2021 2nd International Conference for Emerging Technology (INCET)
Belgaum, India. May 21-23, 2021

COVID 19, Pneumonia and Other Disease Classification using Chest X-Ray images

Abhishek Panwar
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
abhishekpanwar971@gmail.com

Akshay Dagar
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
akshaydagar98@gmail.com

Vishal Pal
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
vishalpal1999@gmail.com

Dr. Vinod Kumar
Department of Computer Science & Engineering
Delhi Technological University Shahbad
Daulatpur, Main Bawana Road, DL - 42, IN
vinod_k@dtu.ac.in

2021 2nd International Conference for Emerging Technology (INCET) | 978-1-7281-7029-9/20/$31.00 ©2021 IEEE | DOI: 10.1109/INCET51464.2021.9456192

***Abstract*— In this report, we propose an alternative way of detecting the Covid-19 disease using Convolutional Neural Networks based deep learning models which will classify between Covid-19 and other similar respiratory diseases such as other types of Viral Pneumonia or Bacterial Pneumonia, using images of Chest X-Rays. The models developed as part of this project achieved approximately 98% accuracy on a test dataset of 300 images.**

Keywords—Covid-19, Pneumonia, Disease Classification

I. INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome Corona virus 2 (SARS-CoV-2). It was first identified in December 2019 in Wuhan, Hubei, China, and has resulted in an ongoing pandemic. As per WHO on 21 March 2021, more than 123 million cases have been reported across 223 countries and territories with more than 2.7 million deaths: more than 69.5 million people have recovered Common symptoms include fever, cough, fatigue, shortness of breath or breathing difficulties, and loss of smell and taste. While most people have mild symptoms, some people develop acute respiratory distress syndrome (ARDS) possibly precipitated by cytokine storm, multi-organ failure, septic shock, and blood clots. The incubation period may range from one to fourteen days. This disease generally spreads through air during close proximity of an infected person, contaminated surfaces etc. Currently the standard method of diagnosis is by real-time reverse transcription polymerase chain reaction (rRT-PCR) from a nasopharyngeal swab. The Chest X-ray imaging may also be helpful for diagnosis in individuals where there is a high suspicion of infection based on symptoms and risk factors, however guidelines do not recommend using it for routine screening currently. In our study here, we trained our models on openly available chest X-ray dataset maintained by physicians and sources like Kaggle. To avoid any sort of overfitting due to shortage of data, we used various renowned image augmentation techniques. Our models had shown high accuracy on Testing dataset. However , we need more data and extensive testing to put it in production. In coming section, we will discuss about our literature overview, our methodology and results.

II. LITERATURE OVERVIEW

Mucahid Barstugan et. al. [5] study presents early phase detection of Coronavirus (COVID-19), which is named by the World Health Organization (WHO), by machine learning methods. The detection process was implemented on abdominal Computed Tomography (CT) images. The expert radiologists detected from CT images that COVID-19 shows different behaviours from other viral pneumonia. Therefore, the clinical experts specify that COVİD-19 virus needs to be diagnosed in early phase. For detection of the COVID-19, four different datasets were formed by taking patches sized as 16x16, 32x32, 48x48, 64x64 from 150 CT images. The feature extraction process was applied to patches to increase the classification performance. Grey Level Co-occurrence Matrix (GLCM), Local Directional Pattern (LDP), Grey Level Run Length Matrix (GLRLM), Grey-Level Size Zone Matrix (GLSZM), and Discrete Wavelet Transform (DWT) algorithms were used as feature extraction methods. Support Vector Machines (SVM) classified the extracted features. 2-fold, 5-fold and 10-fold cross-validations were implemented during the classification process. Sensitivity, specificity, accuracy, precision, and F-score metrics were used to evaluate the classification performance.

In Enzo Tartaglione et. al. [8] insights are provided and also warnings are raised on what is reasonable to expect by applying deep learning to COVID classification of CXR images. A methodological guide and critical reading of an extensive set of statistical results that can be obtained using currently available datasets is provided. In particular,the challenge posed by current small size COVID data and show how significant can be the bias introduced by transfer-learning using larger public non-COVID CXR datasets is taken. Contribution is also made by providing results on a medium size COVID CXR dataset, just collected by one of the major emergency hospitals in Northern Italy during the

peak of the COVID pandemic. These novel data allow for the contribution to validate the generalization capacity of preliminary results circulating in the scientific community. The conclusions shed some light into the possibility to effectively discriminate COVID using CXR.

