Identifying Severity of Covid-19 and Pneumonia from Chest X-rays

Kutala Mohan Prasad *Lakehead University*
Student Id: 1154915
mkutala@lakeheadu.ca Dr. Sabah Mohammed *Dept. of Computer Science*
Lakehead University
sabah.mohammed@lakeheadu.ca

***Abstract*—Coronavirus (COVID-19) is an infectious disease caused by the SARS-CoV-2 virus. Most infected people will experience moderate to severe respiratory illnesses and recover without the need for specialized treatment. However, some will get very sick and will seek medical help. Older people and those with poor health conditions such as cardiovascular disease, diabetes, chronic respiratory disease, or cancer are more likely to develop serious illness. Anyone can get sick with COVID-19 and get very sick or die at any age.Pneumonia is also an infection in one or both lungs caused by bacteria, viruses, or fungi. The infection leads to inflammation in the air sacs of the lungs, which are called alveoli. The alveoli fill with fluid or pus, making it difficult to breathe. So, for detecting the COVID-19 and Pneumonia in now a days become costly and time taking process especially for common people and radiologists.This paper introduces AUTO-ML process to detect the disease(COVID or Pneumonia) inside the human lung along with severity of the disease without any extra cost.**

***Keywords*—COVID-19, Pneumonia, Severity, AUTO-ML, X-Rays, Virus, lungs**

I. Introduction

Since 2020, we see world suffering from dangerous virus called COVID-19 [1]. Once the patient effected by this virus it will act as a slow poison to him to die main reason for this is respiratory problem, this virus will infect all tissues inside the lungs so, slowly it will be very hard to the human to take breath and finally he will die. In mean time every doctor will do chest x-ray scan to measure the severity of the disease and infection inside the lungs. Usually medical guys will know the status of the disease very easily by seeing the reports and X-ray images but the public who don't have idea on medical terms they will not understand the infected area and infected severity.

After having above idea in our mind we just want to do research on available X-Rays to know the severity of the disease and level of infection inside the body so that public can alert and have a chance to consult or ask doctor for remedy for the damage and doctors can easily identify and give proper medication to the patient. I thought to explore X-ray images as doctors frequently use X-rays and CT scans to diagnose pneumonia, lung inflammation, abscesses, and enlarged lymph nodes. Since COVID-19 attacks the epithelial cells that line our respiratory tract, we can use X-rays to analyse the health of a patient's lungs. Given that nearly all hospitals have X-ray imaging machines, it could be possible to use X-rays to test for COVID-19 without the dedicated test kits [2].

A drawback is that X-ray analysis requires a radiology expert and takes significant time which is precious when people are sick around the world. Therefore developing an automated analysis system is required to save medical professionals valuable time.

If we have Image noise for our dataset what we have to do? What is the percentage of noise on the image? How we have to reduce that noise level on the image?

The common types of noise that are present in x-ray images are Poisson noise, salt and pepper noise, and speckle noise. The salt and pepper is seen in the image as white and black pixels respectively. Poison noise is as a result of uneven distribution of x-rays over the receptor surface. Speckle noise on the other hand occurs as a granular appearance in an image which is produced as a result of random fluctuations in the return signal from an object which is not found to be bigger than a single image processing element. Within a specific area, speckle noise is able to increase the mean grey level.

An image pre-processing step can improve the accuracy of machine learning models. Pre-processed images can help a basic model achieve high accuracy when compared to a more complex model trained on images that were not pre-processed. For Python, the Open-CV and PIL packages allow you to apply several digital filters. Applying a digital filter involves taking the convolution of an image with a kernel (a small matrix). A kernel is an n x n square matrix were n is an odd number. The kernel depends on the digital filter [3].

We have different types of techniques to remove the noise inside the image such as

- Mean filter [4]
- Gaussian Filter [5]
- Median Filter (Highly recommended filter to remove Salt and pepper noise) [6]
- Other Filters [7]

II. Related Research

Shivani Guptaa, Atul Guptab they performed a systematic review on noise identification and handling studies published on various conferences and journals between January 1993 to July 2018. They identified 79 primary studies are of noise identification and noise handling techniques. After investigating these studies, they found that among the noise identification schemes, the accuracy of identification of noisy instances by using ensemble-based techniques are better than other techniques. But regarding efficiency, usually single based techniques method is better it is more suitable for noisy data sets [8]. Brett K talks about the EHR data how and where the medical records were collecting and how denoising autoencoders perform dimentionality reduction enabling visualization and clustering for the discovery of the new subtypes of the disease [9]. VisionPro Deep Learning is used to classify these Chest X-rays from the COVID dataset. The results are compared with the results of COVID-Net and various other state-of-the-art Deep Learning models from the open-source community. Segmenting the lungs in the first step, and then doing the classification step on the segmented lungs only, instead of using the entire image [10]. Apart from that deep learning based approach for the identification and localization of pneumonia in Chest X-rays (CXRs) images. Mask RCNN based model gives more accurate pixel wise semantic segmentation than faster RCNN for pneumonia prone regions in the lungs were discussed by SachinKumar and his team [11]. Madhava raja and his team develop a Deep-Learning System (DLS) to diagnose the lung abnormality using chest X-ray (radiograph) images.

