Pneumonia and Hantavirus Detection UsingCNN

Harsh Bhattad[1], Hemant Deshmukh[1], Nawazish Bharwani, Vaishnavi Vyawhare[1], Ms. Sujata Asabe[2]

[1] UG Students, [2] Assistant Professors, Computer Engineering Department, G H Raisoni College ofEngineering and Management, Pune

ABSTRACT

Pneumonia and Hantavirus are the most threatening diseases due to which many people die. There are many reasons behind this but the main reason is that people don't get to know about it in early stages due to lack of experts and medical diagnostic facilities. This research paper focuses on developing a diagnostic system for prediction of hanta virus and pneumonia by using the concept of image processing in conjunction with machine learning. Proposed system will detect the disease from X-Ray images and use preprocessing technique that will remove the noise and disturbance in image. Feature extraction process is applied to extract the useful features of underlying image, and feature selection technique will further optimize the top ranking features. CNN algorithm is then applied to classify the images for detection of respiratory diseases. After detection of disease, report will be generated and submitted to patient.

Keywords— CNN, Machine Learning, Image Processing, Feature Extraction

I. INTRODUCTION

Every year millions of people die due to lung diseases. Hanta virus is one of such disease. Hantavirus spreads due to rodents. Hantavirus is not a common disease because it cannot spread from one human being to another. During the early 1950s cluster of 3000 United Nation troops stationed in Korea were struck by an unknown illness. This was the first time Hantavirus was recognized as an infectious disease. Ten to fifteen percent of the infected people perished and the exact reason was not discovered for more than two decades. Etiologic agents were not discovered but however rodents were considered the main epidemiologic vector. Hantavirus infection was associated with fever, hypotension, renal fever. Earlier it was known as Korean hemorrhagic fever but later it was clinically termed as hemorrhagic fever with renal syndrome(HFRS) and the virus was named as Hanta after the Hantan River of Korea. Over the succeeding years many agents of HFRS were discovered across Asia and Europe.

Pneumonia is another such disease which is very common. It starts as a bacterial, viral or fungal infection. Due to this the alveoli(air-sacs in lungs) gets filled with fluid. The main job of alveoli is to give oxygen to the body(via blood) and take out carbon-dioxide form body. Since the alveoli is filled with fluid our body does not gets enough oxygen to keep it running and this may result is severe illness or even death.

There are many ways by which we can detect pneumonia like Blood culture test, Sputum test, Chest X-Rays and many more. Currently we are using X-Rays and CT Scan to detect lung diseases as they are more cheaper due to technological advances in the bio-medical field. As these images are scans of internal body common man cannot read them. We need experts to read and give accurate results. Using the concepts of machine learning we are proposing to make a system which will detect Pneumonia and Hantavirus from X-Ray images. Our system will take X-Ray image of the lung as an input and generate a report giving the output whether the lung is infectedor not.

The main key in this whole system is the algorithm we are using. CNN has originally been designed for image analysis. It also has an excellent capacity in sequent analysis. The convolution operation uses multiple filters to extract features from the dataset

II. LITERATURE SURVEY

This section of the literature survey reveals some facts based on thoughtful analysis of many authors work as follows.

[1] Zech, John R., Marcus A. Badgeley, Manway Liu, Anthony B. Costa,Joseph J. Titano, and Eric Karl Oermann. "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study." PLoS medicine 15, no. 11 (2018): e1002683.

In this system Authors have used supervised deep learning for detection of pneumonia.The X-Ray image is used as an input to the system and then the system detects the presence of infections and virus using CNN algorithm (DenseNet)Resnet-50. This model detects the virus present in the patient's body in the early stage which will be helpful to start the medications of the Patient to prevent the spreading of the virus in the vital areas of the Lungs. PyTorch0.2.0 is used in this system. The Accuracy of this model is determined as 0.931 , and the dataset used is NH(chest Xray 14) , IU(Open-I) , MSH(mount Sinai hospital). The advantage of this model is it is very accurate and can be used in many hospitals. The Limitationsof this model is CNN does not perform on the external data as compared to internal data.

[2] Rajaraman, Sivaramakrishnan, Sema Candemir, Incheol Kim, George Thoma, and Sameer Antani. "Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs." Applied Sciences 8, no. 10 (2018): 1715.

The dataset used in this model is chestX-Ray 14. The authors have proposed another method for detection of pneumonia. The method of Data collection and Preprocessing Is used in this system for the virus detection in pediatric chest radiographs. The Algorithm used in this system is VGG16. It is a convolutional

neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large- Scale Image Recognition". CAM and gradCAM visualization tools is used for processing. The Accuracy of this model is 96.2% and the advantages of this model is that it is highly accurate.

