ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 8.4
IJAR 2023; 9(6): 189-192
www.allresearchjournal.com
Received: 19-02-2023
Accepted: 23-03-2023

Meragana Venu Madhavi
Department of Computer Science and Business System, R.V.R & J.C. College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Devineni Vignatha
Department of Computer Science and Business System, R.V.R & J.C. College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Pendem Eakshith Roop
Department of Computer Science and Business System R.V.R & J.C. College of Engineering, Chowdavaram, Guntur, AP

Kasukurthi Aravinda
Department of Computer Science and Business System, R.V.R & J.C. College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Peddi Anudeep
Department of Computer Science and Business System, R.V.R & J.C. College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Corresponding Author:
Meragana Venu Madhavi
Department of Computer Science and Business System, R.V.R & J.C. College of Engineering, Chowdavaram, Guntur, Andhra Pradesh, India

Novel classification and prediction of CNN based lung disease using deep learning

Meragana Venu Madhavi, Devineni Vignatha, Pendem Eakshith Roop, Kasukurthi Aravinda and Peddi Anudeep

DOI: https://doi.org/10.22271/allresearch.2023.v9.i6c.10940

Abstract
Corona virus attack, also known as COVID-19, is one of the most fatal and devastating diseases that affects people today. This Corona virus infection has spread over the entire planet due to community transmission. By early illness discovery, even in asymptomatic settings, via proper diagnosis, the patient's mortality rate may be reduced. Thus, it is necessary to build an autonomous detection system that, with its quick and precise findings, stops the corona virus from spreading. COVID-19 recipients are usually detected and given an initial prediction using scans, like chest X-rays and Computed Tomography (CT). Deep learning methods from the medical realm are used to find hidden patterns. In order to make predictions, chest x-ray picture features are extracted with the use of a convolutional neural network (CNN). In order to enhance the health plan, predictions are produced in the patient data using pattern creation. The chest X-ray image characteristics are fused to the CNN model training to provide progress in classification. The testing stage of the model performance evaluation takes into account generalized data. As compared to existing classification state-of-the-art approaches, the suggested CNN-based techniques perform better in terms of classification and illness prediction.

Keywords: Convolutional neural network (CNN), COVID-19, feature extraction, prediction

Introduction
Since December 2019, the outbreak known as COVID-19 has killed more than 800,000 people globally and infected more than 20 million others [1]. The World Health Organization pronounces COVID-19 as a global health emergency on January 30, 2020 (WHO). Chest computed tomography can aid in the clinical diagnosis of COVID-19 (CT). Yet, due to a lack of radiologists and doctors, the number of people being affected by the corona virus is fast rising. The development of automated algorithms is necessary for the categorization of COVID-19-assisted CT images of the chest [2]. The COVID-19 categorization of the chest makes use of various machine learning techniques [3]. In some circumstances, an efficient risk prediction model provides precision medicine recommendations for adjusting clinical therapy to each patient's needs, therefore extending the time to full recovery. Allowing the emergency department may improve patient flow, which lowers waiting times [4].

The purpose of patient outcome prediction involves completing a significant amount of research using several data types, including radiological, laboratory, and clinical aspects. Although a number of writers have reported some encouraging findings, the majority of them are biased for one of two reasons, according to a recent COVID-19 rate of survival study that provides detailed information [5]. Because of the frequent development of serious problems in patients and the first clinical encounter utilized for ML, Many published studies that emphasize the faculty criteria used by the technology lack medical obeying data [6].

The second issue is that many studies are unable to use current epidemiological records throughout people’s study; instead, those who rely on measures of the most recent predictor, which is now available from electronic medical records. The many branches of artificial intelligence (AI) are used to resolve a variety of challenging issues [7]. Thinking, learning, planning, information representation, and searching are all AI subfields. Deep Learning (DL) and Machine Learning (ML) algorithms, which are subsets of AI, produce intelligent models for task recognition.

ML is the primary subset of AI and implies minimal grasp to tackle issues due to its empirical models and fresh computational simulation lifestyle. The next significant model after this ML is the DL subset of ML. In previous years, several individuals have used DL algorithms. Yet in order to solve it, a lot of data is needed. It's encouraging to see ML and DL techniques being used in a variety of experimental projects in the technological, pharmaceuticals, and armed services [8]. These fields in the conflict also presented ML and DL techniques based on cutting- edge AI not long after the COVID-19 crashed. In this study, a quantitative framework that utilizes machine learning has been built to make it easier to arrange.

Materials and Methods

The machine learning discipline has made amazing development during the last ten years. The combination of ML and deep learning techniques greatly increases computational capability [9]. As a result, many applications are embracing machine learning. The COVID-19 pandemic could represent the most prominent use of machine learning in the fight over it. Several researchers are using Machine learning to develop various strategies to combat COVID-19 [10].

The many machine learning applications cover a wide spectrum of challenges related to the virus. For instance, COVID-19 diagnostic employs ML and DL to analyze medical pictures, protects medical staff members from patients who are impacted, and determines the severity of the patient's condition so that future treatment may be determined. In order to accurately anticipate disease outbreaks, disease transmission models are built. These models learn about the transmission list and the effects of the transmission channel by using data from social networking sites and COVID-19 scenario info [11].

The ML and DL are heavily utilized in public surveillance and epidemic protection, as well as in airport security inspections where patients are recognized and epidemics are detected [12]. Predicting the future, clinical applications, personalized medicine, and communication locating are among the four types of machine learning algorithms used by the COVID- 19.

