Detection of Pneumonia in Chest X-Ray Using Deep Learning

Koyel Roy[1], Karamjot Singh[2], Krishnakant Sawant[3], Aakash Sonawane[4]
[1.,2,3,4]Rajarshi Shahu College of Engineering, India

Abstract- *Pneumonia is a type of infection which occurs in the lungs. It is usually caused by viruses or bacteria. It is basically an inflammatory condition of lung which primarily affects small air sacs inside the lung. Physicians and doctors use chest X-rays and different types of tests to diagnose if person is suffering from the disease. We have proposed a system which will help doctors and physicians to check if the person is suffering from Pneumonia or not by taking chest X-ray as an input.We have implemented our system by using convolutional neural network (CNN) which is a class of Deep Learning Network used for image related tasks.*

Keywords- Convolutional Neural Network (CNN), Pneumonia, Deep Learning, Computer Vision

I. INTRODUCTION

Pneumonia affects more than 400 million people globally and results in about 4 million deaths per year. It is important to have an on-time diagnostic for this type of diseases. Medical X-Rays are generally used to diagnose some human body parts such as bones, chest, teeth and so on. By using this technique, medical experts have explored and diagnosed fractures in bones and abnormalities in body part. This has been used for several decades now, where medical experts see the x-Rays with their naked eyes and try to find the abnormalities. This is due to the fact that X-rays are very effective diagnostic tools in revealing the pathological alterations, in addition to its noninvasive characteristics and economic considerations. Chest diseases can be shown in CXR images in the form of cavitation, consolidations, infiltrates, blunted cost phrenic angles, and small broadly distributed nodules [1]. The advantages of chest X-rays include their low cost and are easily operable. Even in developing areas, modern digital radiography (DR) machines are very affordable. Therefore, chest x-Rays are widely used in the detection and diagnosis of the lung diseases[2].

In our model we will develop a system which will be a computer aided detection (CAD) system which can help doctors to detect the suspicious disease. The first attempt to establish a computer-aided detection system was in the 1960s, and studies have shown that the detection accuracy for the chest disease is improved with a X-ray CAD system as an assistant [2]. This system will use following methods: image preprocessing, identifying affected area, extracting features and classification based on features extracted. In image preprocessing system will identify the body part in the X-Ray image and labelling the parts of the image. Then system will identifying affected area and unaffected area separately. After identification of affected area system will extract the properties and features. Based on this features system will give result as if person is suffering from disease or not.

II. RELATED WORK

For detecting Pneumonia many methods like machine learning, data mining and deep learning have been used to implement. Accuracy of the model is based on which type of techniques are used and the size of dataset used to train the model. Different techniques give different accuracy results. When the outcomes of these techniques are compared, it is observed that deep neural networks model gives the best accuracy results for detecting the disease.

In previous work, researchers have used CNNs for training the model. CNN is class of a deep neural network, which is commonly used in analyzing visual imagery. CNN gives the best accuracy results. In one the work researchers have used single-center retrospective review of 35,038 adult posterior-anterior chest radiographs and final reports performed between 2005 and 2015 (56% men, average age of 56, patient type: 24% inpatient, 39% outpatient, 37% emergency department) with a waiver for informed consent[3]. The GoogLeNet CNN was trained using 3 graphics processing units to automatically classify radiographs as normal (n = 11,702) or into 1 or more of cardiomegaly (n = 9240), consolidation (n = 6788), pleural effusion (n = 7786), pulmonary edema (n = 1286), or pneumothorax (n = 1299)[3]. The network's performance was evaluated using receiver operating curve analysis on a test set of 2443 radiographs with the criterion standard being board-certified radiologist interpretation[3].

In another related work, researchers have used thresholding method. For detecting the presence of pneumonia clouds in chest X-rays (CXR) by using only Image processing techniques. For this, they have worked on 40 analog chest CXR spertaining to Normal and Pneumonia infected patients[4].Indigenous algorithms have been developed for cropping and for extraction of the lung region from the images[4]. To detect pneumonia clouds they have used Otsu thresholding which will segregate the healthy part of lung from the pneumonia infected cloudy regions, computing the ratio of area of healthy lung region to total lung region to establish a result[4].

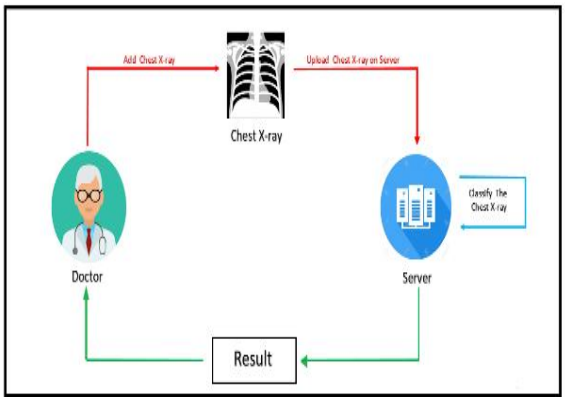
III. LITERATURE SURVEY

Different types of work have been done related to detecting lung diseases using machine learning. Various machine learning techniques are used and have achieved different accuracy for each techniques.

Thresholding methods have been used in related work to detect pneumonia. Image processing techniques have been used to segregate cloudy affected regions of the lungs from non-affected region. Different image processing techniques used were resizing in which the scanned chest X-ray images were resized to dimensions of pixel where the computational time and results given were good and then contrast enhancing functions were used to increase the intensity of cloudy regions. After this process identifying and cropping the lung region was done. Then thresholding method like Otsu thresholding was used to segregate healthy lung areas from affected areas. Then by naked eye we can detect if Pneumonia is present or not.

