Machine Learning Algorithm for Trend Analysis in Short term Forecasting of COVID-19 using Lung X-ray Images

To cite this article: K Sujatha *et al* 2023 *J. Phys.: Conf. Ser.* **2467** 012001

Machine Learning Algorithm for Trend Analysis in Short term Forecasting of COVID-19 using Lung X-ray Images

Sujatha. K[1], N.P.G. Bhavani[2] , D. Kirubakaran[3], N. Janaki[4], G.Victo Sudha George[5], Su-Qun Cao[6],A. Kalaivani[7]

[1]Department of BME/Electrical and Electronics Engineering, Dr. M.G.R. Educational and Research Institute, Chennai-600095, Tamilnadu, India
[2]Dep. of ECE, Saveetha Sch. of Eng., Saveetha Ins. of Med. and Tech. Sci., India.
[3]Dept of EEE, St.Joseph's Institute of Technology, Chennai. India.
[4]Department of EEE, School of Eng., VISTAS, Chennai-600095, Tamilnadu, India.
[5]Dep. of CSE, Dr. MGR Educational & Research Institute,Chennai, Tamil Nadu, India.
[6]Faculty of Elec. Infor. Eng., Huaiyin Institute of Technology, JiangSu, China.
[2]Department of CSE, Rajalakshmi Engineering College, Chennai. India.
Sujatha. K, Email: sujathak73586@gmail.com, sujatha.eee@drmgrdu.ac.in

Abstract. With the development of medical technology, the diagnosis of lung diseases relies more on the determination of medical images. With increasingly huge data, a powerful data processing model is urgently needed to provide favorable support for this field. The goal of this study is to develop a computer-assisted method to identify COVID-19 from X-ray pictures of the lungs at the very beginning of the disease. The architecture is implemented as a software system on a computer that can assist in the affordable and accurate early identification of cardiac illness. The performance of CNN architecture is best among all other classification algorithms to detect COVID-9 from Lung X-ray images. The datasets consist of COVID-19 established cases for 4 weeks which included the X-ray images of the chest. Then the distribution of the data was examined according to the statistical distribution. For this prediction, time series models are used for forecasting the pandemic situation. The performances of the methods were compared according to the MSE metric and it was seen that the Convolutional Neural Networks (CNN) achieved the optimal trend pattern.

Keywords: Convolutional Neural Networks, Lung X-ray Images, COVID-19, Forecasting efficiency

1. Introduction

Since the discovery of the new corona virus in late 2019, the novel corona virus disease 2019 (COVID-19, hereinafter referred to as COVID-19) outbreak has rapidly spread around the world. As of April 19, 2020, more than 200 countries and regions around the world have confirmed cases of new coronary pneumonia, and the cumulative number of confirmed cases has reached 2.16 million. The medical systems of various countries and regions are facing huge challenges, and many countries have experienced shortages of supplies and shortages. Patients with underlying diseases are prone to develop severe symptoms and develop respiratory failure and other symptoms [1]. This feature makes new coronary pneumonia patients often have cluster infections, and a large number of infections in a short period of time. into the hospital. The key to controlling the epidemic is to distinguish patients with common pneumonia from patients with new coronary pneumonia, and accept all patients with new coronary pneumonia to block the transmission path of the virus.

