Optimal method to Detect Pneumonia in Chest X-Rays using deep learning Framework

Ajith R¹, Aravindha Raja G¹, Bahrul Ameen S¹, Kuppan MR¹, Reshma S²

¹Department of IT, Sri Krishna College of Technology
²Assistant Professor, Department of IT, Sri Krishna College of Technology

ABSTRACT

Physicians often use chest X-rays to quickly and cheaply diagnose disease associated with the area. However, it is much more difficult to make clinical diagnoses with chest X-rays than with other imaging modalities such as CT or MRI. With computer-aided diagnosis, physicians can make chest X-ray diagnoses more quickly and accurately. Pneumonia is often diagnosed with chest X-Rays and Kills around 50,000 people each year. With computer aided diagnosis of pneumonia specifically, physicians can more accurately and efficiently diagnose the disease. In this project, we hope to train a model using the dataset described below to help physicians in making diagnoses of pneumonia in chest X-Rays. Our problem is thus a binary classification where the inputs are chest X-ray images and the output is one of two classes: pneumonia or non-pneumonia.

KEYWORDS: Pneumonia, Chest X-Rays, Diagnosis and Binary Classification

1. INTRODUCTION

LUNG diseases are very often examined by using of screening methods. The better is the examination the higher are the chances to help patients in time and save their life’s. Lung cancer is one of the most dangerous diseases, therefore a rapid identification of nodules in the early stadium can reduce deaths. However in-depth analysis of diagnostic images requires large experience and skills. Nowadays due to the recent advancements of clinical and medial sciences, the clinicians dispose of many non-invasive techniques for oncological assessment and diagnosis that are generally preferred to potentially dangerous and more radical approaches (e.g. exploration surgery). The imaging diagnostic tools for medicinal applications are essentially founded on echography, x-rays, magnetic resonance imaging (MRI), computer tomography scan (CTS), and positron emission tomography (PET). X-ray imagery is a paramount exam for a correct diagnosis of potentially cancerous lung nodules. Such a diagnostic method is fundamental for radiologists and clinicians in order to select a correct treatment process. The examination of x-ray images is very important for many lung disorders and diseases even in an early stage, i.e. we can detect pre symptomatic changes in the structure of lung tissues. From X-ray images we can evaluate several pathological conditions, and various research present that even 90% of peripheral neoplasms and 65-70% of central neoplasms can be efficiently
diagnosed from X-ray images in the early stadium of disease. In any case unfortunately the most vulnerable factor is human, since many variables can influence the radiologist ability to detect the illness. The quality of the images can tamper with a clear visibility of all tissues as well as visualize the tissues as opaque or disformed. Sometimes several tissues can be placed in superimpositions with bones which can cover them. Unfortunately, the extreme variability of the matter and the subjective evolution of the illness in each patient can potentially tamper with the correctness of the diagnosis made by a radiologist. Therefore, to improve the accuracy of their diagnosis the medical specialists can benefit from intelligent methods of diagnostic support. Moreover, automated approaches to the process of analysis bring many advantages for faster and accurate diagnosis. Intelligent classification systems, and particularly convolutional neural networks (CNNs) and other deep learning techniques can be used as an efficient aid for lung cancer diagnosis and the consequent determination of a therapeutic approach. Among the possible applications of CNNs based classification techniques we can include solid tumours diagnosis and discrimination between malignant cancers with respect to benign nodules. Deep learning strategies can be partitioned into four classifications: CNN-based methods, restricted Boltzmann machines (RBMs), autoencoders, and Neuro-Fuzzy methods.