However, in Gianluca Maguolo et. al. paper [9], different testing protocols used for automatic COVID-19 diagnosis from X-Ray images in the recent literature are compared and evaluated. It is shown that similar results can be obtained using X-Ray images that do not contain most of the lungs. It is possible to remove the lungs from the images by turning to black the center of the X-Ray scan and training their classifiers only on the outer part of the images. Hence, it is deduced that several testing protocols for the recognition are not fair and that the neural networks are learning patterns in the dataset that are not correlated to the presence of COVID-19. Finally, it is shown that creating a fair testing protocol is a challenging task, and they provide a method to measure how fair a specific testing protocol is. In the future research it is suggested to check the fairness of a testing protocol using their tools and they encourage researchers to look for better techniques than the ones that they propose.

This paper tries to incorporate better strategies in order to overcome aforementioned shortcomings. However, lack of ample sized dataset acts as a bottleneck in obtaining more accurate results.

III. METHODOLOGY

In this section, we have divided our work in four parts. Firstly, we will discuss about the collection of data, data sources and its relevance. Then, we will move to data processing and preparation part where we discuss about how we prepared the data for training. Then, we will see the different model architectures that we used for training the parameters and finally we will see the training and testing of the data for various parameters.

The workflow of our work throughout this process, can be summarised as shown in the figure down below.

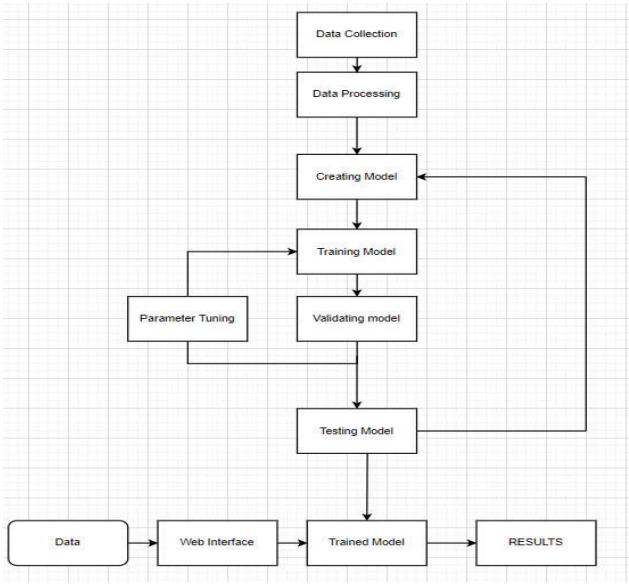

Fig. 1. Workflow Diagram

A. Data Collection

Data is a very important input for any machine learning or deep learning problem. A relevant collection of data helps in analysing a machine learning problem better. In our problem set, we have collected data majorly from two different and prominent sources present online. Firstly, we have data from Covid 19 chest X-ray dataset repository on Github maintained by Joseph Paul Cohen, Paul Morrison, Lan Dao, Karsten Roth, Tim Q Duong and Marzyeh Ghassemi. The data of this repository is collected from various hospitals and physicians. This data was collected from public sources like hospitals and maintained under the project approved by University of Montreal's Ethics Committee. We have chosen the 'AP' and 'PA' view images of chest X-ray for our identification. We further classified it in majorly three classes as covid affected, non-covid pneumonia affected and normal/ other disease affected. Further, we collected the data of Pneumonia and other diseases from Kaggle competition. Apart from these sources, we supplemented our data set from various online sources and studies done before. Finally, after combining these datasets, we were able to gather a total of 4563 images (478 Covid 19 or Covid 19 affected pneumonia images, 2647 other pneumonia images and the rest 1438 were the normal/ other disease images).

B. Data Preparation and Processing

From this combined dataset, we will create a testing dataset containing 100 images belonging to each of the 3 classes (Covid-19 Pneumonia images, Other Pneumonia Images, and Other Diseases/No disease images) and validation dataset containing 78 images belonging to each of the three classes. This will leave us with 300 Covid-19 images, 2469 other Pneumonia images and 1260 images belonging to other diseases/no diseases class

