.

1.4.3 Usage of Existent Methodologies:

The cheapest ways to screen the possibility of pneumonia is through a chest X-ray. With the help of an image of the chest X-ray of a patient, the model can accurately diagnose the possibility of pneumonia quickly and efficiently.

1.4.4 High Accuracy:

According to the performance results of related studies [7], the proposed light weight model should be capable of achieving best mean accuracy above 0.9.

1.4.5 14-layers of CNN with a modified SPP module:

Given the importance of having a good screening model, an accurate light weight deep learning network is proposed by embedding a modified spatial pyramid pooling (SPP) module to the convolutional neural network (CNN).