International Journal of Computer Science & Communication (ISSN: 0973-7391)
Special Issue, Page 100-103, December 2020 ------ available at www.csjournals.com

Survey on Predictive Modeling for Covid-19 and Pneumonia Detection

Srinidhi B [1] Rohan T S [2] Vishala G [3] Surabhi S [4] Dr. Manjunatha Kumar B.H [5]

[1 2 3 4] B.E., Student, Department of CSE, Sri Siddhartha Institute of Technology, Tumakuru

[5] Assistant Professor, Department of CSE Sri Siddhartha Institute of Technology, Tumakuru

Abstract: The Corona Virus (Covid-19) pandemic has taken the whole universe to downside and increasing rapidly day by day. The availability of covid test kits in hospitals are less due to the rapid increases of covid cases. Hence it is necessary to implement an automatic detection model as a quick alternative diagnosis option to prevent the rapid spread of Covid-19, one such possible method for detecting Covid-19 is x-ray imaging. Other disease that have achieved the immense spread in the recent times which are related to the chest, with Pneumonia topping the list, Based on our survey the main objective is to demonstrate the methods for detecting the Covid-19 and pneumonia using the different approaches of machine learning, which inturn contributes to the control of the epidemic by reducing the work pressure of doctor's and frontline radiologists.

Keywords: Corona Virus, Pneumonia, Machine Learning, Convolutional Neural Network

I. INTRODUCTION

The Corona Virus disease emerged in china in December 2019 [1].The Covid-19 virus mainly affects the respiratory system and causes diseases like Middle East Respiratory Syndrome(MERS-COV) and Severe Acute Respiratory Syndrome(SARA-COV). Cough, fever , muscle pain are some of the common symptoms of this diseases. Meanwhile predicting the presence of corona virus in the human body by collecting the temperature reading and blood samples of the patient is also a tough task that still has the time constraint and questions on accuracy. Hence, creating a Machine learning model aids the medical research team in predicting the current status of the patient. Machine Learning model like convolutional neural network is used in the classification, segmentation and lesion detection of medical data. The Magnetic Resonance Imaging (MRI), Computed Tomography (CT) and X-ray imaging all these medical imaging techniques are analysed using Machine learning models. A similar method can be implemented for the detection of pneumonia in a person. Pneumonia is an infection that inflames the air sacs in one or both lungs. The most common symptoms of pneumonia includes difficulty in breathing, fever, cough, chills. Pneumonial pneumonia is the primary virus which causes this disease where the air sacs is filled with fluid or pus. The main trait of this disease that is affecting the diagnosis is its symptoms, most of the patients aren't aware that they might have this symptoms until it has reached a greater stage, with no process to detect it other than scanning. By the time the patient gets diagnosed by this it might be too late. , a machine is considered to be much more accurate since it compares the scanned copy of the current patients with thousands of other scanned copies related to pneumonia to provide an accurate and well proven result.

II. METHODOLOGY CLASSIFICATION

1. CNN – A Convolutional Neural Network(CNN) is one of the technique for image recognition, image classification, object detection , face recognition etc, and also a most efficient and widely used algorithm in deep learning, CNN takes images, text, sound or videos as it's input. The image is given as the input of the convolutional neural layer it process the given input image and classifies as per the categories. The input image given to the computer will be divided into array of pixels and the pixel resolution is based on the resolution of the given input image. The convolutional neural network contains different layers of image classification and in each layer filters are applied, and the output of one layer becomes input for the other layer. There are three layers in the convolutional neural network :

a) Convolutional Layer
b) Pooling Layer.
c) Fully Connected Layer.

a) Convolutional Layer: Convolutional neural network is made up of a multiple layers of neurons. The layers mainly include convolutional layer and pooling layers. Convolutional layer is the base layer of Convolutional neural network. It is responsible for determining the features of the pattern. In this layer, the input image is passed through a filter. The values resulting from filtering consist of the feature map. This layer applies some kernels that slide through the pattern to extract low- and high-level features in the pattern. The kernel is a 3x3 or 5x5 shaped matrix to be transformed with the input pattern matrix. Stride parameter is the number of steps tuned for shifting over input matrix. The output of convolutional layer can be given as:

$$x_j^l = f\left(\sum_{a=1}^{N} W_j^{l-1} * y_a^{l-1} + b_j^l\right) \quad - (1)$$