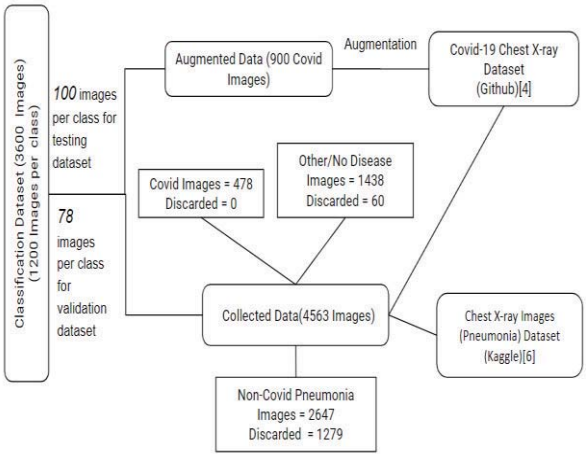

Fig. 2. Dataset Preparation

The leftover training dataset is highly imbalanced and if we use this dataset to train our models, there are high chances of overfitting as the model will be able to obtain a high accuracy by predicting the majority class(es). To balance the collected dataset, we used Image Augmentation techniques. The different augmentation techniques we used are as follows:

- Random Rotations
- Random Shifting (left, right, up, or down)

- Random Gaussian Noise Adding
- Horizontal Flipping (mirror image)
- Random Cropping
- Random Shearing
- Random Scaling

For each Covid-19 Pneumonia Image, we will randomly select 3 of these 10 augmentation functions and apply them to the image to generate 3 augmented copies of the image and save the copies with the original image in our dataset.

After Augmenting, we have 1200 Covid-19 Pneumonia CXR images. However, the dataset is still imbalanced, and we need to reduce the no. of images of the other 2 classes to 1200 each to form a perfectly balanced dataset. Since the other 2 classes contain multiple images from a few persons, we will keep only 1 image per person thus reducing the no. of images belonging to these 2 classes. After that, we will keep only 1200 images from each class to form our final perfectly balanced training dataset.

IV. Model Architectures and Classifiers

In this section, we discuss different model architectures we used. In our work, we mainly use two deep learning models which are 4 Layer CNN (Model 1) and AlexNet [10] (Model 2) Model 1 was further classified on the basis of our dataset. We used Model 1 with the original dataset initially collected once and then we used it with a modified dataset which was supplemented with our augmented dataset.

- Model 1 (4 Layer CNN)
- Model 2 (AlexNet)

The two models were trained with the gathered data as follows:

- Initially, we trained Model 1 with the original dataset. The training was done using Image Data Generators which will read images from the dataset folders and resize them to size (224,224) with 3 channels (RGB) and rescale the images to [0,1]. Data generators were also used for validation images. After training, we only want the model weights which give the best validation accuracy to be saved as our final model weights, hence we use the "Model Checkpoint" callback from the Keras library. We saved the model parameters that result in the best accuracy. Finally, we trained this model for 50 epochs with a default batch size of 32 train images per batch.
- Thereafter, We trained Model 2 with the original dataset. This model is based on AlexNet. We will take the Convolution layers of the AlexNet model and discard the Output layers and Dense (hidden) layers and instead attach our own hidden layers and output layer with Softmax activation to the output of the convolution layers. The complete architecture has already been defined in the previous section on "Model Architectures". The training of this model will be done similar to that of model1. The only exception being that we will fine tune the AlexNet weights while training instead of using randomly initialized weights as done for model 1.
- Finally, we used the same Model 1 however we trained it differently using Image Data generators with augmentation. Basically, we will randomly take images from the training dataset folder and apply random augmentations on them, thus generating augmented versions of those images, on the fly, while training. Note that we will not use the original image, only the augmented ones, while training. Since we have already augmented our original dataset, this will be kind of a double augmentation. Augmenting like this can help us reduce overfitting our model since the dataset contains augmented copies of the original 300 images and hence there will be a lot of similarities. The rest of the training was similar to other training above.

V. Experiment Results and Discussion

In this section, we will see the experiment results collection from testing.

We tested the trained models on the test dataset containing 100 samples for each of the 3 classes, the results are as follows:

TABLE I. No. of Images of Different Classes Used for Training and Testing Dataset

Class Name	*No. of Images*	
	Training	***Testing***
Covid 19	1200	100
Non –Covid Pneumonia	1200	100
Normal/Other diseases	1200	100

(Note: 0 = Covid-19 Pneumonia, 1 = Other Diseases/No Disease, 2 = Other Pneumonia)

The results of the three models are as follows:

Finally, we will take the weighted average of the predictions of the 3 models to get final predictions. This weighted prediction function given below is the final classifier we used in our work. The weighted average is taken in the following manner:

