ISSN 2395-1621

Chest Disease Detection from X-rays using CNN

Jincy Biju, Nikhil Padole, Vitthal Karad, Snehal Vibhute.

All India Shri Shivaji Memorial Society's Institute of Information Technology, Pune
Savitribai Phule Pune University, Pune

ABSTRACT

The world is changing so fast that the pressure on fitness is increasing, the terrible changes in weather, the surroundings, the lifestyles manner of human, also grows the threat as well as diseases for human beings. One of the them is chest related diseases. About millions of humans succumbed in 2018 to chronic obstructive pulmonary disease (copd), prompted especially due to smoking and pollutants, even as four hundred people died from bronchial asthma. With such a lot of chest related diseases humans can get, here is just one example of disease we will save if we discover them out earlier. With the help of technology, the sooner identity of illnesses, specially chest related problems, we may be able to hit upon in advance and greater appropriately, which could save many human lives as well as reduce the pressure at the system. The fitness gadget has no longer advanced in time with the improvement of the populace. With the increasing power of computer systems as well as the big quantity of facts being launched to the general public, this is a good time to contribute to solving this hassle. We hope that our solution can make a contribution to reducing clinical fees, the improvement of computer technology for clinical tasks. We are very fortunate to recognize that there may be a massive set of x-ray image records on Kaggle.

Keywords: x-rays, CNN, max pooling, image processing.

ARTICLE INFO

Article History

Received: 8th March 2020

Received in revised form :

8th March 2020

Accepted: 10th March 2020

Published online :

11th March 2020

I. INTRODUCTION

Deep learning is a machine learning method inspired by the deep structure of a human brain. The deep structures are characterized by multiple hidden layers allowing the abstraction of the different levels of the features [1][2]. One of the most powerful deep neural networks is the Convolutional Neural Network (CNN) that can include multiple hidden layers performing the extraction of features of the input data. Basically, the neural network consists of three layers: convolutional layers, subsampling or pooling layers and full connection layers [3]. CNN is the brand-new emerging trend in market place. Using CNN, a lot of useful applications can be developed. Mostly it is used in image classification, pattern recognitions, predictions, image recognition, etc [4]. It is used in a lot of medical applications which could be very useful for doctors as well as patients. As there is a lot of data available about various diseases, these could be used to predict various diseases in patients [5]. Using x-rays doctors could predict the disease in no time. And this will help in faster treatments of patients helping save time and decreasing the risk of life.

II. ALGORITHM USED: CNN

Convolutional Neural Networks (CNN) is a layered network of neurons used for pattern detection. CNN receives input and detects patterns through hidden layers [6]. In CNN, each input image will pass through a series of convolution layers with filters (kernels), pooling, fully connected layers (FC) and apply SoftMax function to classify an object with probabilistic values between 0 and 1. There are three main layers in CNN:

Convolution layer: In this layer, the entered photo of length r*c is convolved with a filter out of length a*a. Every block of the input matrix is independently convolved with the kernel and generated a pixel inside the output. The result of the convolution of the entered photo and filter is used to generate n output image functions also known as feature map. CNN can include multiple convolutional layers; the inputs and outputs of subsequent convolutional layers are the characteristic vector [7].

There's a group of n filters in every convolution layer. This layer can perform operations such as edge detection, blur and sharpen by applying various filters [8].

Sub-sampling layer: It is used to reduce the spatial dimensionality of the feature map extracted from the previous layer. It reduces the number of parameters when images are too large. Spatial pooling also called as subsampling or down sampling. It reduces dimensionality of each map but retains important information. Types of pooling layers are max pooling, average pooling and sum pooling.

Fully connected layer: The final layer of a CNN is a traditional feed ahead network with one or greater hidden layers. The matrix is flattened into vector and fed into FC layer.

Other terms used:

Stride: number of pixels shift over the input matrix.

Padding: sometimes filter does not fit the input matrix perfectly then padding is used. Padding can be done in two ways:

a. Pad pic with zero.

b. Drop the part of image where filter did not fit. Also known as valid padding.

