© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 03 | NCASIT- 2022 | April 18th 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Refereed & Peer Reviewed

LUNG DISEASE DETECTION FROM X-RAYS USING CNN

Nischint Makode
Dept. of Computer Engineering,
SVPCET, Nagpur, MH, India

Aditya Dabhade
Dept. of Computer Engineering,
SVPCET, Nagpur, MH, India

Onam Dumbhare
Dept. of Computer Engineering,
SVPCET, Nagpur, MH, India

Abstract—**Lung diseases are becoming more commonplace throughout the world. Some of the major diseases include chronic obstructive pulmonary disease, pneumonia, asthma, tuberculosis, fibrosis, etc. A multitude of distinct image processing and machine learning models have been developed for this cause. Various types of deep learning methods including convolutional neural network (CNN), vanilla neural network, capsule network and visual geometry group based neural network (VGG) have been implemented for lung disease diagnosis. For implementation of the research, Jupyter Notebook, Tensorflow, OpenCV, and Keras are utilized. The model is applied to NIH chest x-ray image dataset obtained from the Kaggle repository. Complete and sample editions of the dataset are kept in view. For the use of full dataset, CNN exhibits a validation accuracy of 90%. Whereas the use of sample dataset yields a much lower training time at the cost of a slightly less validation accuracy. Thus, the proposed CNN framework will make the diagnosis of lung diseases an easy task for medical practitioners as well as for experts.**

I. *KEYWORDS- CNN, X-RAY, LUNG DISEASE*

1. INTRODUCTION

Health effects of diseases are drastically spiking due to changes in environment, lifestyle, weather conditions and other components. The susceptibility of catching lung diseases is huge, specifically in developing countries, where people are affected by poverty. It has now become a necessity that an effective and efficient diagnostic system should be developed for aiding in detection of lung diseases as the number of premature deaths are on a rapid increase due to diseases like asthma, pneumonia etc. [1, 2]

A novel coronavirus disease 2019 (COVID-19) has been responsible for causing severe damage to lungs and problems related to respiration since December 2019. Therefore, the need for early diagnosis has never been felt more important than now. This can be achieved by using the principles of in deep learning and machine learning. [3] Through this research paper, a general direction can be provided to medical practitioners and research personnel for using deep learning principles for detecting lung diseases. For achieving this, a dataset with large number of lung x-ray images can be used. This system can aid in accurate diagnosis of lung diseases, which can save a lot of endangered people and in turn reduce the disease rate.

For the development part, the NIH chest x-ray image dataset is obtained from the Kaggle repository which is a completely open-source platform. An algorithm is established in this paper, and it is successfully implemented for the dataset mentioned above for the classification of lung diseases. [4, 5]

This paper is sectioned as follows: Section 2 gives a description of a few related works on lung disease classification and detection. A brief analysis of the used dataset is given in Section 3. The implementation of this research is described in Section 4. Section 5 concludes the paper by providing the appropriate results and discussion related to the paper.

2. RELATED WORK

A multitude of research propose a lot of deep learning models that were used for the detection of lung cancer and other lung diseases. [6-11] Ref. [6] emphasises on the detection of thorax diseases. Ref. [7] proposes a 3D deep learning CNN with multiscale prediction strategies for detecting the lung nodules from segmented images. But it cannot classify diseases according to their types and the multiscale prediction approaches are used for small nodules. Ref. [8] defines a fully CNN for the reduction of false positive rate in classification of the lung nodules. A much more effective R-CNN is given in Ref. [9] for diagnosis of affected lung nodules. The R-CNN gives promising results for object detection. The cumulation of deep CNN architecture and dual path network (DPN) is proposed in Ref. [10] for classification and extraction of features of the nodules. The importance of artificial intelligence (AI) is given in Ref. [11] including a state of art in the classification of chest x-ray and diagnostic analysis.

3.DESCRIPTION OF X-RAY IMAGE DATASET USED.

This section includes various aspects of the dataset which include description, exploration, visualization, and view position of the data samples. These are elaborated as follows.

Dataset description.The dataset sample [8] file used consists of a random sample (5%) of the complete dataset. It contains 5606 images. Each image in the dataset has a resolution: 1024 X 1024. Also, it consists of a comma separated values (.csv) file which contains the patient data and class labels for the dataset sample being used.