Count_Class0 = (No. of Models with majority class prediction = 0)

Count_Class1 = (No. of Models with majority class prediction = 1)

Count_Class2 = (No. of Models with majority class prediction = 2)

Final Prediction = ((P1*Count_ClassP1) + (P2*Count_ClassP2) + (P3*Count_ClassP3)) / (Count_ClassP1 + Count_ClassP2 + Count_ClassP3)

TABLE II. CLASSIFICATION ACCURACY, PRECISION AND RECALL (IN PERCENTAGES) FOR DIFFERENT DEEP LEARNING MODELS IN COVID-19, PNEUMONIA AND OTHER DISEASES/NORMAL CLASSIFICATION ON TRAINING AND TESTING DATASET

Model Name	***Closs***	**Testing/Training**		
		Accuracy	***Precision***	***Recall***
Model 1	Covid	98.67	98.00	98.00
	Pneumonia	97.00	95.00	95.96
	Normal/other Disease	97.67	97.00	96.04
Model 2	Covid	98.00	98.00	96.07
	Pneumonia	97.67	94.00	98.94
	Normal/Other Disease	98.33	99.00	96.11
Model 3	Covid	98.00	95.00	98.95
	Pneumonia	95.00	88.00	96.70
	Normal/Other Disease	95.67	100.00	88.49
Final Result	Covid	99.00	99.00	98.02
	Pneumonia	98.33	95.00	100.00
	Normal / Other Disease	98.67	100.00	96.15

Thus the overall accuracy for Model 1 is 96.67 %, for Model 2 is 97.00% , for Model 3 is 94.34%.The final results show that our classifier is performing well on a test dataset of 300 images and achieving 98% accuracy with 100% accuracy for “Other Diseases/No disease” class and 99% accuracy for “Covid-19 Pneumonia” class.

Due to lack of data for Covid-19, we cannot test it on a larger dataset, however testing the classifier on a dataset of 1000 images each of classes 1 and 2 and 100 images of class 0 (Covid-19) produced almost similar results with slightly reduced accuracy (97%) which is bound to happen on larger datasets).

VI. CONCLUSION AND FUTURE WORK

AlexNet and 4 layer Convolutional Neural Network have been proved to be really worthy and provided us the commendable results in predicting the Covid 19 cases using only the X ray images. Using them, we are able to achieve almost 98% accuracy which is comparatively higher than many other parallel research work going in the field. But since the data available is quite less right now, we cannot completely rely on these models and their results for actual implementation in the Medical industry. So, to conclude we can say that a more exhaustive availability of the required data is needed.

As future work, we plan to test our classifier on more data as it becomes available so as to produce more reliable results as per the medical industry acceptance standards. Till then, we are also looking forward to training some parallel CNN models on the available data in near future and testing parallel CNN models for the same in order to achieve better results.

REFERENCES

[1] R. M. Pereira, D. Bertolini, D. Bertolini, C. N. S. Jr. and Y. M. Costa, "COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios," *Computer Methods and Programs in Biomedicine,* vol. 194, 2020.

[2] A. I. F. AI, AI2, CZI, MSR, Georgetown, NIH and T. W. House, "COVID-19 Open Research Dataset Challenge (CORD-19)," 2020. [Online]. Available: https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge. [Accessed 2020].

[3] J. P. Cohen, P. Morrison and L. Dao, "COVID-19 Image Data Collection," *arXiv,* p. 4, 25 March 2020. K. Elissa, “Title of paper if known,” unpublished.

[4] J. P. Cohen, "covid-chestxray-dataset," 2020.

[5] S. ozturk, U. Ozkaya and M. Barstugan, "Classification of Coronavirus (COVID-19) from X-rayand CT images using shrunken features," *International Journal of Imaging Systems and Technology,* p. 11, 2020.

[6] Kaggle, "COVID-19 chest xray," 2020.

[7] Wikipedia, "Coronavirus disease 2019," 2019. [Online]. Available: https://en.wikipedia.org/wiki/Coronavirus_disease_2019. [Accessed 2020].

[8] E. Tartaglione, C. A. Barbano, C. Berzovini, M. Calandri, and M. Grangetto, “Unveiling COVID-19 from CHEST X-Ray with Deep Learning: A Hurdles Race with Small Data,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 18, p. 6933, Sep. 2020.

[9] G. Maguolo and L. Nanni, "A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images," *ArXiv,* vol. abs/2004.12823, 2020.

[10] A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097-1105, 2012.