

Image Processing Techniques using CXR Images**A comparison of datasets available on study of Chest X-ray images
regarding COVID-19 analysis****Anup Adhikari, Bhawana Poudel, Saroj Giri^{1,1,1}**^a*anup.adhikari@gandakiuniversity.edu.np*^b*bhawana.poudel@gandakiuniversity.edu.np*^c*saroj.giri@gandakiuniversity.edu.np*

Abstract*Keywords:*

1. Introduction*1.1. Background*

The current corona virus disease 2019 (COVID-19) pandemic is very lamentable because the second wave was more dangerous than the first wave. It is seen that countries like Nepal, India is one of the most affected countries in the second wave of severe acute respiratory syndrome corona virus 2 (SARS-CoV-2). The virus is spreading very fast and can be contracted at all ages, which can lead to serious illness. As a highly contagious viral disease caused by SARS-CoV-2, COVID-19 has wreaked havoc on the world's demography, killing over 2.9 million people globally, making it the most significant global health epidemic since the 1918 influenza pandemic [?]. Patients older than 60 years, small children and persons with medical problems, should be considered at a higher risk According to the estimates of the World Health Organization, there are approximately 306,091,673 COVID-19 cases world wide and 831748 cases in context of Nepal. When this virus attacks the human body, there may be two scenarios: mild and severe. At the onset of the corona virus infection, one issue is certain: the virus has a negative effect on lung health. As a result, doctor's advise patients to keep track of their oxygen levels with oxygen meter so that any abnormalities can be detected and treated early [?]. The virus normally attacks the lungs in the human body and causes pneumonia in severe cases. Subsequently, it decreases the oxygen level instantly. Because this virus has no cure thus far, the only solution before a vaccine is to prevent the spread of the virus. Therefore, tests and trace is the only solution thus far. Normally, the polymerase chain reaction (PCR) test is widely used in medical science for testing. As it is time-consuming and costly. Therefore, an alternative testing is required so that infected people can be identified quickly and quarantined or isolated. To date, some deep learning approaches have been used to identify viruses. However, the results of these deep learning techniques are not sufficient to deal with a medical-related diagnosis system.

X-radiation or X-ray is an electromagnetic form of penetrating radiation. These radiations are passed through the desired human body parts to create images of internal details of the body part. The X-ray image is a representation of the internal body parts in black and white shades. X-ray is one of the oldest and commonly used medical diagnosis tests. Chest X-ray is used to diagnose the chest-related diseases like pneumonia and other lung diseases [?], as it

provides the image of the thoracic cavity, consisting of the chest and spine bones along with the soft organs including the lungs, blood vessels, and airways.

Although rapid point-of-care COVID-19 tests are expected to be used in clinical settings at some point, for now, turnaround times for COVID-19 test results range from 3 to more than 48 hours, and probably not all countries will have access to those test kits that give results rapidly [?]. Deep ConvNets were applied in several image recognition applications with high accuracy, and this increased its reliability for future research

A research [?] was conducted to classify CXR images into three groups: a transfer learning-based CNN model was used for COVID-19, non-COVID-19, and regular pneumonia. CXR images have been used as a sample dataset because X-ray equipment is low cost and time-efficient, as well as small and available in almost every clinic. Therefore, fewer developing countries can benefit from this research. This system will help detect corona virus from CXR images within the shortest possible time. One of the most common radiological tests is chest radiography. CXR analysis involves the detection and localization of thoracic illnesses. This will reduce the pressure on PCR testing, which is costly and time-consuming. False negatives were a common issue in PCR tests results, which is not helpful for the current situation.

1.2. Rationale of the Study

The virus is spreading very fast and can be contaminated to all ages, which can lead to serious illness. It has been assumed that third wave may arise soon. So, this research may help to detect patient with infection. Normally, the polymerase chain reaction (PCR) test is widely used in medical science for testing. PCR test cannot be available at all possible so X-ray methods can be a good alternative.

However, because the number of cases is increasing rapidly, it has become nearly impossible to perform enough tests through PCR, as it is time-consuming and costly. Therefore, an alternative testing is required so that infected people can be identified quickly and quarantined or isolated. To date, some deep learning approaches have been used to identify viruses. However, the results of these deep learning techniques are not sufficient to deal with a medical-related diagnosis system. In context of Nepal it can play an important role to detect people diagnosed with COVID-19 [?]. The financial costs of the laboratory kits used for diagnosis, especially for developing and underdeveloped countries, are a significant issue when fighting the illness.