[3] Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." arXiv preprint arXiv:1711.05225 (2017).
Authors have proposed a model for the detection of Pneumonia. In this model the Image downscaling is done to 224*224. It also contains process of Normalization. It is based on standard deviation and mean. Normalization is a scaling technique in which values are shifted and rescaled in such away that they end up ranging between 0 and 1. It is alsoknown as Min-Max scaling. Data Augmentation is also usedfor performing Random Horizontal Flip and the algorithmused in this model is DCNN. Data augmentation is a techniquewhich we use when we don't have sufficient data, by using this technique we can artificially expand the size of a trainingdataset by creating modified versions of images in the dataset.The accuracy of this model is 0.76 and the dataset used is ChestX-ray14. Advantages of this model is that it indicates the presence of 14 different pathology classes with only one limitations of presence of only frontal radiographs.

[4] Xin Li, Fan Chen, Haijiang Hao, Mengting Li "A Pneumonia Detection Method Based on Improved Convolutional Neural Network" 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2020).
Authors have improved the CNN algorithm in this model. CNN algorithm is improved using the lenet-5 architecture. It is based on the original classical lenet 5 model with additional convolutional layer and pooling layer. It also have feature integration layer. The LeNet-5 architecture consists of two sets of average pooling layer and convolutional layer. This is followed by a flattening convolutional layer. After that comes two fully-connected layers. In the end there is finally a softmax classifier. Two public datasets were used and the results were excellent. Images are processed to 64*64. This helps in decrease the learning time and increase the batch size. The accuracy of this model is 98.44%. The dataset used is RSNA joint Kaggle medical image pneumonia recognition competition , and the advantage of this model is it gives increased accuracy.

[5] In this system the authors have used neural network models to detect the pneumonia from x-rays. They used the chest x-ray images as input. Four models, known as CNN, VGG16, VGG19, InceptionV3 were constructed using CNN and transfer learning methodologies for this model. This model was trained on pediatric chest X-Ray images. The Accuracy obtained of this model is 92.8% with a sensitivity of 93.2% and a specificity of 90.1%. The testing data is increased by data augmentation techniques and hence, the model was tested on a greater number of images. Kaggle's kernels were used as GPU for training, testing and validation of the model. In this model, confusion matrix was plotted by comparing the obtained values.

[6] In this model, the authors have used a deep learning algorithm which is based on CNN to identify and classify the presence of Pneumonia inside a patient's lungs using Chest X-ray images as input. Visual Geometry Group (VGG16) is used in this model and obtained the average accuracy of 68%, average specificity of 69% and the current state of art accuracy is 51%. This model discusses the design and implementation of deep learning model using convolutional neural network. This model is created to provide a new paradigm for diagnosing the pneumonia from Chest X-Rays. The Authors analyze open available through the National Institute of Health consisting of 112K chest X-rays of more than 30,000 unique patients. The model is proposed to increase the accuracy in detecting the pneumonia virus using the Chest X-Rays. For deep learning to be incorporated in the areas like detecting pneumonia, high accuracy is very important.

III. PROPOSED SYSTEM AND ALGORITHM

1) Data

1.1] Data Collection

The biggest challenge for us was to collect the X-Ray images for Hanta Virus. We couldn't find a readily available dataset for hanta Virus. We were able to find several X-Ray images from different websites for our model training. We cropped away some unwanted information from the images like website logos and some texts. For Pneumonia and normal lung images we got a dataset readily available from kaggle.com. We omitted some images to get a balanced data set in the end this will help us avoid overfitting.

1.2] Training

First we gave labels to the images by using opencv and numpy libraries. Now We had labelled data for both Hanta Virus and Pneumonia. we spited the labelled data into train and test set. Prior to feeding these images to our CNN Network we resized these images to 50X50 size by using OpenCV2 Python Library. Now our data was ready to proceedfor feeding it to CNN.

1.3] System Architecture Diagram

In the System Architecture Diagram(below) we can see that CT scan images of lungs is given as the input to the system. The first step is the pre-processing which has three parts(RGB to Grayscale, Thresholding and Edge detection). In Grayscaling we convert the colour image to gray scale by taking the RGB value of each pixel and then taking its mean. In Thresholding we highlight the region of intrest(in our model it is the lung) by converting remaining pixels to white. In Edge detection we extract the main part of the input image that was highlighted during thresholding and leave the remaining part. Here the preprocessing is completed. In the next step the pre-processed image is passed to the algorithm. We are using CNN algorithm since it is more accurate as compared to other algorithms. CNN consists of four layers(explained below in detail). After the result is predicted by the model it is then labeled(the label here is name of the disease). Based on the the respected result report will be generated.