Activation function: They determine the output of the neural network using mathematical equations. Softmax function or sigmoid function is used to classify the output.

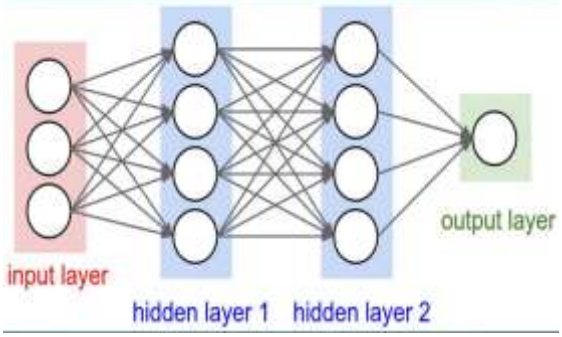

Fig. 2.1 Convolutional Neural Network Layers

III. SYSTEM ARCHITECTURE

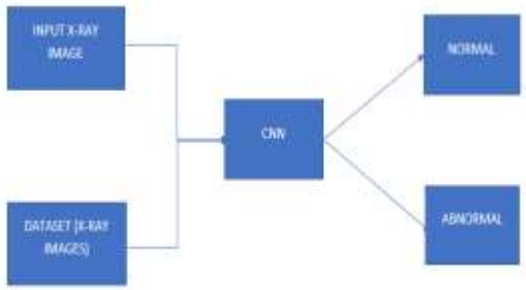

Fig. 3.1 System Architecture

Our system will predict which chest disease is the patient suffering from, with the help of the patient's x-ray. CNN model will be trained using the x-ray dataset and input will be the x-ray image. Here we will divide the problem into many small problems. In this project, we will diagnose whether the patient is ill and what disease is it. That is, there are two small problems: The binary classification problem with two classes is normal and abnormal and the binary classification problem for each disease is the corresponding disease or not. We will use image processing techniques to process data, we can also use Spatial Transform to get the main information from the image [9]. As we know that the traditional CNN network is very powerful in image processing [10], so we will first test to see if CNN is good for this dataset or not. For this project, we will try to achieve accuracy 70 % of the binary classification problem that is disease or not, and we will try to continue to upgrade the accuracy of the algorithm. We will split the dataset into training data, validation data and testing data. In the sample dataset, we will divide the set into 60 % training data and 20 % validation data and 20 % testing data because this is a small dataset with 5,606 sample x-ray images, and full dataset have proportionally 80 % training data and 10 % validation set and 10 % testing data because this is a huge set of 1,12,120 samples. Training the CNN model then test the CNN with the original architecture to make sure CNN is detecting the patterns of the diseases, then accelerate the convergence by using a pre-trained model.

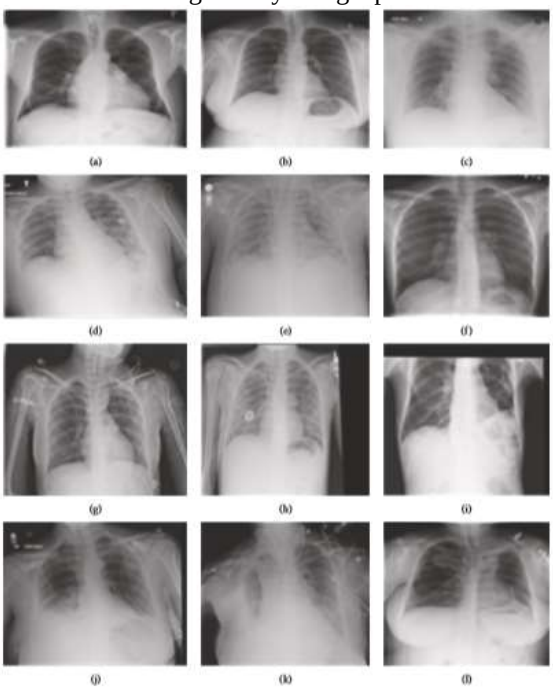

Fig. 3.1 Chest diseases x-ray image samples

Evaluation Matrix Model will be based on the indexes specified below-

TP: True Positive - The number of people affected is expected to be affected

FP: False Positive - The number of people who are sick is predicted to be unwell

FN: False Negative - The number of people without the disease is predicted to be wrong