Where x_j^l is the j^{th} feature map in layer l, w_j^{l-1} indicates j^{th} kernels in layer l-1, y_a^{l-1} represents the a^{th} feature map in layer

l-1, b_j^{l-1} indicates the bias of the j^{th} feature map in layer l, N is number of total features in layer l-1, and(*) represents vector convolution process.

b) Pooling Layer: This layer is usually incorporated between a successive convolutional layers. The second layer after the convolutional layer is the pooling layer. Pooling layer is usually applied to the created feature maps for reducing the number of feature maps and network parameters by applying corresponding mathematical computation. In this study, we used maxpooling and global average pooling. The max-pooling process selects only the maximum value by using the matrix size specified in each feature map, resulting in reduced output neurons. There is also a global average pooling layer that is only used before the fully connected layer, reducing data to a single dimension. It is connected to the fully connected layer after global average pooling layer. The other intermediate layer used is the dropout layer. The main purpose of this layer is to prevent network overfitting and divergence.

c) Fully Connected Layer: Fully connected layer is the last and most important layer of Convolutional neural network. This layer functions like a multilayer perceptron. Rectified Linear Unit (ReLU) activation function is commonly used on fully connected layer, while Softmax activation function is used to predict output images in the last layer of fully connected layer. Mathematical computation of these two activation functions are as follows:

$$ReLU(x) = \begin{cases} 0, x < 0 \\ x, x \geq 0 \end{cases} \quad - (2)$$

$$\text{Soft max}(x_i) = \frac{e^{x_i}}{\sum_{y=1}^{m} e^{x_y}} \quad - (3)$$

Where $x_i$and m represents input data and the number of classes respectively.

2. KE Seive Algorithm.

The KE Seive Algorithm is a non-iterative and adopts a new approach, which separates N data points of n-dimension by at least one hyperplane.

The number of hyperplanes, q needed approximately to separate N data points is the order of q = $\log_2$ (N) [6] provided N < 2^n and the computational complexity of this algorithm in Big-

O asymptotic notation is approximately,

O((n.N$\log_2$(N))+ (n^3 $\log_2$ (N)))

Where N is the data points and n is the dimension of space.

For training the samples we need a training set. Assume a n dimensional space with N train points. And G space and S space are the two additional space with points and no points in the space respectively. The above can be represented in an equation form

$1 + a_1 X_1 + a_2 X_2 + a_3 X_3 + .. + a_n X_n = 0$ ----(1)

Then initialization of the plain has to be done and then we need to calculate the orientation vector(OV).The OV is calculated with respect to the points in the G space and S space .It gives the information about the points whether it is lied on positive or negative side. Then we need to collect the neighbours separately and repetition done for all the n points in the dimensional space. Later separation of the neighbours in the S space has to be done by drawing a new plane in S space. Whenever we add a new plane we need to update the OV with respect to the new hyperplane. This process is repeated until all the training set is completed. Dot product technique[Dot product of Ov and train points] is used in order to test the data.

III. RELATED WORK

There are few studies on the detection of covid-19 and pneumonia disease.

Antonios Makris et al[4] proposed a classification model using CNN(1). The pre trained transfer models are used in convolutional neural network in order to obtain high prediction accuracy using x-ray images as it's input. The Dataset consists are of 3 categories i.e chest x-ray of covid affected patients, normal chest x-ray and pneumonia affected chest x-ray. Covid-19 samples are from Dr. Joseph Cohen's Github repository. Chest x-ray of patients with Middle East Respiratory Syndrome(MERS-COV) Severe Acute Respiratory Syndrome(SARA-COV) and covid-19 are included in this repository. In addition, 112 normal and 112 pneumonia chest x-ray images were selected from kaggle's repository. The dataset used for this article is evenly distributed regarding the number of cases and consists of three classes (Covid-19, pneumonia, normal) and it is freely available in Antonis Makris's repository. The x-ray images are divided into two classes, normal and Covid-19. Further to overcome the lack of input data, a transfer learning technique is applied by employing the ImageNet dataset. The proposed model is called as DeTraC and has three phases. In the first phase an ImageNet pre-trained CNN is employed for local feature extraction. As in the second phase a stochastic gradient descent optimisation method is used for training and at last the class-composition layer is adapted for final classification of the x-ray images using error correction criteria applied to a softmax layer. By using the VGG16 pre-trained ImageNet network the accuracy obtained for this is 95.12% using chest x-ray images, Fine-tuning process is used on the VGG16 network architecture. The network is represented with weights pre-trained on ImageNet, VGG16 has 13 convolutional and three fullyconnected (FC) layers. The last set of layers which has the FC layers together with the softmax activation function that is called as "head". Afterwards, the FC layers are eliminated and the final POOL layer is taken as a feature extractor as portray. Finally, a new FC head layer is randomly initialized and placed over the top of the original architecture. It is worth mentioning, that the