(i) Conventional chest radiographs

(ii) Chest radiograph treated with a threshold filter. The initial experimental evaluation is carried out using the traditional DLS, such as AlexNet, VGG16, VGG19 and ResNet50 with a SoftMax classifier. The results confirmed that, VGG19 provides better classification accuracy (86.97%) compared to other methods [12]. To develop the system to auto detect the covid-19 positive cases such that they selected few different methods in each level and tested the dataset against the model. Convolutional Neural Network with a classification accuracy of 83.02 % and a superior AUC of 0.907, which would mean a better ability to detect the COVID19 using this method.100% accuracy on the validation set using the feed-forward neural network, and this is using as inputs the flattened image and the texture features. Feature-based feed forward NN with an 84.02% classification accuracy and an AUC of 0.850 [13].

III. Methods to Deduct Disease and Severity

In this section, we first introduce the overall architecture of the detection of noise, disease and its severity. Then, we discussed about the AUTO-ML process that we used to run and present the results.

A. Noise Detection Architecture

By using chest x-ray dataset we build a model to convert the given images to HSV and Gray scale images by using OpenCV library. With the help of same library we will apply the different types of noise detecting and removing algorithms such that it will remove different type of noisy data on the image. In the next step I will store all the output images to different folders based on the filter name as shown in the figure 1 so, that it will helpful while applying the detecting algorithm and severity model. The main advantage of doing this is to test the same image by removing different noisy data on it if we do so I hope I will get good results with more accuracy. In the same way, we will take sample of testing images to test the model on both diseases.

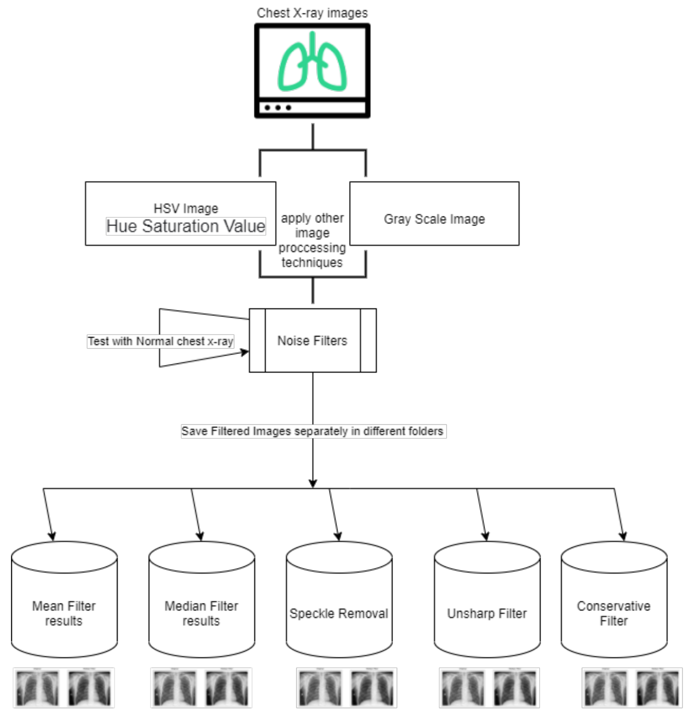

Fig. 1. Noise filtering process on chest X-ray images

B. Architecture For Disease And Severity

In the Second phase of my research, we developed a model to detect the disease affected to the human based on the chest x-ray it may be Covid-19 or Pneumonia. This model will give Total accuracy for given dataset and the table which specifies number of the x-rays those are effected by specific disease.In the next step as shown in figure 2, severity can be calculated by using RALO dataset given by the radiologists [14]. This severity model will take specific disease chest x-ray as an input and will give severity scale in the form extent of lung and degree of opacity.

C. AUTO-ML

Utilization of Machine Learning is common in all the fields including health care but most hospitals are not currently deploying machine learning solutions. One reason for this is that health care professionals often lack the machine learning expertise that is necessary to build a successful model. In order to make machine learning techniques easier to apply and to reduce the demand for human experts. Most of the models wants a human interaction in each stage to add, deploy and run the model. Don't we run the machine learning models without human interaction in such a way that a model should take input from the user and run on its own way to give output? Yes, we have automated machine learning (AutoML) has emerged as a growing field that seeks to automatically select, compose, and