Since Artificial neural networks (ANN) have been popular in the short term COVID-19 forecasting from X-ray images of lungs by the research community over the past decade, several ANN based models is being developed for obtaining effective and efficient use of resources. The knowledge of applying combination of dynamic clustering method with deep learning ANN methods with new to short term COVID-19 forecasting is proposed here. An effective new combination Dynamic Clustering Method with deep learning Artificial Intelligence (AI) methods along with a new weighted aggregation mechanism is my plan of study as one of the best and effective techniques in terms of precision for short term forecasting of COVID-19 [2, 3]. In[4] used the keras deep learning library to build a network model to complete the discrimination of lung X-ray images. During the training of the network model, the data augmentation operation is performed on the data using the machine learning image augmentation library. In[5] designed a progressive multi-scale information fusion mechanism for the problem of many noise elements in pneumonia image samples and the indistinguishability of positive and negative samples. The existing convolutional neural network (CNN) structure was improved, and an end-to-end pneumonia image automatic recognition model was established. The model selects the corresponding fusion strategy for the output of different layers of CNN, and comprehensively utilizes the information of different layers of the image. In[6] proposed a multi-channel pneumonia pretrained network model to facilitate the diagnosis of COVID-19 lung X-ray images. Three Resnet-based network models were pre-trained, namely classification network models for normal or diseased, pneumonia or non-pneumonic, and COVID-19 or non-COVID-19 individuals. Finally, the three models were integrated and fine-tuned using X-ray images from 1579 normal patients, 4245 pneumonia patients, and 184 COVID-19 patients. In[7] proposed a convolutional neural network with Faster-RCNN structure to detect the new coronavirus (SARS-CoV-2) infection foci, solving the high false negative rate of reverse transcription polymerase chain reaction (RT-PCR) detection results problem and improve the detection sensitivity. The experimental study selected 420 CT images and 2697 features of 7 SARS-CoV-2 infected patients, and 200 CT images of healthy individuals for analysis. In [8] proposed an improved deep residual network to classify lung computed tomography (CT) images, and reduced the neural network's requirement for large amounts of data by means of transfer learning. In [9] used the LUNA16 data set and proposed a method for detecting pulmonary nodules based on convolutional neural networks, and adopted a set of preprocessing schemes for lung CT images. The accuracy of the proposed model reached 92.3%. In [10] obtained 50 X-ray images of new coronary pneumonia from the new coronary pneumonia data set, obtained 50 normal lung X-ray images from the chestX-rayimages data set, and constructed a data of 100 X-ray images Set, using the ResNet50 model for classification, the accuracy rate reaches 98%. In [11] proposed a model combining UNet++ and ResNet50 for the diagnosis of new coronary pneumonia based on a CT image dataset containing 723 positive and 413 negative tests for COVID-19, achieving an accuracy of 97.4%. . In [12] systematically reviewed the application of current artificial intelligence technology in the diagnosis of new coronary pneumonia, and analyzed 87 papers on new coronary pneumonia research, covering all processes of medical image analysis technology involved in new coronary pneumonia, including Image acquisition, segmentation, diagnosis, etc. Although CT examination is a relatively advanced medical imaging examination technology, it can help radiologists make detailed analysis [13]. However, plain X-rays are cheap and more convenient than CT examinations, which can help radiologists quickly determine whether a patient is a patient with new coronary pneumonia. Therefore, it is very necessary to use deep learning methods to automatically identify plain X-ray images. .

In this paper, a novel coronavirus pneumonia detection method based on an improved convolutional neural network is proposed. The detection method uses image preprocessing technology combined with related algorithms of convolutional neural network in deep learning to detect novel coronavirus pneumonia. The goal of this suggested study plan is to develop an accurate integrated forecasting framework with a focus on a model for short-term load that is easily connectable to other forecasts.

This suggested study effort dissects the main approaches to short-term load forecasting that have been used recently and gathers the essential components to establish a novel way for analyzing short-term load forecasting issues and creating short-term load forecasting models.

2. Related background

2.1. Convolutional Neural Network Components

Imaging detection relies on the evaluation of radiologists, but a large number of patients pouring into the hospital in a short period of time will lead to manpower shortage and work fatigue of radiologists. problems, and automated testing for COVID-19 could help doctors address those issues. In recent years, with the disclosure of medical image data, the application of deep learning in medical imaging has become more and more extensive, especially the convolutional neural network has achieved excellent results in medical image processing [4].

The convolutional neural network (CNN) was created on the basis of the neural network algorithm, which is a standard technique in the field of machine learning. Although the automatic training of the model is completed using the gradient descent approach, there are still significant discrepancies between the convolutional neural network and the neural network algorithm. Convolutional neural networks can process images straight from input, whereas neural networks must first flatten the image data. During the training process of the model, the convolutional neural network can also use the unique structure in the model to abstract and enhance the features of the image. In addition, in terms of the parameters contained in the model, the convolutional neural network adopts the receptive field and weight sharing method, which greatly reduces the number of its own parameters.

Figure 1, shows CNN consists of five major parts, including: input layer, convolution layer, pooling layer, fully connected layer, and output layer. Sometimes for convenience, the fully connected layer and the output layer are combined as a neural network layer. The specific functions of most of them are as follows.