Following are the15 classes of which one is “No findings” and another 14 diseases: Atelectasis - 508 images, Pneumonia – 62 images, Hernia - 13 images, Edema - 118 images, Emphysema - 127 images, Cardiomegaly - 141 images, Fibrosis - 84 images, Pneumothorax - 271 images, Consolidation - 226 images, Pleural Thickening - 176 images, Mass - 284 images, Effusion - 644 images, Nodule - 313 images, Infiltration - 967 images, No Finding - 3044 images

Dataset exploration. Following are the patient data and class labels of the complete dataset. Patient ID Finding labels such as disease type, Image index, View position: X-ray orientation, Patient gender, Patient age, Original Image Height, Original Image Width, Original Image Pixel Spacing_x, Follow-up, Original Image Pixel Spacing_y

The data contains important records for the set of data constructed as: gender, age, snapshot data, view position as well as lung X-ray images. We will use this key information to train the CNN model. View position.

Posterior-anterior (PA) position: It is a standard position used for locating a normal mature chest radiograph. Patient erects standing with the anterior position of chest implemented alongside the anterior of the film.

Anterior-posterior (AP) position: It is conducted when the patient is immobilized, debilitated, or incapable to cooperate with the PA process.

4. IMPLEMENTATION

Data pre-processing.

The dataset is a collection of numerous chest x-ray images. Furthermore, few additional information such as gender and age distribution can be acquired from the dataset. The pre-processing that was carried out is mentioned as follows.

For images: Initially, rescale all the images for making the size smaller which leads to faster training stage. Then, every image is transformed to RGB and gray. Finally, numpy arrays are used for reading the images which are normalized by isolating the image matrix using 255.

DFD & its explanation.

Install all the Required Libraries and Packages.

- Loading the required dataset.
- Classification of diseases.
- Split the dataset into training and testing part.
- Initialization and creation of model using CNN.
- Train the model.
- Save the model in the working directory.
- Load the saved model for disease detection.
- Using model to determine disease for any file in the dataset.
- Display disease probability percentage for any image using the model.

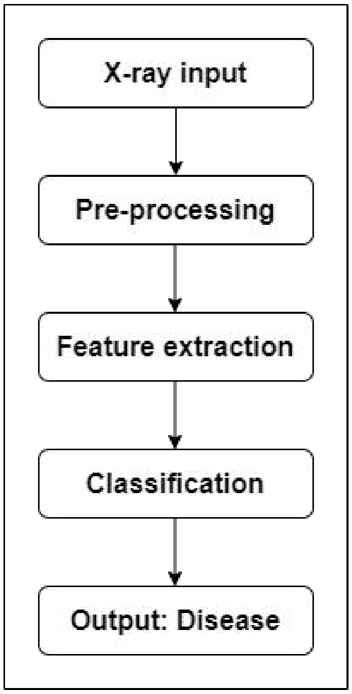

Fig 1. Data flow diagram representing the workflow.

After the use of convolution, nonlinear and pooling layers, it is important to follow them by a fully connected layer. This layer gives an N dimensional vector, where N is the number of classes from which the model selects the desired class [12].

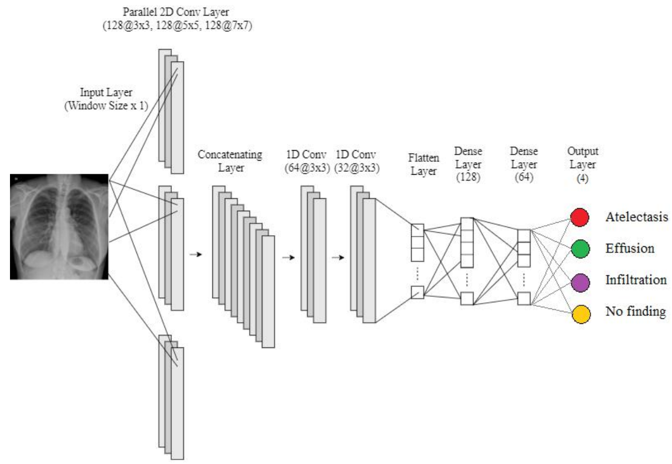

Fig 2. Convolutional neural network (CNN) architecture.