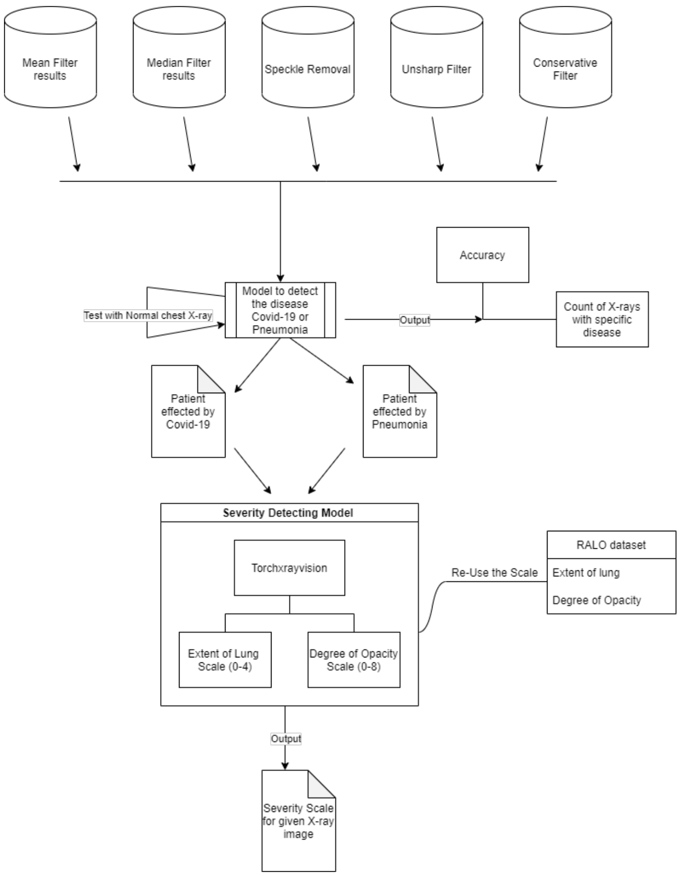

Fig. 2. Detecting the severity of the disease infected chest X-ray image

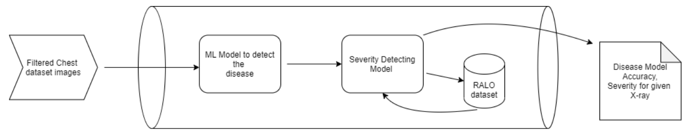

Fig. 3. Auto ML Pipeline

parametrize machine learning models, so as to achieve optimal performance on a given task and/or dataset [15].

The Auto Machine Learning is technique that machine will run the code by taking the input from user. As shown in figure 3, pipeline it will take set of filtered chest images as input and it will run couple of Machine Learning models to detect and report the severity of the disease. There are several key challenges to applying machine learning in the healthcare space that make it very difficult to deploy AutoML solutions [15]. An important challenge in any machine learning problem is assembling a high-quality, representative, and diverse dataset. Ideally, the machine learning model would be trained with data that exactly matches the format and quality of data that would be used at a later point.

Using Machine learning model in health care department is not popular now a days because of not having ML experts in health care department and the models are not giving the accurate performance and the main reason for this is lack of data. If the person wants to run the ML model he should have idea on what we are running and how we have to run the model. But if we have a model or Auto ML technique it will be easy to run and get the results.

IV. Research Materials and Questions

In this section, we discuss the method for the analyses I have carried out and worked to get the severity and detection of the disease.

A. Analysing the Chest X-ray Image

X-ray images as a doctors frequently use X-rays and CT scans to diagnose pneumonia, lung inflammation, abscesses, and enlarged lymph nodes. Since COVID-19 attacks the epithelial cells that line our respiratory tract, we can use X-rays to analyse the health of a patient's lungs [16]. To analyse the X-ray images we should know the different parts of the X-rays for example if the lungs filled with total air then what it will show in X-ray, what if we have bone fractures inside the lungs. What will be the identification of other tissues, normally Air appears black, fat appears gray, soft tissues and water appear as lighter shades of gray, and bone and metal appear white. The denser the tissue, the whiter it will appear on x-ray. Denser tissues appear radiopaque, bright on the film; less dense tissues appear radiolucent, dark on the film. Chest x-ray review is a key competency for medical students, junior doctors and other allied health professionals. Using A, B, C, D, E is a helpful and systematic method for chest x-ray review [16]: A: airways
B: breathing (the lungs and pleural spaces)
C: circulation (cardiomediastinal contour)
D: disability (bones - especially fractures)
E: everything else, e.g. pneumoperitoneum

B. Research Questions

RQ1: Removing or adding noisy data as a radiographer, your goal from day to day, from patient to patient, is to complete an imaging exam that provides sufficient information for an accurate clinical diagnosis. The patient's anatomy has created variations in the X-ray intensity that the imaging system uses to create the image. But overlaying this image "signal" is the inherent statistical "noise" associated with the X-ray production process. As you will see, there are two main challenges. One is that the X-ray process is governed by fundamental laws of nature that we cannot alter and whose characteristics introduce unavoidable "noise". The second is that the multiple processes in the X-ray image capture process also generate noise, but they are amenable to optimization through careful detector design. In an chest X-ray test where only a small amount of radiation has been used to create the image (low exposure), the distracting visual appearance of the statistical noise (sometimes known as "salt and pepper" noise) relative to the size of the signal variations generated by the patient's anatomy, can reduce the visibility of subtle, clinically important features. This can lower the radiologist's diagnostic confidence.