1.3. Literature Review

The novel coronavirus SARS-CoV-2 has been in the ecosystem of the world after it was discovered in Wuhan, China [?] with the name of coronavirus disease 2019 (COVID-19); previously 2019-nCoV. Though Corona Virus has been prevalent in the world before Wuhan also, this novel strain has posed a serious devastating effects on global well-being of human population, by creating mild and critical symptoms to infected individuals [?].

The advance development of swab-test facility by real-time reverse transcription-polymerase chain reaction (RT-PCR) could detect SARS-CoV-2 RNA from respiratory specimens. This screening approach is a time-consuming and difficult manual process that takes 4-6 hours to acquire results, which is quite a while contrasted with the rapid spreading pace of COVID-19 [?]. The fastest antigen test, Rapid Diagnostic Test(RDT), would not be relied upon to confirm the SARS-CoV-2 presence by many of the countries and is used only in case of public events and fairs.

The other diagnosis methods of COVID-19 include clinical symptoms analysis, epidemiological history assessment and radiographic images (Computed Tomography(CT)/ Chest Radiograph(CXR)). The clinical symptoms analysis includes tests of fever, cough, dyspnea and respiratory failure. Epidemiological history assessment involves the governance body taking decisions on critical geographical areas for prominence of COVID-19 infected population. The radiological imaging technique could be an important diagnostic tool for COVID-19 to assess the presence of SARS-CoV-2 and damages done by the virus [?]. Furthermore several groups have reported deep learning techniques using X-ray images for detecting COVID-19 pneumonia [?]. A public database was created for the analysis and development of machine learning techniques for different scientific and medicinal pioneers [?].

1.4. Objectives of the Study

The objective of the research study is to:

- To detect COVID 19 detection from CXR images using CNN.

1.5. Scope and Limitations of the Study

1.6. Theoretical Framework

2. Methods and Methodology

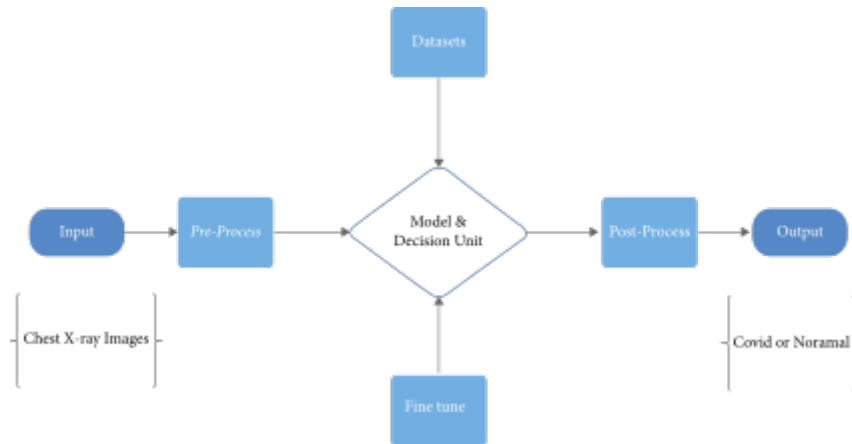


Figure 1. Block Diagram [?] of the Proposed System

2.1. Data Collection/Input:

The data set is collected from different sources. We have proposed a framework that can effectively detect COVID-19 cases by evaluating the chest x-ray images. The image data are collected from different websites such as Radiopedia.org and Figure1.com. The random images are studied and divided into Normal image and Covid image. This data set is further converted into training and testing with the split of 70% data for training of proposed deep learning model and 30% data for testing purpose.

2.2. Data pre-processing

Due to inconsistency of the dataset, the X-ray images are of different sizes. So, the all the images are converted to the same size. Also, splitting dataset process and data augmentation technique comes under data pre-processing. Final input to the proposed model is prepared.

Algorithms:

Convolutional Neural Network

Convolutional Neural Network is one of the popular deep learning algorithms which are used for the analysis of image data. CNN make the use of learned features with input data, and uses convolutional layers, and make this architecture well suited for processing the images.

There are three types of layers in Convolutional Neural Networks.

1) Convolutional Layer: The convolutional layer, considered as a main layer of a CNN, performs the operation called “convolution”. Kernels in the convolutional layer are applied to the layer inputs. In CNN, the input layer neurons is connected to the neuron hidden layer.

2) Pooling Layer: The pooling layer is used to reduce the dimensionality of the feature map. There will be multiple activation pooling layers inside the hidden layer of the CNN.

3) Fully-Connected layer. The input to the fully connected layer is the output from the final Pooling or Convolutional Layer, which is flattened and then fed into the fully connected layer.

3. Testing CNN:

After making the model, the external X-ray images are tested using the classifier model. And the image is categorized as Covid infected or not.