body of the network, that is the convolutional layers have been "repressed", i.e. only the FC head layer is trained. The reason for this behaviour is that the convolutional layers have already learned discriminative filters while the FC head layer is randomly initialized from scratch and random values are able to destroy the learned features. As a result, the best performance in this article is obtained for VGG16 that is 95% (Accuracy).

Ali Narin et al[3] proposed a classification model using CNN(2).Five pre trained convolutional neural network based model have been proposed for the detection of covid-19 ,pneumonia infected patients using chest x-ray radiography. The dataset consists of chest x-ray images of 341 covid-19 patients have been obtained from the open source Github repository shared by Dr.Joseph Cohen et al[10].The repository consists of chest x-ray images of mainly patients with acute respiratory distress syndrome (ARDS),Covid19, Middle east respiratory syndrome(MERS), pneumonia, severe acute respiratory syndrome(SARS).2800 normal chest X-ray images were selected from "chestxray8" database. In addition, 2772 bacterial and 1493 viral pneumonia chest x-ray images were used from kaggle repository called "Chest X-ray images (Pneumonia)"[11]. Here they have built Convolutional Neural Network (CNN) based ResNet50,ResNet101,ResNet152,InceptionV3 and Inception-ResNetV2 models for the classification of covid-19 chest x-ray images into three different binary classes(Binary class 1=covid-19 and normal, Binary class 2=covid-19 and viral pneumonia, Binary class 3=covid-19 and bacterial pneumonia).In addition they have applied a transfer learning technique that was realized by using ImageNet data to overcome the insufficient data and training time. python programming language was used to train the proposed deep transfer learning models. The dataset used was randomly split into two independent datasets with 80% and 20% for training and testing respectively. performance results shows that ResNet50 yields the highest accuracy among the 5 models.

Fatima M. Salman et al[12] proposed a model using CNN(3) model for the detection of covid-19. 130 covid-19 affected chest x-ray and 130 normal chest x-ray are used as the dataset for this work. The COVID-19 X-ray images are collected from the GitHub repository shared by Dr. Joseph Cohen[10], a postdoctoral fellow at the University of Montreal. The Normal X-ray images of pneumonia collected from Kaggle repository and Open-i repository. The COVID-19 excludes the MERS, SARS, and ARDS Images. Based on the deep learning architectures like VGG16, ResNet50 and InceptionV3 they have used the deep feature extraction for the dataset in order to detect the presence of covid-19. Finally, they trained, validated and evaluated the performance of the proposed model; firstly they have created a CNN model which includes 3 major steps as mentioned at the start of the paper. After creating the CNN model, they have trained the model for 20 epochs and using data augmentation for preventing the model from overfitting. A total of 30 x-ray samples were selected randomly from each group (Covid-19 and healthy) for the test set. The accuracy of the testing came out 100%.

Dr. K Eshwaran et al[5] proposed a model uses two datasets [7] [8], which has been combined to classify 3 classes as a whole. A total of 108 chest X-ray images of COVID-19 have been taken from this dataset [7]. The other two types of chest X-ray images pneumonia and normal are taken from the Kaggle dataset [8].The KE Seive Algorithm is used to train the 4771 images and tested on 1193 images. Firstly they have converted the images for 224 x 224 standard size and then they have applied the histogram equalization. This pre-proposed vector has been fed into the feature extractor of the pre-trained VGG-19[9] based on the convolutional neural network model. This model resulted with the accuracy of 98.49%

Formulae:

Accuracy (ACC) = T P + T N / n - (1)

Precision (P) = T P / T P + F P - (2)

