ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 42-47

Article Title: An Analysis of Lung Disease Detection Using Deep Learning

An Analysis of Lung Disease Detection Using Deep Learning

A. Angel Mary[1], K. K. Thanammal[2]

[1] Research Scholar, Department of Computer Science and Research Centre, S. T. Hindu College, Nagercoil, Tamil Nadu, India.

[2] Associate Professor, Department of Computer Science and Research Centre, S. T. Hindu College, Nagercoil, Tamil Nadu, India.

ABSTRACT

Learning about Lung Diseases and their characterization is one of the most interesting research topics in recent years. With the various uses of medical images in hospitals, pathologies, and diagnostic centres, the size of the medical image datasets is also expanding expeditiously to capture the diseases in hospitals. Though a lot of research has been done on this particular topic still this field is confusing and challenging Deep learning techniques have recently achieved an impressive result in the field of computer vision along with Medical Engineering. In this paper, we proposed and evaluated a deep convolutional neural network, designed for classifying the Chest Diseases. The proposed model consists of Convolutional layers, ReLU Activations, Pooling layer, and fully connected layer. Last full connected layer which consists of fifteen output units. Each output unit will predict the probability of one of the fifteen diseases. The stacked ensemble learning classifier contains random forest and SVM in the first stage and logistic regression in the second stage for lung disease detection. The performance of the proposed method is studied in detail for more than one lung disease such as pneumonia, Tuberculosis (TB), and COVID-19. The performances of the proposed method for lung disease detection using chest X-rays compared with similar methods with the aim to show that the method is robust and has the capability to achieve better performances. In all the experiments on lung disease, the proposed method showed better performance and outperformed similar lung disease existing methods. This indicates that the proposed method is robust and generalizable on unseen chest X-rays data samples. To ensure that the features learnt by the proposed method is optimal, t-SNE feature visualization was shown on all three lung disease models.

Keywords: Artificial Intelligence & Neural Networks, Chest X-ray, Dataset Analysing with Caps Net, Deep learning, Lung Disease Classification, Multichannel, ReLU Modelling.

1 Introduction

This project conduct a study and analysis of data set, then apply Machine Learning and Deep Learning to predict that the patient has a lung disease. This project is a binary classification with input is patient's data (X-ray images, View Position) and output is found for diseases or not. The difficulty is a new dataset, and this will be one of the pioneers to learn it, analysis is a large dataset but has never been processed full, data has a lot of noise, and X-ray of the lung is

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 42-47

not likely to provide enough information to assess whether a patient may be ill. It will use Machine Learning as well as Deep Learning to process data as well as create models for diagnosing patients. My key point here will be: combining the processing of patient information with data from X-rays, using CNN with the well-known pre-trained model, first time using the Caps Net [910] network for data in this form. According to Science, an increase in average global temperature of two degrees had posed much danger to humans. People with Asthma, Chronic obstructive pulmonary disease (COPD), and lung cancer are more prone to climate change. COPD was the third most significant cause of mortality globally in 2019, accounting for around 3.23 million fatalities. A study has shown that heat waves caused due to climate change result in an increase of respiratory diseases among children. Some studies also relate the cyclone to rise in respiratory diseases. Climate changes also causes change in the plant flowering season leading to extended pollen seasons and in turn leading to more human exposure. Pollinosis patients suffer from asthma caused by the thunderstorms occurring during the pollen seasons. Several studies have demonstrated effects of ozone over respiratory symptoms, including shortness of breath, lower respiratory tract infections and acute and transient decreases in lung function. Pneumonia can be widely classified into two types: Bacterial and Viral. Bacterial pneumonia has more dangerous symptoms than viral pneumonia and needs antibiotic therapy for recovery. Scientists have identified that one of the main reasons for pneumonia is the increasing levels of air pollution. Studies show that in most countries, the disease is overlooked in the case of elderly patients and is untreated until it reaches a deathly point. Researchers feel that there is a pressing need to develop precise computer-aided diagnosis systems to assist in diagnosing the disease at an early stage, especially in the case of children. There are many tests in the market to diagnose pneumonia, such as chest X-rays (CXR) and chest magnetic resonance imaging (MRI).