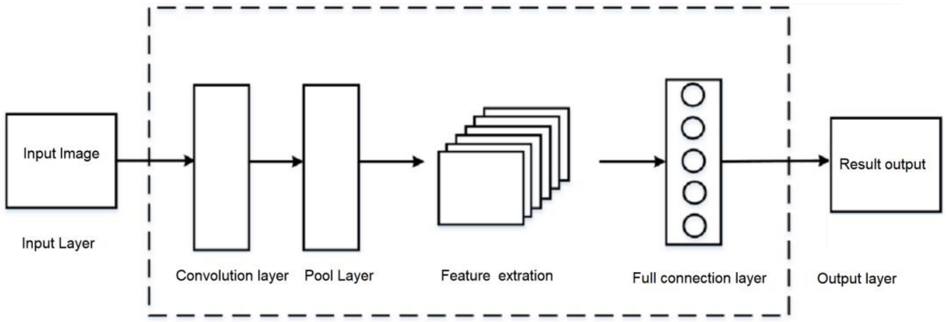

Figure1. Theconvolution neural network architecture

2.2. Image scaling

The resolution of lung X-ray images is not uniform across datasets. This paper uses an interpolation algorithm to scale different training samples to a fixed size required by a deep learning model. Figure 2 shows the principle is shown in . In the figure, Q_{11}, Q_{21}, Q_{12}, and Q_{22} are the four pixel points of the original image, and point P is the result value after the an interpolation operation.

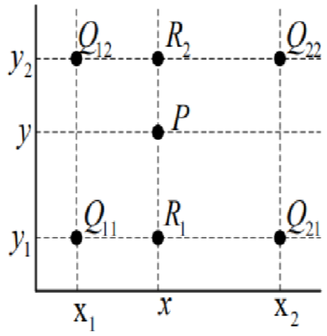

Figure 2. Schematic diagram of linear interpolation

In the figure, R_1 and R_2 are the result values after the linear interpolation operation by

$$f(P) \approx \alpha f(M) + \beta f(N) \quad (1)$$

Where taking point R_1 is determined by two adjacent pixel points Q_{11} and Q_{21}. Among them, the weight of point Q_{11} represents the ratio of the $\frac{x_2-x}{x_2-x_1}$, distance between Q_{11} and R1 to the x-axis distance between Q_{21} and Q_{11}.

The weight of Q_{21} is the same, and the specific by

$$f(P) \approx \frac{x_2-x}{x_2-x_1} f(Q_{11}) + \frac{x_2-x}{x_2-x_1} f(Q_{21}) \quad (2)$$

After determining R1 and R2, perform linear interpolation again to determine point P by

$$f(P) \approx \frac{y_2-y}{y_2-y_1} f(R_1) + \frac{y_2-y_1}{y_2-y_1} f(R_2) \quad (3)$$

3. Proposed Method

The Machine Learning has substantially enhanced the therapy, dosage levels, forecasting, prediction and has been a key element in vaccine growth process. The inevitability of SIR model has been employed to estimate the end-time of the second wave of the pandemic in different states in India. The number of patients who will get newly infected with COVID-19, mortality rates, and estimations of recovered COVID-19 in the upcoming 10 days may all be accurately predicted using the Linear Regression Analysis (LR). The total active cases of corona across India were accurately forecasted using Prophet model analysis. In addition, the ARIMA and SARIMA analysis found suitable for predicting the occurrence of new and total deaths of Covid-19 during the later part of the year 2020. The models such as AIC, MAPE, MAE have been applied to forecast on epidemiologic data to avert the epidemiological pattern of the occurrence in India. The FB Prompet model has performed on forecasting the future prevalence of Pandemic situation in India.

The training of CNN for COVID-19 forecasting is depicted in Figure 3. Figure 4, shows the high level features are extracted from the presented set of inputs. The low level features are extracted in the first Convolutional Layer. The adding of extra Convolutional Layers helps to extract high level features. Valid Padding is a case where dimension reductionality takes place in comparison with the inputs and on the other hand the dimension remains same or increases and is called as Same Padding. The 168x4 is dimensionally reduced to 5x5x4 by applying a 5x5 kernel over it.