Another model developed an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists[5]. The algorithm, CheXNet, a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000frontalviewX-ray images with 14 diseases[5]. They have compared the performance of CheXNet with the test set which is annotated by four practicing academic radiologists.

IV. PROPOSED SYSTEM



In proposed system the model which is developed is a web application, deployed on the server. This application can be accessed by doctor to upload the scanned images and to get the result. The application will have a doctor login. Before login he has to sign up to the app with particular credentials. Only after login he can access the application. If patient comes to doctor, doctor can add information of patient to the app, like personal information, symptoms which he or she is suffering from. Then doctor after entering the information, can scan the chest X-ray of the patient and upload it to the app. After uploading the scanned image, the backend of the application (at server) will process the image. After processing the uploaded image, the application will give the result if the scanned chest X-ray image is healthy or is affected by Pneumonia. Then doctor can help patient with further treatment.

V. DATASET

We were grateful to have a public dataset available on Kaggle containing the Chest X-Rays of pneumonia affected lungs and healthy lungs. This helped in training the CNN and obtaining a good model.

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 (JPEG) and 2 categories (Pneumonia/Normal).

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. All chest X-ray imaging was performed as part of patients' routine clinical care.

For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans. The diagnoses

for the images were then graded by two expert physicians before being cleared for training the AI system.

Dataset	Normal	Pneumonia	Total
Training	1341	3875	5216
Validation	8	8	16
Test	234	390	624

VI. NETWORK ARCHITECTURE

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 148, 148, 32)      896
_________________________________________________________________
activation_1 (Activation)    (None, 148, 148, 32)      0
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 74, 74, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 72, 72, 32)        9248
_________________________________________________________________
activation_2 (Activation)    (None, 72, 72, 32)        0
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 36, 36, 32)        0
_________________________________________________________________
conv2d_3 (Conv2D)            (None, 34, 34, 64)        18496
_________________________________________________________________
activation_3 (Activation)    (None, 34, 34, 64)        0
_________________________________________________________________
max_pooling2d_3 (MaxPooling2 (None, 17, 17, 64)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 18496)             0
_________________________________________________________________
dense_1 (Dense)              (None, 64)                1183808
_________________________________________________________________
activation_4 (Activation)    (None, 64)                0
_________________________________________________________________
dropout_1 (Dropout)          (None, 64)                0
_________________________________________________________________
dense_2 (Dense)              (None, 1)                 65
_________________________________________________________________
activation_5 (Activation)    (None, 1)                 0
=================================================================
Total params: 1,212,513
Trainable params: 1,212,513
Non-trainable params: 0
_________________________________________________________________
```

The images are scaled down to 150 x 150 and fed as an input to the network. Reducing the image dimension results into information loss. On the other hand, processing high dimensional images is computationally expensive. The network consists of 3 convolution layers, 3 max pooling layers and 2 fully connected (FC) layers. All the convolution layers and the FC layers (except the last FC layer) have the activation function ReLU. Sigmoid activation function is applied to the last FC layer. Dropout of 0.5 is added between FC 1 and FC 2. Stride of (2,2) and kernel size of (2,2) is used in MaxPooling Layers to half the spatial dimension of the activation maps. Dropout is used to overcome the problem of overfitting.

VII . EXPERIMENTAL RESULTS

We predicted the accuracy of our model on the test dataset consisting of 624 chest x-ray images. The model provided with an accuracy of 81.25%.

VIII. CONCLUSION

Thus, we conducted a comparative analysis of various CNN architectures. We calculated the accuracy for all the architectures and the best architecture is selected. Hence, at the end it is predicted whether a patient is suffering from Pneumonia or not.

IX. ACKNOWLEDGEMENT

It gives us great pleasure in presenting this paper on ‘Detection of Pneumonia in Chest X-Rays Using Deep Learning’. We would like to take this opportunity to thank our guide Prof. Dr. P. B. Kumbharkar for giving us all the help

and guidance we needed. We are really grateful to him for his kind support. Valuable suggestions by him were very helpful.

REFERENCES

[1] "Deep Convolutional Neural Networks for Chest Diseases Detection", Rahib H. Abiyev and Mohammad KhaleelSallamMa'aitah.

[2] "Computer-aided detection in chest radiography based on artificial intelligence: a survey",Chunli Qin, Demin Yao, Yonghong Shi and Zhijian Song.

[3] "Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs", Cicero, Mark MD, BESc, et. Al.

[4] "Detection of Pneumonia clouds in Chest X-ray using Image processing approach", Abhishek Sharma, Daniel Raju, SutapaRanjan.

[5] "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rayswith Deep Learning", Pranav Rajpurkar, et. Al.

[6] http://cs231n.github.io/transfer-learning/

[7] http://cs231n.github.io/convolutional-networks/

[8] "A Survey on Transfer Learning", Sinno Jialin Pan and Qiang Yang, Fellow, IEEE.

[9] "Using Deep Learning for Image-Based Plant Disease Detection Sharada" – Prasanna Mohanty, David Hughes and Marcel Salathé.

[10] "Actively Transfer Domain Knowledge,",Proc. European Conf. Machine Learning and Knowledge Discovery in Databases, X. Shi, W. Fan, and J. Ren, Sept. 2008