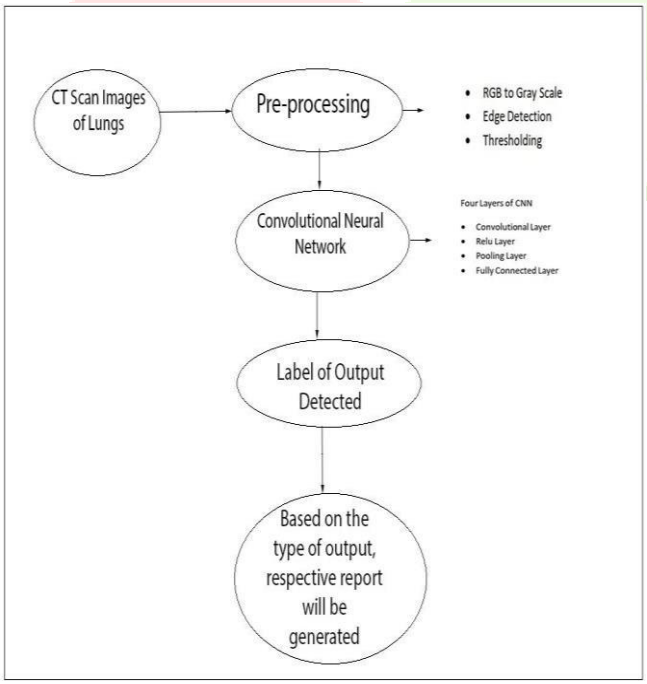


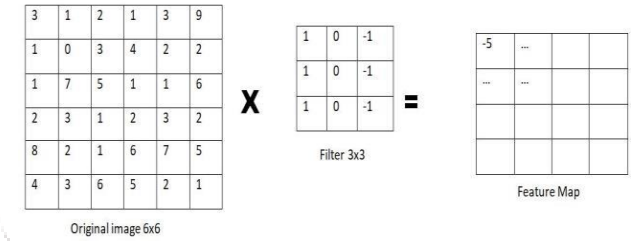
Fig. 1 System Architecture Diagram

2) CNN

In older days image classification algorithms required human man power for extraction of features which used too much time. Modern Deep learning algorithms like CNN does not require tremendous time for feature extraction. CNN automatically learns the features during the training phase through the data. CNN consist of four layers. Convolutional layer, relu layer, pooling layer, fully connected layer. CNN has two parts first in which neurons are connected to only a few neurons ahead of them and in the second part i.e. in fully connected layer the neurons are connected to all the neurons ahead.

- Convolutional Layer

This layer is the core layer of the Convolutional Neural Network algorithm. This layer is responsible for feature extraction. Several filters are selected and applied to the input image that result in creation of feature map. The filters are learned by CNN during the training phase. CNN can extract more information from the image that manually designed filters may not be able to find. Every filter in the neural network is randomly taken from some distributions. This helps in making all filters to learn different features. The filter has a small size which is multiplied to a part of the input image. Element wise matrix multiplication is done between filter and part of image. These values are kept in a matrix called feature map. The filter is moved pixel by pixel throughout the image to get feature map. Different filters will create different feature maps. CNN updates the filter through backpropogation.



The activation function used in convolution layer is ReLU. ReLU activation outputs the maximum value between 0 and the input it got. Due to ReLU activation function the value of gradient will be either 0 or 1. This solves the prolem of vanishing gradient and exploding gradient as we wont have values greater than 1 or values close to zero. ReLU also helps us to converge faster.

- Max-Pooling layer

The output of the Convolutinal layer is given to pooling layer. Here we define a filter with a size. We have taken filter size as 3. The filter takes the maximum value and passes that value. This filter is passed throughout the feature map. This reduces the dimention of the feature map abstracts the features present in that region. Pooling function shrinks the number of

parameters to learn for the network. Max Pool makes the model to detect feature when they are present on different location, scale and rotation.

- Fully Connected Layer

In this layer all the neurons are connected to all the other neurons. This layer predicts the final label of the input. Here we have taken softmax as the activation function which will classify the image to Pneumonia, Hanta Virus or Normal.

- Dropout Layer

We have also added a dropout layer which takes dropout rate as a parameter. This cuts off all the connections of the neurons present in that layer. Dropout rate implies that how many neurons should be dropped out. The Neurons are selected randomly for each iteration. This acts as an regularizer to the Neural Network and helps to avoid overfit.

We iterate through all these layer a couple of times except the dropout layer to design the network. After the first fully connected layer the dropout layer is added. In the end the f1-score is measured by the classification report.

IV. RESULT & DISCUSSION

The Model performed fairly well on the training dataset. The Model can be improved if we get more hanta virus data in the future. Due to unavailability the data model struggled with external data. Pneumonia is a symptom of Hantavirus. Though Hantavirus is very deadly disease but it does not spreads from one human to another and so there are very less chances of being a victim to it.Pneumonia is a symptom of Hantavirus. Though Hantavirus is very deadly disease but it does not spreads from one human to another and so there are very less chances of being a victim to it.