RQ2: novel coronavirus disease (COVID-19) is the most challenging problem for the world. The disease is caused by severe acute respiratory syndrome coronavirus-2 (SARS-COV-2), leading to high morbidity and mortality worldwide. The COVID-19 test is typically done on respiratory samples obtained by a nasopharyngeal swab. However, a nasal swab or sputum sample may also be used. Results are generally available within hours based on the test type. Chest X-rays scans may be helpful to diagnose COVID-19 in individuals with a high clinical suspicion of infection. Radiology experts will have possibility to detect the COVID-19 virus using chest X-rays. For this detecting we have to analyse the negative chest X-rays along with normal or positive X-rays to identify whether the two lungs are effected or not. People with these symptoms may have COVID-19:

- Fever or chills
- Cough
- Shortness of breath or difficulty breathing
- Fatigue
- Muscle or body aches
- Headache
- New loss of taste or smell
- Sore throat
- Congestion or runny nose
- Nausea or vomiting
- Diarrhea

RQ3: Detecting Pneumonia Severity using chest X-rays Pneumonia is an infection that causes inflammation in one or both of the lungs. It can be caused by a virus, bacteria, fungi or other germs. The infection is usually acquired when a person breathes in air carrying germs. Your doctor may conduct a physical exam and use chest x-ray, chest CT, chest ultrasound, or needle biopsy of the lung to help diagnose your condition. Your doctor may further evaluate your condition and lung function using thoracentesis, chest tube placement or image-guided abscess drainage. Most at risk for developing pneumonia are young children or people over the age of 65. An x-ray exam will allow your doctor to see your lungs, heart and blood vessels to help determine if you have pneumonia. When interpreting the x-ray, the radiologist will look for white spots in the lungs (called infiltrates) that identify an infection. This exam will also help determine if you have any complications related to pneumonia such as abscesses or pleural effusions (fluid surrounding the lungs). Pneumonia can sometimes lead to serious complications, such as respiratory system failure, spread of infections, fluid surrounding the lungs, abscesses or uncontrolled inflammation throughout the body (sepsis). The condition can also be fatal, so it is important to seek immediate medical attention if you are experiencing these symptoms. Patients with pneumonia could have the following symptoms:

- cough that produces phlegm or sometimes blood
- fever
- shortness of breath or difficulty breathing
- chills or shaking
- fatigue
- sweating
- chest or muscle pain

V. Experiment Evaluation

A. Data preparation and Augmentation

We have used a large publicly available chest x-ray datasets in those folders we have a team of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh along with their collaborators from Pakistan and Malaysia in collaboration with medical doctors have created a database of chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images. This COVID-19, normal and other lung infection dataset is released in stages. In the first release they have released 219 COVID-19, 1341 normal and 1345 viral pneumonia chest X-ray (CXR) images. In the first update, they have increased the COVID-19 class to 1200 CXR images. In the 2nd update, they have increased the database to 3616 COVID-19 positive cases along with 10,192 Normal and 1345 Viral Pneumonia images [16] [17]. A dataset of 2373 Chest X-ray (CXR) images from Stony Brook Medicine. Each CXR has been scored by two radiologists. There are multiple CXRs per patient and temporal information is included [18]. Severity can be taken from RALO (Radiographic Assessment of Lung Opacity) dataset in which Radiological scoring was performed by three blinded experts [18]. A dataset of 3,875 images for training of pneumonia with 1341 normal chest x-rays. Testing with 390 images and 234 normal images for it. They have used 9 pneumonia images and 8 normal images for validation of model [19].

B. Data Augmentation

Data augmentation is a method that enables practitioners to significantly improve the variety of data for training models without actually collecting new data. The technique of increasing the number of training samples by transforming the images without losing semantic information is referred to as data augmentation, and it enables the image data to be free of bias. To prevent bias in the image data due to the similarity of the underlying image, basic modifications such as horizontal flipping, color space augmentations, and random cropping are frequently employed for model training [20, 21]. Additionally, data augmentation enables the model to learn a more diverse collection of features, which increases the dataset's size and assists in preventing the model from becoming overfitted.

The objective of the image pre-processing stage is to smother unwanted twists present in the picture, resize and normalize the image for further processing. There is numerous image pre-processing technique found in the previous literature based on the requirement of model building. Among them, image resizing, image normalization, and covert level to categorical are generally used techniques. In this study, images were resized to ensuring the same size and the same pixel using the Pillow python package. This study considers 64 x 64 pixel values for images. Besides, image normalization is how we adjust the pixel intensity to make the picture increasingly natural. Typically, most of the image pixels integrate values between 0 and 255. But, due to network architecture, it is better to perform all values between 0 and 1, which will be a good fit for the model building. This technique reduces the computational complexity during training the model. However, using below Equation 5 images were normalized values.

$$X_{norm} = X - \frac{X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X_{min}, X_{max} refer to the minimum and maximum pixel values.