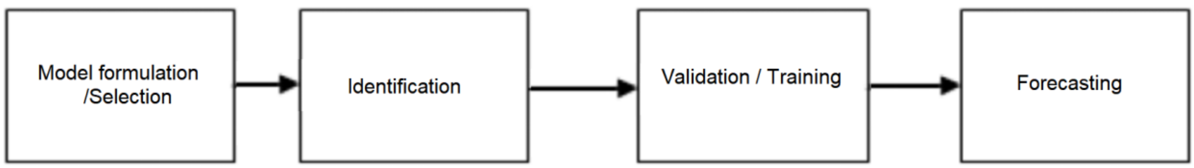

Figure 3. The main process of CNN for COVID-19 forecasting from lung X-ray images

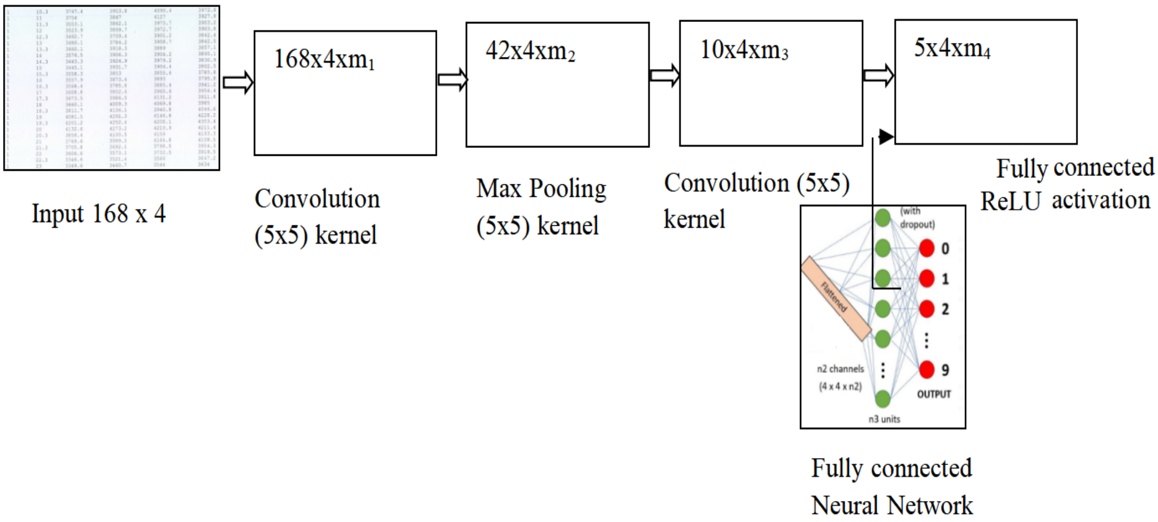

Figure 4. Proposed Architecture for CNN

The main process as follows:-

(1) Input layer: Before inputting data into the network model for training, this paper needs to consider the dimension, size, and number of channels of the data sample. First, we determine the type of CNN according to the dimension of the data sample. The data used in the experiment in this paper are all color images, so 2D-CNN is selected as the training model. After the input dimension is determined, since the sizes of the images collected in the datasets are not exactly the same, the CNN model needs to uniformly specify the size of the input sample data.

(2) Convolution layer: The convolution layer contains several n*n convolution kernels, each of which can extract different features such as image level, edge, color, etc., so the CNN model uses the convolution layer to abstract data . In the process of network model training, the model uses the back-propagation algorithm to adjust the size of each parameter of the convolution kernel in the convolution layer. Therefore, by continuously adjusting the size of the parameters in the convolution kernel, the model can abstract different information, and the model can combine more information features by increasing the number of convolution kernels, so as to rely on more information to evaluation the corresponding category of the object. In the deep learning network, the convolution layer can further extract the abstract data of the previous layer by cascading. Generally speaking, the number of convolution layers in the model determines the model's ability to extract composite features from the input image by

$$X_j^l = f\left(\sum_{i\in M_j} X_j^{l-1} \times K_{ij}^l + b_j^l\right) \quad (4)$$

Where X_j^l represents the jth feature response map of the l_{th} layer, b_j^l is the bias, and K_{ij}^l is the convolution kernel function. In the curly brackets is the image convolution process of this layer, in addition, it also includes f, the activation function. In the process of processing, the nonlinear fitting ability of the model is enhanced by adding bias and activation functions.