2. RELATED WORKS

The convolutional neural network (CNN) is proven to be very effective in image recognition and classification tasks. The development of CNNs starts from, LeNet, AlexNet [6], ZFNet [7], VGG [8], Inception [9] [10], ResNet [11], Inception-ResNet [12], Xception [13], DenseNet [14], and NASNet [15]. There are many studies on the use of deep CNNs to detect abnormalities in chest x-rays. For instance, M. T. Islam et al., [16] use several CNNs to detect abnormalities in chest x-rays. There is also a study by X. Wang et al., on the use of CNNs to detect thoracic pathologies from chest x-ray images. Their study also provides a large dataset as is the case in this study [17]. Among current research, some studies on the application of Densely Connected Convolutional Networks (DenseNet) [14] to detect thoracic pathologies such as ChexNet [18] and the Attention Guided Convolutional Neural Network (AG-CNN) [19]. Both studies train the neural network on a very large chest x-ray image dataset. Lung cancer prediction with CNN faces the problem of small sample size. Indeed, CNN contains a large number of parameters for adjustment on a large image dataset. In practice, researchers often pre-train CNNs on ImageNet, a standard image dataset containing more than one million images. The trained CNNs are then adjusted on a specific target image dataset. Unfortunately, the available lung cancer image dataset is too small for this transfer learning to be effective, even with a data augmentation trick.

3. PROPOSED FRAMEWORK

The chest X-ray is one of the most commonly accessible radiological examinations for screening and diagnosis of many lung diseases. A tremendous number of X-ray imaging studies accompanied by radiological reports are accumulated and stored in many modern hospitals’ Picture Archiving and Communication Systems (PACS). Machine learning has been used in various fields of medical industry to derive patterns from data. We demonstrate the effectiveness of this iterative refinement framework via extensive experimental evaluations on the publicly available data. And propose a deep learning framework using algorithms to extract knowledge from image data.
4. DATASET

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images and 2 categories (Pneumonia/Normal). Source: Mendeley. Chest X-ray images (anterior-posterior) were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children’s Medical Center, Guangzhou. Format: .JPEG. The diagnoses for the images were then graded by two expert physicians before being cleared for training the AI system.

5. CONVERT IMAGES TO NUMPY ARRAYS

Often in machine learning, we want to work with images as NumPy arrays of pixel data. This can be achieved using the imread () function that loads the image an array of pixels directly, the imshow () function that will display an array of pixels as an image. first loads the image and then reports the data type of the array.
we're applying a technique called Data Augmentation. We're changing the sizes of the images to 226 x 226 and we'll flip the images horizontally as well so that we can have more data(images) to train on. Data augmentation: to make slight variations to our data so that we have more data, without losing semantic meaning in our data. Data augmentation is a powerful technique which helps in almost every case for improving the robustness of a model. Augmentation can be much more helpful where the dataset is imbalanced.

6. DEEP LEARNING

A CNN structure uses a Feed-Forward Neural Network. CNN uses MLP, Multi-Layer Perceptron's, to do this convolutional process. Deep learning CNN techniques, became well known based on an outstanding, or winning performance, of Image Recognition. CNNs need a minimal amount of preprocessing. The convolutional layer uses multiple filters where each filter moves sequentially across the input data or image to make a 2-dimensional activation map based on each filter. Accuracy is one metric for evaluating classification models. The fraction of predictions our model got right.

6.1. RESULTS AND DISCUSSION

The simplicity of the proposed method is one of the main advantages. Moreover, it has obtained very good results. In further research we want to apply this methodology to other pulmonary diseases, which have similar symptoms. From our early research we can conclude that this kind of approach can be easily adjusted to other requirements due to the use of fuzzy rules without any loss in precision

<table>
<thead>
<tr>
<th>CLASSIFIER</th>
<th>ACCURACY</th>
<th>LOSS</th>
<th>Epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN model 1</td>
<td>87%</td>
<td>0.28</td>
<td>20</td>
</tr>
<tr>
<td>CNN model 2</td>
<td>93%</td>
<td>0.2</td>
<td>3</td>
</tr>
</tbody>
</table>

Table.1.1 Summary of Results

Fig.6. Accuracy of CNN

7. CONCLUSION

Data Analytics is the procedure of retrieve a pattern from large data set in connection with machine learning, data base, and statistics. Machine learning can be a very good help in deciding the line of treatment to be followed by extracting knowledge from such suitable databases. Our project can assist in proper treatment for Automated methods to detect and classify human diseases from medical images.
REFERENCES