If we include more Data and fine tune again then the system can also detect other lung diseases like Lung Cancer, Covid- 19, etc. given Enough data is available When there will be technological advances in the filed of medical science and the patients will have access to the soft copies of their X-Rays we can also take this system online to implement it on a large scale and make it available to everyone.

V. CONCLUSION AND FUTURE WORK

Pneumonia-screening CNNs achieved better internal than external performance. The trained models performed better on the pooled data from different sites with different pneumonia prevalence, but not on external data. CNNs robustly identified hospital system and department within a hospital, which can have large differences in disease burden and may confound predictions. Pneumonia is a symptom of Hantavirus. Though Hantavirus is very deadly disease but it does not spreads from one human to another and so there are very less chances of being a victim to it.

If we include more modules then the system can also detect other lung diseases like Lung Cancer, Covid-19, etc. When there will be technological advances in the filed of medical science and the patients will have access to the soft copies of their X-Rays we can also take this system online to implementit on a large scale and make it available to everyone.

References

[1] Zech, John R., Marcus A. Badgeley, Manway Liu, Anthony B. Costa,Joseph J. Titano, and Eric Karl Oermann. "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross- sectional study." PLoS medicine 15, no. 11 (2018): e1002683.

[2] Rajaraman, Sivaramakrishnan, Sema Candemir, Incheol Kim, George Thoma, and Sameer Antani. "Visualization and interpretation of convolutional neural network predictions in detecting pneumonia in pediatric chest radiographs." Applied Sciences 8, no. 10 (2018): 1715.

[3] Rajpurkar, Pranav, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding et al. "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning." arXiv preprint arXiv:1711.05225 (2017).

[4] Xin Li, Fan Chen, Haijiang Hao, Mengting Li "A Pneumonia Detection Method Based on Improved Convolutional Neural Network" 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC 2020).

[5] Gaurav Labhane, Rutuja Pansare, Saumil Maheshwari, Ritu Tiwari, Anupam Shukla. Detection of Pediatric Pneumonia from Chest X-Ray Images using CNN and Transfer Learning. 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE-2020), 07-08 February 2020, (IEEE Conference Record # 48199)

[6] Mohammed Aledhari, Shelby Joji, Mohamed Hefeida, and Fahad Saeed Optimized CNN-based Diagnosis System to Detect the Pneumonia from Chest Radiographs. 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).

[7] 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; 2017 [cited 1 July 2018].

[8] Wang G, Giannakis G B, Chen J. Learning ReLU Networks on Linearly Separable Data: Algorithm, Optimality, and Generalization[J]. IEEE Transactions on Signal Processing, 2019, 67(9):2357-2370.

[9] Kermany D S , Goldbaum M , Cai W , et al. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning[J]. Cell, 2018, 172(5):1122-1131

[10] Lakhani, Paras, and Baskaran Sundaram. "Deep learning at chest radiography: automated classification of pulmonary

tuberculosis by using convolutional neural networks."Radiology 284, no. 2 (2017):574-582.

[11] Sharma, Abhishek, Daniel Raju, and Sutapa Ranjan. "Detection of pneumonia clouds in chest X-ray using image processing approach." In 2017 Nirma University International Conference on Engineering (NUiCONE), pp. 1-4. IEEE, 2017.

[12] Gang, Peng, Wang Zhen, Wei Zeng, Yuri Gordienko, Yuriy Kochura, Oleg Alienin, Oleksandr Rokovyi, and Sergii Stirenko. "Dimensionality reduction in deep learning for chest x-ray analysis of lung cancer." In 2018 Tenth International Conference on Advanced Computational Intelligence (ICACI), pp. 878-883. IEEE, 2018.

[13] Stephen, Okeke, et al. "An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare." Journal of healthcare engineering 2019 (2019).

[14] Chan, Heang-Ping, Berkman Sahiner, Lubomir Hadjiyski, Chuan Zhou, and Nicholas Petrick. "Lung nodule detection and classification." U.S. Patent Application 10/504,197, filed September 22, 2005.

[15] Van Ginneken, Bram, Mikkel B. Stegmann, and Marco Loog. "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database." Medical image analysis 10, no. 1 (2006): 19-40.

[16] Neuman, Mark I, Lee, Edward Y, Bixby, Sarah, Diperna, Stephanie, Hellinger, Je_rey, Markowitz, Richard, Servaes, Sabah, Monuteaux, Michael C, and Shah, Samir S. Variability in the interpretation of chest radiographs for the diagnosis of pneumonia in children. Journal of hospital medicine, 7(4): 294-298, 2012.

[17] Hopstaken, RM, Witbraad, T, Van Engelshoven, JMA, and Dinant, GJ. Inter-observer variation in the interpretation of chest radiographs for pneumonia in community-acquired lower respiratory tract infections. Clinical radiology, 59(8):743-752, 2004.