Precision=TP/(TP+FP)

Recall=TP/(TP+FN)

Precision represents the number of people whose disease is positively predicted in the total number of people who were predicted to be infected. Recall represents the number of people who correctly predicted illness on the total number of people actually infected. Both of these indicators are very important in predicting the disease, and we need an index that can be both Precision and Recall. From precision and recall, we can calculate F score.

IV. EXPERIMENTAL RESULTS

We have proposed a model that will detect chest diseases from a patient's chest x-ray image.

Table 1

Chest Diseases	Percentage
Atelectasis	9.06
Cardiomegaly	2.515
Effusion	11.487
Infiltration	17.25
Mass	12.30
Nodule	12.89
Edema	2.104
Pneumothorax	4.83
Fibrosis	1.49
Pneumonia	1.105
Emphysema	2.265
Consolidation	4.03
No disease	55

V. LIMITATIONS

a. Limited dataset used.
b. Requires lot of processing time.
c. Only 14 diseases will be detected.
d. Newly discovered diseases won't be detected.

VI. APPLICATIONS

a. Early diagnosis of chest diseases.
b. As diseases are detected faster than the traditional approach, which helps in lowering the risk of patient's life.
c. No need to carry x-rays to the doctor.

VII.CONCLUSION AND FUTURE WORK

The end result will be showing the patient's health condition i.e. whether he/she is suffering from any chest related disease or not. This will help doctors save time in detecting the abnormality and start proper treatment for the patient by lowering the risk of further complexities. In this proposed work, the specificity obtained will be 100 percent which shows that that there is no false positive detection. Also, the accuracy, availability and reliability of the proposed system is high when compared to previously available conventional neural network-based systems.
In Future, this system could be modified for detection of more diseases. The accuracy of the system could be improved.

REFERENCES

[1] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, R. M. (2017). Chestxray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, pages 3462–3471. IEEE.
[2] Y.LeCun, Y.Bengio, and G.Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[3] Deep Learning Based Medical X-ray Image Recognition and Classification Department of Computer Science and Engineering BRAC University, Dhaka 1212, Bangladesh December, 2018.
[4] Andrew Ward, Nicholas Bambos. Quantum Annealing Assisted Deep Learning for Lung Cancer Detection.
[5] Deep Convolutional Neural Networks for Chest Diseases Detection Rahib H. Abiyev and Mohammad Khaleel Sallam Ma'aitah Depar Turkey Correspondence should be addressed to Mohammad Khaleel Sallam Ma'aitah; mohammad. maaitah@neu.edu. Published 1 August 2018.
[6] Lung Cancer Detection and Classification Using Deep CNN: S. Sasikala, M. Bharathi, B. R. Sowmiya International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-2S December, 2018.
[7] Automatic diagnostics of tuberculosis using convolutional neural networks analysis of MODS digital images: Lopez-Garnier S, Sheen P, Zimic M (2019) Automatic diagnostics of tuberculosis using convolutional neural networks analysis of MODS digital images. PLoS ONE 14(2): e0212094.
[8] A. R. Abas, "Adaptive competitive learning neural networks," Egyptian Informatics Journal, vol.14, no.3, pp.183–194,2013.
[9] Full Resolution Image Compression with Recurrent Neural Networks: George Toderici, Damien Vincent, Nick Johnston, Sing Jin Hwang, David Minnen, Joel Shor, Michele Coveli IEEE Conference Paper.
[10] Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers' Health Examination Data Seok-Jae Heo 1, Yangwook Kim, Sehyun Yun , Sung-Shil Lim , Jihyun Kim, Chung-Mo Nam 1,3, Eun-Cheol Park, Inkyung Jung, and Jin-Ha Yoon Received: 27 November 2018; Accepted: 9 January 2019; Published: 16 January 2019.