(3) Pooling layer: Convolutional layer output can be reduced by adding a pooling layer, which is typically applied after the convolutional layer. By reducing the size of feature images, the computational efficiency of subsequent operations of the model is improved. At the same time, this layer can also retain the main feature information of the image. For the linearly transformed lung X-ray images, the model uses the operation characteristics of the pooling layer in the training process to clearly distinguish the relationship with the original lung X-ray images, effectively enhancing the robustness of the network model.

In the process of iterating the network model, when the SoftMax function completes the category probability prediction for the input data, we need to define an evaluation standard for the classification result of the SoftMax classifier to measure the performance of a model. We usually use this The evaluation criterion is designed as a loss function is expressed by

$$j(\theta) = \left[\sum_{i=1}^{m}\sum_{j=1}^{k} y^{i}\log\frac{e^{\theta_j^T x^i}}{\sum_{l=1}^{k} e^{\theta_j^T x^i}}\right] \quad (5)$$

(4) Fully connected layer: The fully connected layer has the same structure and function as the traditional neural network. After the previous convolution operation and pooling operation respectively complete the abstraction and dimensionality reduction of image features, this layer classifies the finally generated image feature matrix. and return tasks. This layer is usually constructed in a fully connected manner, so the number of parameters in this layer accounts for the largest proportion of the entire network model. However, compared with traditional neural networks, CNN greatly reduces the number of parameters in this layer by introducing convolutional layers and pooling layers, which improves the efficiency of model training.
(5) Output layer: The output layer is the output of the entire CNN model. For the classification task, the output layer plays the role of a classifier in the entire network model. This layer uses the feature vector output by the fully connected layer to complete the classification task, so the number of labels for the final classification is the number of the final neural nodes. Usually, the model will select the SoftMax function as the output layer. We can get the probability value of each label with the SoftMax function, and then select the label value with the highest probability as the result value of the final classification. At the same time, most machine learning algorithms also have classification capabilities, so they can be used as the output layer of the network model. Subsequent chapters will describe in detail, here is a brief introduction to the principle of the SoftMax function.
As shown in Figure 5, the area in theframe is the pooling area. According to the above maximum pooling rule, this paper will select 7 as the representative value in this feature range. On the entire feature image, by setting the pooling area size and the step value of the pooling process to complete the operation of the entire feature image. Similarly, the operation method of average pooling is to add and average the values in the area, the result is 5, and other areas are calculated in turn.

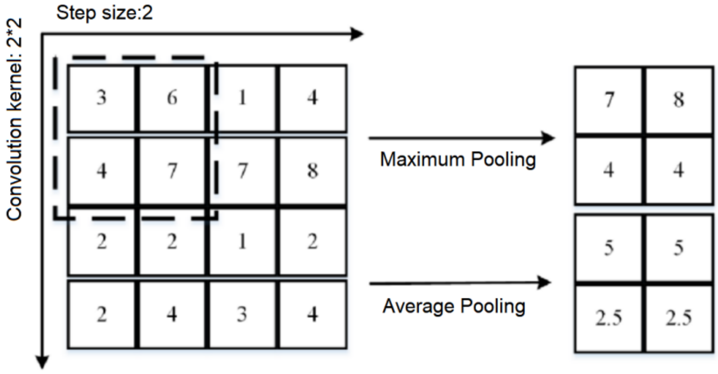

Figure 5. Convolution kernel for Pooling rule

4. Results and Discussion

In the normal environment, the All India Medical Council knows the number of patients affected by COVID-19. Due to rapid modernization, forecasting the patients affected with COVID-19 in the city is very difficult to find and offer a proper treatment along with isolation. To overcome this problem, an artificial neural network approach has been adopted. The accuracy of model training depends to a large extent on the dataset. The data related to COVID-19 is obtained from the All India Medical Council. The data includes data extracted from the X-ray images of the lungs which includes normal and abnormal images, month, day, peak hours (time1 & time2), year, number of COVID-19 affected patients, and COVID-19 patients cured. The collected data is for the period June 2020 and May 2021. In order to avoid the problem of overfitting in the later stage, we use data augmentation to expand the dataset. The pictures of the four categories are shown in Figure 6.