2 Background

The world is changing so fast that the pressure on health is increasing, the bad changes of climate, the environment, the life way of human, ... also increase the risk as well as diseases for people. One of the issues that we will focus on in this article is lung diseases. About 3.2 million people succumbed in 2015 to chronic obstructive pulmonary disease (COPD), caused mainly by smoking and pollution, while 400,000 people died from asthma. With so many lung diseases people can get, here is just one example of diseases we can save if we find them out earlier. With the technology machine and computer power, the earlier identification of diseases, particularly lung disease, we can be helped to detect earlier and more accurately, which can save many people as well as reduce the pressure on the system. The health system has not developed in time with the development of the population.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 42-47

Article Title: **An Analysis of Lung Disease Detection Using Deep Learning**

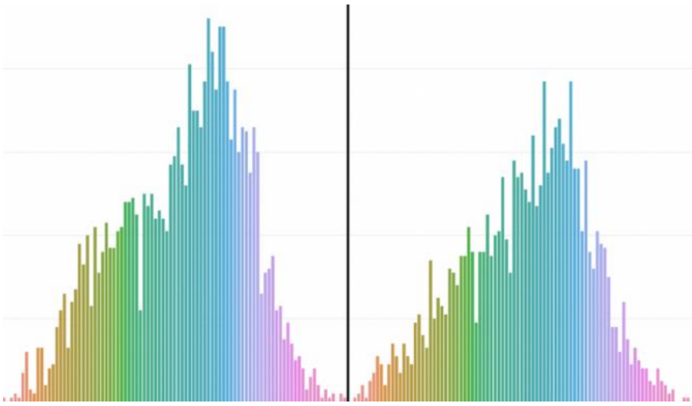

Figure 1: *Sample Dataset*

The detailed surveys on lung disease detection are studied and reported. There are many machine learning and deep learning approaches along with image processing employed for lung disease detection and classification. Overall, ImageNet-based fine-tuned models showed better performances. The literature has several studies on deep learning to identify lung illness. Pneumonia is one of the lung disease types that has led to many deaths among children. Computer-assisted diagnostics methods have shown promise in terms of increasing diagnostic accuracy and have reduced the time consumed in analysing the chest X-rays by the doctors.

(a) Normal (b) Pneumonia (c) COVD-19

Figure 2: *Chest X-Ray images of affected lungs*

3 Methodology

The dataset consists of many X-ray images. Moreover, some additional information such as age or gender distribution can be obtained from the dataset. The pre-processing steps used in this work are mentioned in the following.

For images:

- ❖ At first rescale all images for the purpose of reducing size leading to faster training stage.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 42-47

- All the images are transformed to RGB and gray, and are mutually conducted for various models.
- The numpy array uses for reading the images at that time is normalized by separating the image matrix using 255.

For additional information:

- Redefine some of the specific features.
- Normalize the age field to the numeric system then along with the year, at that time normalization field.
- Eliminate the outliers in the age attribute.
- There are two essential attributes, this paper will conduct as 'view position' and 'patient gender' in indiscriminate both datasets

In this work, the Caps Net from the main Hinton architecture is modified to make it fit for the lung image dataset. A basic Caps Net architecture for lung X-ray images analysis.

Main portions of this model can be summarized as follows.

- Convolution layer with filters = 256, strides = 2, kernel_size = 9, activation = 'relu', padding = 'same'. This layer was improved as of the original classifier from strides = 1 to strides = 2, the image was 28 × 28, as well as the data was 64 × 64, the output of this classifier will be considerably compacted. With strides = 2, we will acquire less features than strides = 1, subsequently we have improved the strings, consequently we consider that the output of lung images have been considerably concentrated.
- Primary capsule with dim_capsule = 8, strides = 2, kernel_size = 9, n_channels = 32, padding = 'same', simply variations with Hinton's structure in which the padding 'valid' is exchanged with 'same'.
- Diagnosis capsule (we change the similar name in which Hinton situates) with n_class = num_capsule, dim_capsule = 16, stable of the set routings.

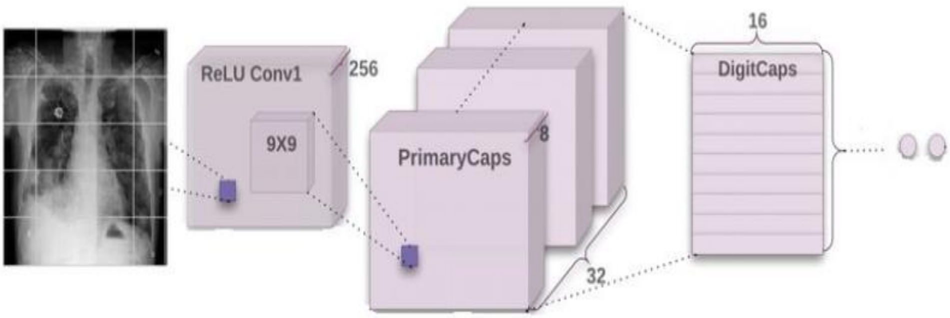

Figure 3: *Network for Lung X-Ray Images Prediction.*

The evaluation metrics used here will be precision, recall and F-beta scores (beta is 0.5) for binary classification – found diseases or not. In this case F score is better than accuracy because