C. Severity of the disease using RALO dataset

The model to check the severity of the disease for example negative and positive case x-rays if we take based on the disease the model will identify if its positive case it will define the severity of that disease inside other model. Here severity can be taken from RALO (Radiographic Assessment of Lung Opacity) dataset in which Radiological scoring was performed by three blinded experts: two chest radiologists (each with at least 20 years of experience) and a radiology resident. They staged disease severity using a score system, based on two types of scores (parameters): extent of lung involvement and degree of opacity [18]. Extent of lung: 0 = no involvement 1 = ¡25% involvement 2 = 25%-50% involvement 3 = 50%-75% involvement 4 = >75% involvement. Degree of Opacity: 0 = no opacity 1 = ground glass opacity 2 = consolidation 3 = mix of consolidation and ground glass opacity (>50% consolidation) 4 = white-out The total opacity score ranged from 0 to 8 (right lung and left lung together).

As stated in the algorithm below we will take each Chest X-ray image and will calculate the geographic and opacity of the disease if it is 2 dimensional image and will return the corresponding values to the given image.

Algorithm 1 Deducting Severity of the given Chest X-ray

Input: Human Chest X-ray
Output: Disease Severity from Chest X-ray

```
if (Image is 2D image) then
    By using RALO dataset
    Calculate the Geographic Extent of Infection
    Calculate the Opacity of Infection
end if
return Severityvalues
```

D. Classification of Pneumonia using ResNet50V2

ResNet50V2 is a modified version of ResNet50 that performs better than ResNet50 and ResNet101 on the ImageNet dataset. In ResNet50V2, a modification was made in the propagation formulation of the connections between blocks. ResNet50V2 also achieves a good result on the ImageNet dataset. The ResNet50V2 architecture consists of several residual blocks with each block having several convolutional operations. The implementation of skip connections makes the ResNet50V2 better than VGG. The skip connections between layers add the outputs from previous layers to the outputs of the stacked layers. This allows the training of deeper networks. One of the problems that ResNet50V2 solves is the vanishing gradient problem [22].In recent years, many researchers have focused on using CNN, which consists of three principal layers, namely, a convolutional layer, a pooling layer and a fully connected layer. The ResNet [23] architecture has two types of layers, namely, conv block and identity block, and these serve as shortcuts in residual blocks and are included in an order

ResNet50V2 takes an image of size 224 × 224 pixels. Pre-processing of the images was performed automatically by calling ‘preprocess input’ from the ResNet50V2 model in TensorFlow. The ‘preprocess input’ is fed into the ‘ImageDataGenerator’ from TensorFlow (Keras). ‘ImageNet’ weights are used for training. Out of the 5,242 images, 80% of the images are used for training. The VisionPro Deep Learning suite automatically selects the other 20% images for validation. Both the training and validation sets are randomly selected by the VisionPro Deep Learning suite. The user just needs to specify the train validation split. The maximum number of epoch counts was selected to be 10. There are options of selecting the minimum epochs and patience for which the model will train, but this was not selected.

E. Classification of COVID-19 using DenseNet121

DenseNet (Dense Convolutional Network) [24, 25] is an architecture which focuses on making the Deep Learning networks go even deeper, while at the same time making them more efficient to train, using shorter connections between the layers. DenseNet is a convolutional neural network where each layer is connected to all other layers that are deeper in the network, that is, the first layer is connected to the 2nd, 3rd, 4th and so on, the second layer is connected to the 3rd, 4th, 5th and so on. It does not combine features through summation but combines the features by concatenating them. So, the ‘ith’ layer has ‘i’ inputs and consists of feature maps of all its preceding convolutional blocks. It therefore requires fewer parameters than traditional convolutional neural networks.

F. Performance Parameters

To quantitatively evaluate the proposed model we use accuracy, sensitivity, specificity, precision, ROC [26,27] as performance metrics. **Accuracy:** The accuracy can be calculated in terms of positive and negative classes:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP (True Positives) is the number of correctly classified instances of the class under observation, TN (True Negatives) is the number of correctly classified instances of rest of the classes, FP (False Positives) is the number of miss-classified instances of rest of the classes and FN (False Negatives) is the number of miss-classified instances of the class under observation.

Recall/Sensitivity: it is the ratio of TP and TP + FN

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

Specificity: it is the ratio of TN and TN + FP Highlight

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

Precision: it is the ratio of TP and TP + FP Highlight

COVID-19 images		Pneumonia images		Total number of images
Train	*Test*	*Train*	*Test*	
2630	1030	3018	2224	8902

TABLE I
DATA SPLIT ACCORDING TO DATASET AND CLASS

		Predicted Class	
		Covid -19	Healthy
True Class	COVID-19	944	3
	Healthy	64	19

TABLE II
FINAL CONFUSION MATRIX TABLE OF THE TEST SET FOR THE CLASSIFICATION OF CHEST X-RAY IMAGES AS HEALTHY OR COVID-19 POSITIVE FOR DENSENET201

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

Receiver Operating Curve (ROC): plots the true positive rate (TPR) against the false positive rate (FPR) as shown in Figure 8.