4.3. Architecture of CNN & its explanation.

Convolutional neural networks (CNN) use some features of the visual cortex. This enables CNN to be used for image classification. The main objective of image classification is accepting an input image and the following definition of its class [12].

In additional depth: the image is passed through a series of convolutional, nonlinear, pooling layers and fully connected layers, and then generates the output [12]. The Convolution layer is the first layer. This layer accepts the image matrix and converts it into a smaller input matrix by choosing filters which produce convolution by summing the multiplications made by the filter.

The nonlinear layer is used after every convolution operation. It consists of an activation function, which gives the nonlinear property which makes the network sufficiently intense and makes it able to model the response variable.

The pooling layer comes after the nonlinear layer. It works with the width and height of the image and performs down sampling operation on them. This reduces the image volume. This helps in getting rid of large size detailed images which are no longer needed for further processing.

5. RESULTS

As a result of testing the model, an accuracy of 90% was achieved of correct classification samples after 30 epochs. The only limitation was the long waiting time for the 30 epochs to come to an end. For portraying this, two plots were built. The first shows dependence of accuracy on the number of epochs. The second shows dependence of loss on the number of epochs.

6. CONCLUSION

In this work, we figured out the use of CNN for image classification. A CNN model was created and trained to classify chest x-ray images. This model works very efficiently even with a small number of images. The dependence of accuracy on the number of epochs were measured. Our future scope is to experiment with other different neural network designs to assess how a higher accuracy can be obtained from them.

7. ACKNOWLEDGEMENT

The authors would like to thank the National Institute of Health (NIH) for uploading their datasets in the Kaggle repository.

8. REFRENCES

[1] Bharati S, Podder P, Mondal R, Mahmood A, Raihan-Al-Masud M. Comparative performance analysis of different classification algorithm for the purpose of prediction of lung cancer. Advances in intelligent systems and computing, vol. 941. Springer; 2020. p. 44757. https://doi.org/10.1007/978-3-030-16660-1_44.

[2] Coudray N, Ocampo PS, Sakellaropoulos T, et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med 2018;24:1559–67. https://doi.org/10.1038/s41591-018-0177-5.

[3] Mondal MRH, Bharati S, Podder P, Podder P. "Data analytics for novel coronavirus disease", informatics in medicine unlocked, 20. Elsevier; 2020. p. 100374. https://doi.org/10.1016/j.imu.2020.100374.

[4] NIH sample Chest X-rays dataset. https://www.kaggle.com/nih-chest-xrays/sample. [Accessed 28 June 2020].

[5] NIH full Chest X-rays dataset. https://www.kaggle.com/nih-chest-xrays/data. [Accessed 28 June 2020].

[6] Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. ChestX-Ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: 2017 IEEE Conference on computer Vision and pattern recognition (CVPR); 2017. p. 3462–71. https://doi.org/10.1109/CVPR.2017.369.

[7] Gu Y, Lu X, Yang L, Zhang B, Yu D, Zhao Y, Gao L, Wu L, Zhou T. Automatic lung nodule detection using a 3D deep convolutional neural network combined with a multi-scale prediction strategy in chest CTs. Comput Biol Med 2018;103:220–31.

[8] Setio AAA, Traverso A, de Bel T, Berens MSN, van den Bogaard C, Cerello P, Chen H, Dou Q, Fantacci ME, Geurts B, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge. Med Image Anal 2017;42:1–13.

[9] Zhu W, Liu C, Fan W, Xie X DeepLung. Deep 3D dual path nets for automated pulmonary nodule detection and classification. In: Proceedings of the IEEE winter conference on applications of computer vision (WACV); 12–15 March 2018. p. 673–81. Lake Tahoe, NV, USA.

[10] Kong W, et al. YOLOv3-DPFIN: a dual-path feature fusion neural network for robust real-time sonar target detection. IEEE Sensor J 1 April 2020;20(7):3745–56. https://doi.org/10.1109/JSEN.2019.2960796.

[11] Kallianos K, Mongan J, Antani S, et al. How far have we come? Artificial intelligence for chest radiograph interpretation. Clin Radiol 2019;74(5):338–45. https://doi.org/10.1016/j.crad.2018.12.015.

[12] Ksenia Sorokina. Image Classification with Convolutional Neural Networks. https://medium.com/@ksusorokina/image-classification-with-convolutional-neural-networks-496815db12a8.