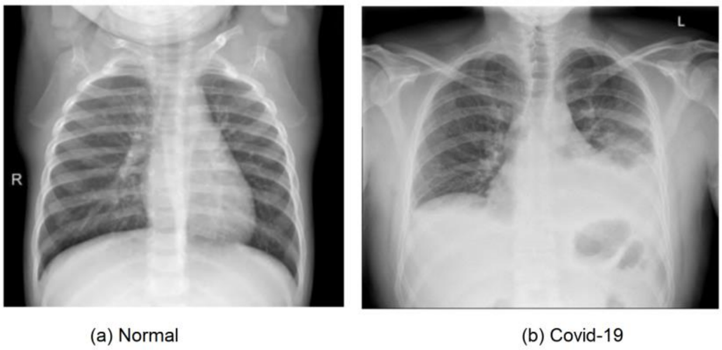

Figure 6. Types of lung X－ray images

After scaling transformation processing, we assume that the input image size is m*n and the output is h/w. The scaling ratio of the output to the length and width of the original image input is m/h and n/w respectively, so if you select (P_i) , j point as a certain point of the output image, you can know the corresponding point in the input image according to the ratio is (Q_i) m/h, jn/w. According to the interpolation algorithm and the four adjacent pixels found at point Q, ($P_{i,j}$) can be calculated. From Figure 7, it can be seen that after the original image is transformed by an interpolation algorithm, although the lung contour of the lung X-ray image is compressed, the main features contained in it are preserved. On the other hand, the contrast of the transformed lung X-ray image is slightly lower than that of the original image, so it is necessary to add image contrast enhancement processing in the follow-up.

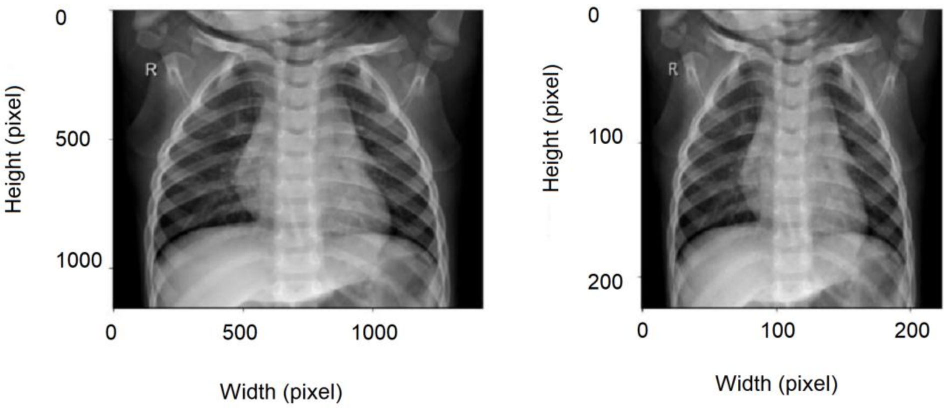

Figure 7. Comparison before and after image reduction

The Max pooling concept is used which depends on the maximum intensity present in the image matrix. The noise removal is also done here. The flattening layer normalizes the final output which is feed to a conventional ANN with a fully connected layer for classification of non-linear set of high level features. The output from the flattened layer is fed to a feed forward neural network trained with redial based neural network. The Softmax classification will facilitate the classification of low level feature set whose output is depicted in Figure 8. Figure 8, the convolved feature set is spatially reduced in the Pooling layer which enables dimension reduction thereby decreasing the computational complexity. The CNN model will extract the dominant feature set which are rotationally and position based non-variance, thus having an effective model for training. The results show that using the improved model to identify the positive and negative X-ray images of pneumonia, the accuracy rate reaches 95.11%, the precision reaches 90.75%, and the recall rate reaches 90.28%. Therefore, the improved model can greatly improve the diagnostic performance of pneumonia images.