G. Feature Extraction Steps

For extracting features from the X-ray images, we use the transfer learning concept discussed in Section II. Firstly, we select different CNN architectures that achieved excellent performance on the ImageNet dataset. Secondly, we choose different configurations, previously trained on ImageNet, from the selected CNN architectures. Thirdly, we remove any fully connected layers from these configurations, leaving only convolutional and pooling layers. These two types of layers are responsible for extracting features from the image, while the fully connected ones are responsible for classifying the features and, consequently, the image. Thus, removing these layers is necessary to turn a CNN into a feature extractor. After this step, the new output of the adapted CNN is a set of features extracted from an input image.

For each CNN configuration, we create a dataset composed of sets of features extracted from each image of the original datasets. In order to build a dataset, we first resize each image according to the input size required by the selected CNN. Then, each resized image is used as input to the CNN, and its set of features is extracted and stored in the corresponding dataset.

Table 1 shows the total 8902 dataset images divided for each model 2630 covid images took for training model same for pneumonia with 3018. We took 1030 images for Covid-19 testing and 2224 for pneumonia model testing.

1) Model Training: In this step, we use 80% of the dataset to perform the training of the model. We consider the setup for the hyperparameters presented in Table I to find the configuration of the classifiers on the training set. The classifiers that were configured for a random search perform a 20-iterations search. The hyperparameters for all classifiers, except for the Bayes classifier, are determined after 10-fold cross-validation. Then, each classifier has optimal hyperparameters, which are saved on the computer.

2) Model Testing: In this step, we perform a test in the remaining 20% of the sub-dataset using the saved classifiers. The system determines one class for each sample of the dataset. In addition, the metrics are calculated in this step.

VI. RESULTS AND DISCUSSION

In this section, we investigate the results achieved by combining the features extracted by both CNNs, applying transfer learning, and the classifiers.

Based on the dataset from table 1 we have tested pneumonia and we got the result that was obtained based upon the designed convolutional neural network model was found to be exceptional. The Training accuracy of the model aimed to be 100while the Testing accuracy of the model was found to be 97.98%. For the hypothesis of the model, the average precision and recall of the model was considered and it was found that the precision and recall of both the classes were quite decent. Figure 4 shows the confusion matrix that has been generated based on the predictive performance of convolutional neural network model. Based on the confusion matrix it can be observed that the model underwent a very little percentage of mispredictions and also demonstrated an overall accuracy of 91.5% with an average precision and recall of 90% and 93% respectively.

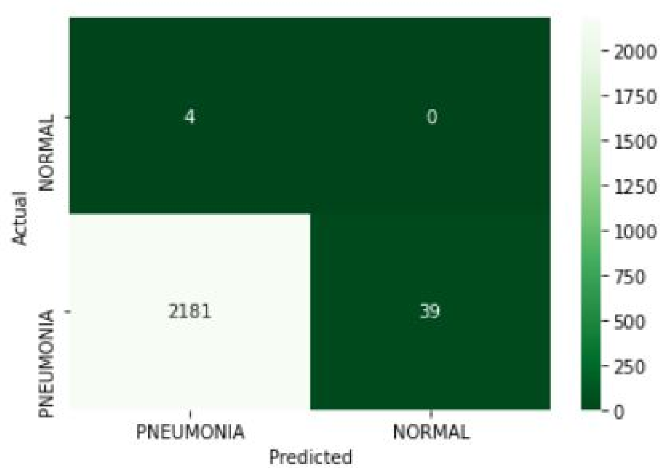

Fig. 4. Confusion Matrix for Pneumonia Dataset

Table II presents the confusion matrix for the COVID-19 model DenseNet201 testing in such a way that we got overall accuracy of 94.85% with an average precision and recall of 92% and 96% respectively.As per the results, the model got deducted 944 test images has covid out of 1030 images and 19 images as healthy images and remaining we got the mixed results.

After deducting the disease, the following Figure 5 shows how the severity of the pneumonia x-chest ray will display as a result when compared to normal image in Figure 7. As discussed in section V we conclude that this chest X-ray image patient has high severity immediate attention or consultation with doctor is require. On the same side we have severity of the Covid-19 is deducted in the Figure 6 when compare to the normal image this patient has moderate to high severity of the infection its better to consult a Doctor as soon as possible.