Classification Performance Evaluation:

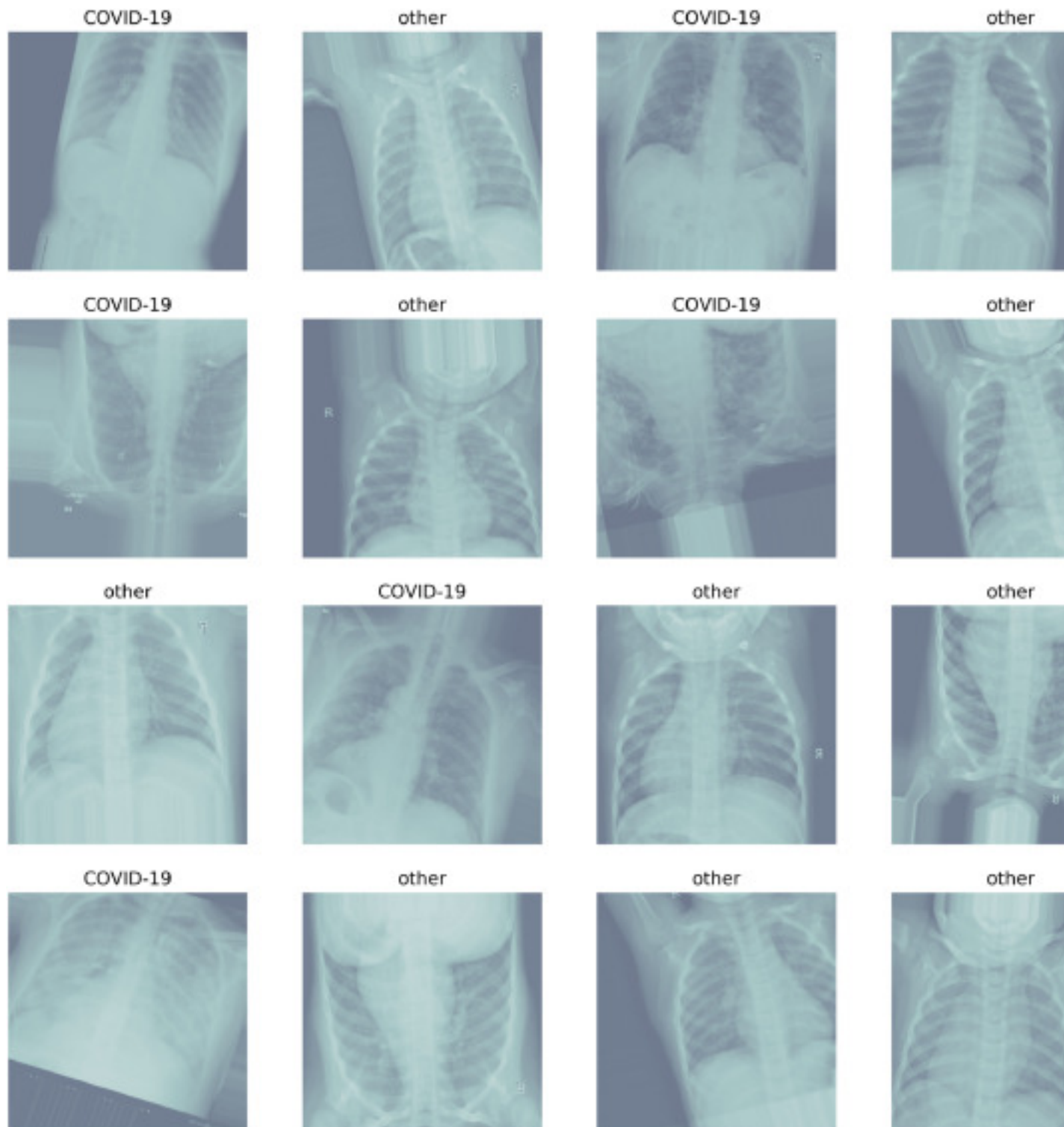


Figure 2. Figure: Sample of the labelled X-rays after data augmentation taken from the combined data set of COVID-19 patients and normal patients

Figure 3. Convolutional Neural Network

Figure 4. Convolutional Neural Network [?]

After diagnosing the X-rays in the test set, the proportion of each diagnosis is obtained. The performance evaluation indices are:

Accuracy:

Accuracy is a ratio of correctly predicted observation to the total observations. It is represented as:

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

Precision:

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate. It is represented as :

$$\text{Precision} = \frac{TP}{TP+FP}$$

Recall:

Recall is the ratio of correctly predicted positive observations to the all observations in actual class. It is represented as:

$$\text{Recall} = \frac{TP}{TP+FN}$$