Recall (Sensitivity) = T P / T P + F N - (3)

F1–score=2(Precision × Recall Precision + Recall) - (4)

Specificity = T N / T N + F P - (5)

Where TP- True positive

TN - True Negative

FP - False Positive

FN - False Negative

IV. PERFORMANCE AND COMPARISON

Method	Accuracy (%)	Recall (%)	Specificity (%)	Precision (%)	F1-Score (%)
CNN(1) Covid-19	95.0	96.0	98.0	96.0	96.0
Pneumonia	95.8	91.0	98.0	95.0	93.0
CNN(2) Covid-19	96.1	91.8	96.6	76.5	83.5
Viral Pneumonia	99.5	99.4	99.5	98.0	98.7
Bacterial Pneumonia	99.7	95.6	100	100	97.7
CNN(3)	100	100	100	100	100
K E Seive(4) Covid -19	98.07	100	100	100	100
Pneumonia	99	99	99	99	99

V. CONCLUSION

Diseases have been a part of the human way of life from time immemorial and humans have always strived to find means to cure it and achieve a healthier way of life. For each cure to be found and tested on a disease or the carrier of the virus causing the disease, there is a set of procedures to be followed. where diagnosing the disease accurately takes the important place in all of these procedures. A medical

treatment can be proceeded depending on the diagnosis of whether the patient is affected or not. Diseases transferred due to viruses are spreading drastically across the world; the main reason for the spread shall be the lack of accurate detection of the symptoms. Pneumonia does not have detectable symptoms and the symptoms used to detect covid-19 are inaccurate. By the time the patient understands a detectable symptom and gets diagnosed it might be too late and the disease might have spread from the host to different parts of the body extensively. Thus, our medical team is in direct need of a toolkit that can diagnose the presence of virus in a patient in a fast, low cost and efficient manner. Thus the different machine learning models satisfies this requirement by taking the x-ray scans of a patient's chest and analysing it to further classify it as affected or not. After this survey we came to know that CNN is a better model for the detection of covid-19 and pneumonia which yields more accuracy than the other models of machine learning.

REFERENCES

[1] Origin of Covid-19, a report by journal of Advance Research. https://www.sciencedirect.com/science/article/pii/S2090123220300540

[2] Covid-19 symptoms report by WHO https://www.who.int/healthtopics/coronavirus#tab=tab_3

[3] Alinarin, ceren kaya ,Ziynet Pamuk "Automatic Detection of coronavirus Disease using x-ray images and deep convolutional neural networks https://arxiv.org/abs/2003.10849

[4] Antonios Makris, Ioannis Kontopoulos, Konstantinos Tserpes,"Covid-19 detection from chest x-ray images using deep learning and convolutional neural network https://www.medrxiv.org/content/10.1101 /2020.05.22.20110817v1

[5] S Sai Thejeshwar,Chaitanya Chokkareddy, Dr K Eswaran ,"Precise prediction of covid-19 in chest x-ray images using Ke sieve Algorithm https://www.medrxiv.org/content/10.1101

[6] /2020.08.13.20174144v1.full.pdf

[7] K Eswaran. "A non iterative method of separation of points by planes in n dimensions and its application". https://arxiv.org/abs/1509.08742

[8] Joseph Paul Cohen and Paul Morrison and Lan Dao, COVID-19 image data collection, arXiv preprint arXiv:2003.11597,2020 https://github.com/ieee8023/covidchestxray-dataset

[9] Chest X-Ray Images (Pneumonia) dataset in Kaggle - https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[10] K. Simonyan, A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition" https://arxiv.org/abs/1409.1556

[11] Cohen JP, Morrison P, and Dao L. COVID-19 Image Data Collection. https://arxiv.org/abs/2003.11597

[12] Mooney P. Chest X-ray Images(Pneumonia). Kaggle epository: https://www.kaggle.com/ paultimothymoo ney / chest-xray-pnemonia,2018

[13] Fatima M Salman,Samy s.abuNaser,Eman Alajrami,Bassem S.AbuNasser,Belal .M.Ashqar,"Covid-19 detection using artificial intelligence" http://dstore.alazhar.edu.ps/xmlui/bitstream/handle/123456789/587/IJAER200304.pdf?sequence=2&isAllowed=y