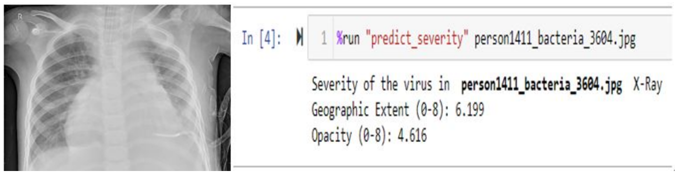

Fig. 5. severity In Pneumonia Chest X-Ray

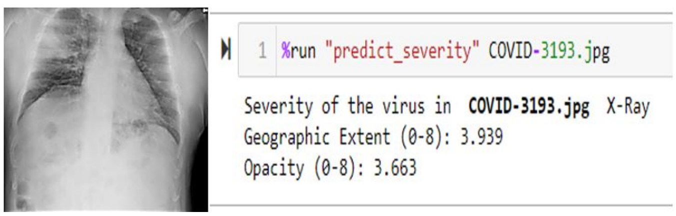

Fig. 6. severity In Covid-19 Chest X-Ray

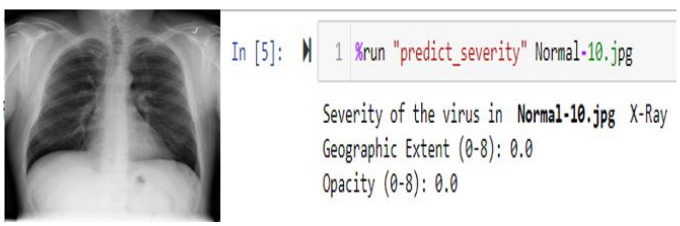

Fig. 7. severity In Normal Chest X-Ray

Figure 8 shows the ROC curve of the Covid-19 dataset, the relationship between clinical sensitivity and specificity for every possible cut-off (as discussed in Section V). As a result, we got highest accuracy rate of 95% of the covid-19 class. Thus every point on the ROC curve represents a chosen cut-off even though you cannot see this cut-off. What you can see is the true positive fraction and the false positive fraction that you will get when you choose this cut-off.

VII. Conclusion

In Conclusion, we have been studied widely to identify and handle noise in data sets.Identification and handling noise is important for researchers and practitioners to accurately handle their data and predict future trends. After removing the noise inside the image we proposed work that implements a Resnet50V2 and DenseNet121 architecture to examine the chest radiograph images. The task considered in this work is to examine and classify the image dataset into normal/pneumonia and normal/Covid-19 category. Apart from the classification we have identified the severity of the disease inside each and every image that can understand easily that can help to

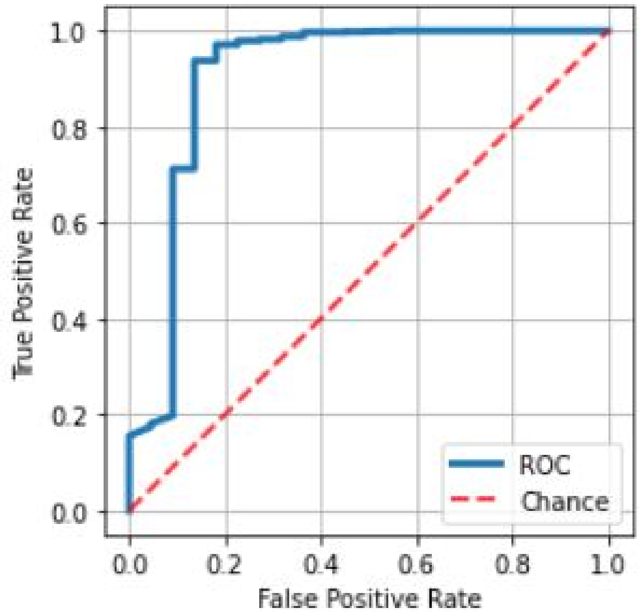

Fig. 8. ROC Curve of Covid-19 Dataset

choose the right treatment and for prevent the quick spread of the diseases. The proposed method has not undergone in deep clinical study. Thus, it does not replace a medical diagnosis since a more thorough investigation could be done with a larger dataset. Under those circumstances, our work contributes to the possibility of an accurate, automatic, fast, and inexpensive method for assisting in the diagnosis of COVID-19 and pneumonia through chest X-ray images.

In Future, We have to work on different number of layers it might be automated or manual layers for the models to get more accuracy and we are also planning to work on the web application that will take the image as an input and give the severity of the specific disease and also we will to try display the severity of the disease inside the lung with different color so, that illiterate people can also understand it more easily.

VIII. Acknowledgement

I would like to show my gratitude to the Dr. Sabah Mohammed from Department of Computer science, Lakehead University for sharing his pearls of wisdom with me during the course of this COMP-5800 Research Project. The data and code of this project will see in below GitHub Repository https://github.com/KutalaMohanPrasad/Mohan_COMP5800YDG_Research_Project

References

[1] World Health Organization (2020) Naming the coronavirus disease (covid-19) and the virus that causes it. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-(covid-2019)-and-the-virus-that-causes-it

[2] World Health Organization (2020) The continuing 2019-ncov epidemic threat of novel coronaviruses to global health - the latest 2019 novel coronavirus outbreak in wuhan, China. https://pubmed.ncbi.nlm.nih.gov/31953166/

[3] W. Luo, "Efficient removal of impulse noise from digital images," IEEE Trans. Consumer Electron., vol. 52, no. 2, pp. 523-527, May 2006.

[4] S.W Chiu, J.H Wang, K.H Chang, T.H Chang, C.M Wang, C.L Chang, C.T. Tang, C.F. Chen, C.H.Shih, H.W. Kuo and L.C. Wang, 2014. A fully integrated nose-on-a-chip for rapid diagnosis of ventilator- associated pneumonia. IEEE transactions on biomedical circuits and systems, 8(6), pp.765-778.