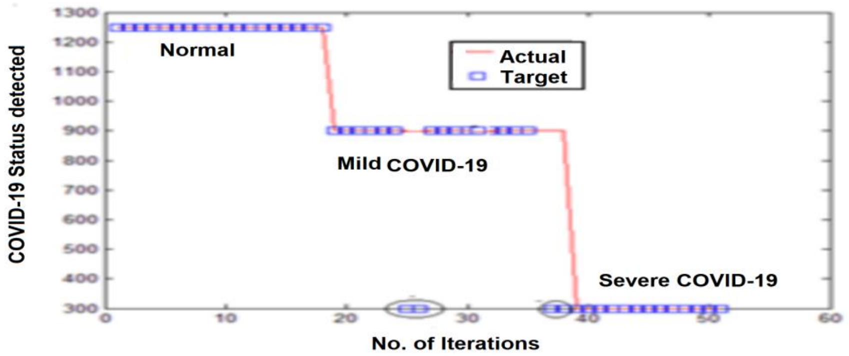

Figure 8. COVID-19 Detection using X-ray images of lungs

A comparative analysis is also done to substantiate that CNN is capable of forecasting the COVID-19 from the lung X-ray images as compared to the conventional Back Propagation Algorithm (BPA) in Table 1.

Table1. Comparison of Forecasting Efficiency

S. No	**Algorithm**	**Forecasting efficiency (%)**
1.	CNN	98.99
2.	BPA	96.7

5. Conclusion

With the development of modern medicine, the diagnosis of lung diseases relies more on the determination of medical images. With increasingly huge data, a powerful data processing model is urgently needed to provide favorable support for this field. As a new branch in the field of machine learning, deep learning has been widely used in many medical image processing fields, providing new methods and approaches for the intelligent detection of medical images. The influence of COVID-19 is predicted over the medium term one to several months in advance. Last but not least, the long-term projection predicts COVID-19 for a time frame of one to several years in the near future. Because of its resilience, CNN's COVID-19 forecasting has been given consideration despite the employment of current conventional approaches. Additionally, the ability to generalize is the primary justification for using CNN to forecast the COVID-19. The COVID-19 data was gathered over the course of 24 hours. Nearly 168 patterns are included, including details like the day, hour, and COVID-19 patients. The forecasting efficiency is 98.99% by CNN during the testing of this method.

References

[1] Eastin, C., Eastin T., Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *Journal of Emergency Medicine*, 2020.**58**(4):713-714.

[2] Wang, Y., Wang, Y., Chen, Y. & Qin, Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. J. *Med. Virol.*2020. **92**, 568–576 .

[3] Guang Chen, et al., Clinical and immunological features of severe and moderate coronavirus disease 2019. *J Clin Invest.* 2020.130(5):2620-2629.

[4] Khan A.I., et al., Coronet: a deep neural network for detection and diagnosis of COVID －19 from chest x －ray images. Computer Methods and Programs in Biomedicine, 2020. 196: 105581.

Journal of Physics: Conference Series **2467** (2023) 012001 doi:10.1088/1742-6596/2467/1/012001

[5] Heimdal I, et al., Human Coronavirus in Hospitalized Children with Respiratory Tract Infections: A 9-Year Population-Based Study from Norway. *J Infect Dis.* 2019. **219**(8):1198.
[6] Alimadadi A., et al., Artificial Intelligence and Machine Learning to Fight COVID－19. Physiological Genomics, 2020. 52(4):200－202.
[7]. Monto A.S., et al., Coronavirus Occurrence and Transmission Over 8 Years in the HIVE Cohort of Households in Michigan. J Infect Dis. 2020. 222(1):9
[8] McIntosh K., et al., Seroepidemiologic studies of coronavirus infection in adults and children. *Am J Epidemiol.* 1970. **91**(6):585
[9] Nickbakhsh S., et al., Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019. J. Infect Dis. 2020. **222**(1):17.
[10] Clark A., et al., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 working group. Global, regional, and national estimates of the population at increased risk of severe COVID-19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob Health*. 2020.(8):1003-1017.
[11] Song Y., et al., Deep learning enables accurate diagnosis of novel coronavirus(covid-19) with CT images. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021. **18**(4):2775-2780.
[12] Wang, S., et al., A deep learning algorithm using CT images to screen for corona virus disease. European Radiology, 2021. **31**(1):6096–6104.
[13] Hu J., Shen L.,Sun G., Squeeze-and- excitation networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020. **42**(8):2011-2023.