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 42-47

with binary classification found diseases or not, the classes are imbalanced. For example, consider you have a trivial classifier that just guesses the majority class, it will obtain 80% accuracy when there is an 80/20 split and 50% accuracy when there is a 50/50 split. These indexes will be evaluated on a separate testing data set from the original dataset. These indicators will be evaluated for all diseases – found disease or not. If the case is positive and not negative, then the indicators: (1) True Positive - The number of people affected is expected to be affected. (2) False Positive - The number of people who are sick is predicted to be unwell. (3) False Negative - The number of people without the disease is predicted to be wrong. Optimized Convolutional Neural network and Grey Wolf Optimization algorithm. A confusion matrix (also called an error matrix) is a table that is used to describe the performance of a classifier. It provides the number of true positives, true negatives, false positives, and false negatives. – True positive (TP): lung disease chest x-ray identified as lung disease chest x-ray. – True negative (TN): healthy chest x-ray identified as healthy chest x-ray. – False positive (FP): healthy chest x-ray misclassified as lung disease chest x-ray. – False negative (FN): lung disease chest x-ray misclassified as healthy chest x-ray.

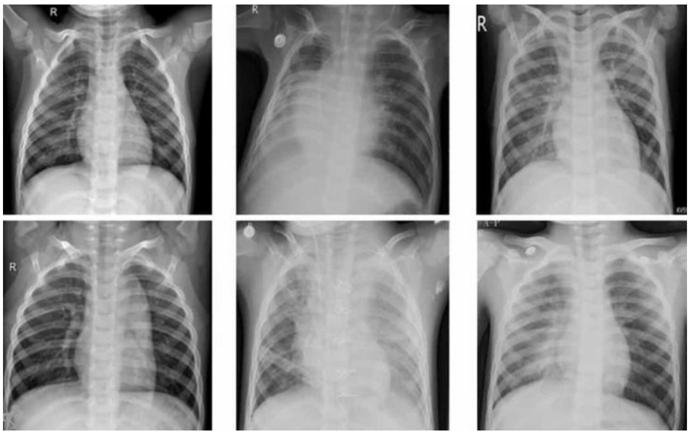

Figure 4: *Classifies Pneumonia on Chest X-Ray*

In the case of weighted-average, each class has a contribution to the final average that is weighted by its size. Both macro and weighted metrics compute the precision, recall, and F1-score for each class in the dataset and return the average. The comparison studies, shows that Vinayakumar Ravi technology is best.

4 Conclusion

During development, a validation set was used to evaluate the model. The following is a change chart of loss in training algorithms corresponding to sample dataset and full dataset Sample dataset: CNN + VGG + data + STN. From the above three charts, it is easy to see that the vanilla CNN is running the worst, it overfilled very early and stopped because of my Early Stopping checkpoint model. CapsNet seems to be working but convergence is too slow. This

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
pp. 42-47

Article Title: An Analysis of Lung Disease Detection Using Deep Learning

work has proposed a framework to classify lung diseases from chest X-rays. The framework leverages multichannel EfficientNet deep learning-based stacking ensemble approach. EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 models serve as feature extractor and followed by the features of EfficientNet models are concatenated. The concatenated features are passed into more one fully connected layer to learn the optimal features and a stacked ensemble approach is used for classification. It uses random forest and support vector machine in the first stage for prediction and followed logistic regression in the second stage for classification. The proposed model has performed better than the EfficientNet models such as EfficientNet-B0, EfficientNet-B1, and EfficientNet-B2 for pediatric pneumonia, COVID-19, TB lung disease classification.

Reference

1. Year: 2021, “CDC—data and statistics—chronic obstructive pulmonary disease (COPD)”.

2. Zhiwei Xu; Perry E. Sheffield; Wenbiao Hu; Hong Su; Weiwei Yu; Xin Qi; Shilu Tong, Year: 2012, “Climate change and children’s health-A call for research on what works to protect children”, Int. J. Environ. Res. Public Health, Vol: 9, no: 9, pp. 3298 – 3316.

3. Gennaro D'Amato; Lorenzo Cecchi; Gennaro Picardie, Year: 2008, “Thunderstorm-related asthma: not only grass pollen and spores”, J. Allergy Clin. Immunol, Vol: 121, no: 2, pp. 537 – 538.

4. Kingsley Kuan; Mathieu Ravaut; Gaurav Manek; Huiling Chen; Jie Lin; Babar Nazir; Cen Chen; Tse Chiang Howe; Zeng Zeng; Vijay Chandrasekhar, Year: 2017, “Deep Learning for Lung Cancer Detection: Tackling the Kaggle Data Science Bowl 2017 Challenge”.

5. Andrew Ward; Nicholas Bambos, “Quantum Annealing Assisted Deep Learning for Lung Cancer Detection”.