[5] S. Zhang and M.A. Karim, "A new impulse detector for switching median filter," IEEE Signal Process. Lett., vol. 9, no. 11, pp. 360-363,Nov. 2002

[6] Kenny Kal Vin Toh, Student Member, IEEE, Haidi Ibrahim, Member, IEEE, and Muhammad Nasiruddin Mahyuddin, Member, IEEE "Salt-and-Pepper Noise Detection and Reduction Using Fuzzy Switching Median Filter"

[7] J. S. Lee,L. Jurkevich,P. Dewaele,P. Wambacq A. Oosterlinck "Speckle filtering of synthetic aperture radar images" Published online: 19 Oct 2009

[8] Shivani Guptaa, Atul Guptab Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review

[9] Semi-Supervised Learning of the Electronic Health Record for Phenotype Stratification Brett K. Beaulieu-Jones

[10] Arjun Sarkar, Joerg Vandenhirtz, Jozsef Nagy, David Bacsa, Mitchell Riley Identification of Images of COVID-19 from Chest X-rays Using Deep Learning IEEE trans

[11] Amit KumarJaiswal PrayagTiwari SachinKumar DeepakGupta AshishKhanna Joel J.P.C.Rodrigues Identifying pneumonia in chest X-rays: A deep learning approach

[12] Nilanjan Dey , Yu-Dong Zhang , V. Rajinikanth, R. Pugalenthi , N. Sri Madhava Raja "Customized VGG19 Architecture for Pneumonia Detection in Chest X-Rays" IEEE Trans

[13] Varela-Santos, Sergio, and Patricia Melin. "A New Approach for Classifying Coronavirus COVID-19 Based on Its Manifestation on Chest x-Rays Using Texture Features and Neural Networks." Information Sciences 545 (2021): 403–14. https://doi.org/10.1016/j.ins.2020.09.041.

[14] Joseph Paul Cohen Stanford University; Beiyi Shen; Almas Abbasi; Mahsa Hoshmand-Kochi; Samantha Glass; Haifang Li; Matthew P Lungren; Akshay Chaudhari; Tim Q Duong " Stonybrook Radiographic Assessment of Lung Opacity (RALO) dataset is here: Pneumonia severity scores for 2373 images" https://zenodo.org/record/4634000.YX2YGp7MJPY

[15] Waring, Jonathan, Charlotta Lindvall, and Renato Umeton. "Automated Machine Learning: Review of the State-of-the-Art and Opportunities for Healthcare." Artificial Intelligence in Medicine 104 (2020): 101822. https://doi.org/10.1016/j.artmed.2020.101822.

[16] Jones, Jeremy. "Chest X-Ray Review: ABCDE: Radiology Reference Article." Radiopaedia Blog RSS. Accessed December 6, 2021. https://radiopaedia.org/articles/chest-x-ray-review-abcde.

[17] Paul M (2020) Kaggle chest X-ray images (pneumonia) dataset. https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia/

[18] Joseph Paul Cohen Stanford University; Beiyi Shen; Almas Abbasi; Mahsa Hoshmand-Kochi; Samantha Glass; Haifang Li; Matthew P Lungren; Akshay Chaudhari; Tim Q Duong " Stonybrook Radiographic Assessment of Lung Opacity (RALO) dataset is here: Pneumonia severity scores for 2373 images" https://zenodo.org/record/4634000.YX2YGp7MJPY

[19] James C. Church, Yixin Chen, and Stephen V. Rice Department of Computer and Information Science, University of Mississippi, "A Spatial Median Filter for Noise Removal in Digital Images", IEEE, page(s): 618-623, 2008.

[20] Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):60. doi: 10.1186/s40537-019-0197-0.

[21] Wang J, Perez L. The effectiveness of data augmentation in image classification using deep learning. Convolutional Neural Netw Vis Recognit 2017;11:1–8.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," CoRR, vol. abs/1409.1556, pp. 1–14, Sep. 2014

[23] B. M. Rocha et al., "A respiratory sound database for the development of automated classification," in Precision Medicine Powered by pHealth and Connected Health. Singapore: Springer, 2018, pp. 33–37.

[24] Veit A, Wilber M, Belongie S. Residual networks behave like ensembles of relatively shallow networks. (2016) In: Advances in neural information processing systems. Cambridge: MIT Press.

[25] S Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings—30th IEEE conference on computer vision and pattern recognition, CVPR 2017. (2017). https://doi.org/10.1109/CVPR.2017.

[26] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," Radiology, vol. 143, no. 1, pp. 29–36, 1982.

[27] M.A. Khan , S. Kadry , M. Alhaisoni , Y. Nam , Y. Zhang , V. Rajinikanth , M.S. Sar- fraz , Computer-Aided Gastrointestinal Diseases Analysis From Wireless Cap- sule Endoscopy: A Framework of Best Features Selection, IEEE Access 8 (2020) 132850–132859 .