The Deep learning systems correctly forecast the number of new infections. For forecasting time series, a conventional technique like ARIMA models performs slowly compared to recurrent neural networks [13]. Researchers employed recurrent neutral networks to forecast the number of new infections and the spread of diseases and to account for shifting long-short-term memory networks [14]. Thus, it is the most crucial application of the machine learning model.

It is possible to improve medical diagnosis by using computer vision. In many image recognition tasks, living thing precision are roughly comparable to model accuracy. As a result, the machine vision program can unquestionably identify virus indicators using COVID-19 diagnosis chest X- ray images. Nevertheless, many nations do not currently have enough medical facilities for testing, and the only means to identify or diagnose the virus is through a costly process for chest X-ray scans.

According to earlier research, utilizing the deep learning techniques, COVID-19 may be recognized with a 99% using various artificial intelligence-based software programs. In numerous domains, researchers have used machine learning algorithms as a guide for future findings [15]. Variation auto encoders are used to assess chemical compositions, developing novel medications as a consequence. The development of the COVID-19 vaccine may be achieved by combining the existing flu vaccinations with auto encoders.

Novel classification and prediction of CNN based lung disease using deep learning

The proposed CNN uses X-ray images and is derived from the Lung dataset. The technique for lung disease classification and prediction based on CNN is demonstrated in the following Fig. 1. The suggested framework has two modules: CNN and preprocessing.

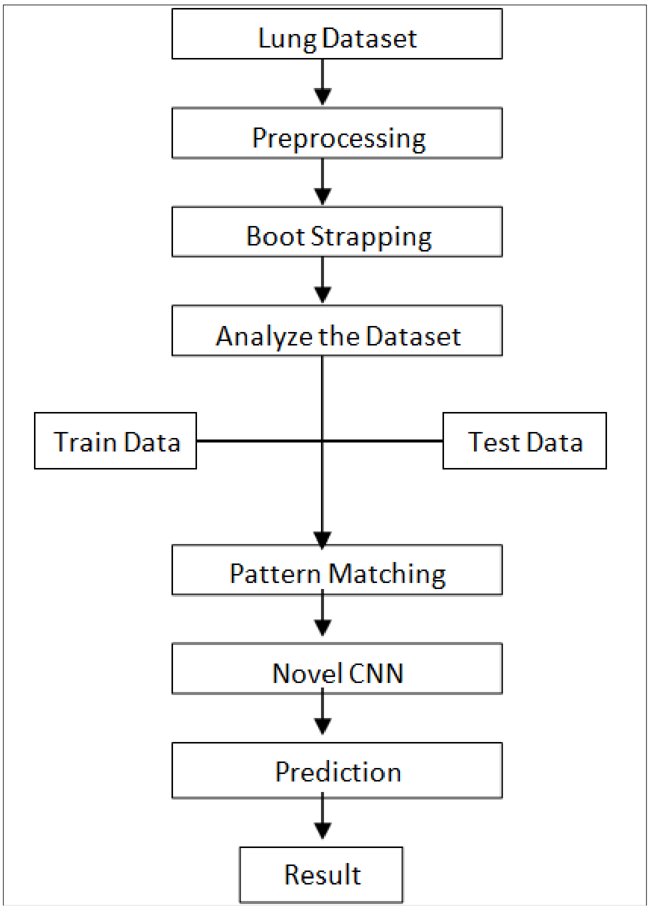

Fig 1: Framework of CNN's Novel Classification and Prediction

The X-ray datasets are extracted using two datasets in this approach. The 79 photos in dataset 1 come from the viral and bacterial pneumonia separately (2020). Dataset 2 at Kaggle contains 28 photos of regular persons and 78 photographs of COVID-19 patients (2020).The size of the training dataset has an impact on how well deep learning models perform. The COVID- 19 data set is quite tiny compared to the enormous dataset used by deep learning modules, making it impossible to calculate the generality and toughness of these models. The Keras Image Data Generator class is used to resolve this issue by fusing the CNN network with massive X-ray images. The image generator class offered by Keras defines the image augmentation settings.

The real-life information contains a significant quantity of noisy data as well as missing data. So, it is necessary to remove this noisy data through the preprocessing stage in order to make accurate predictions. The suggested model's framework is depicted in Fig. 1. The data that was gathered also includes missing information and noise. Filling in the missing numbers and eliminating the noise yields efficient and precise outcomes. The activities involved in transformation include aggregation, normalization, and smoothing, which alter one type of data into another type of data. Before preprocessing the data, the data must be merged from several sources.

The suggested method is simulated by utilizing the intervention effectiveness ratings to ascertain the intervention's potential relative impact. After becoming familiar with the simulation settings, this may be done for a particular nation or people. Effectiveness ratings are given access to disease what-if scenario temporal progression patterns, such as the second peak in June.

The block exposes the configurability parameters, which are adjustable between 0.1 and 1.0.For proven instances, daily sliding window analysis for the first and second peaks and beginning point detection are necessary. The following information, which was developed at the simulator's core, is used in the conclusion of the probability distributions parameters. The supervised learning method is used to estimate the intervention effectiveness score, which is also utilized to determine the population of study data. In order to improve the model, the parameters are changed iteratively. This significantly minimizes the mistakes in the forecast data when using historical data.

Results and Discussion

Precision, accuracy, F-score, recall, specificity, sensitivity, and AUC are the performance indicators used to assess this study. The aforementioned performance measurements are defined by four terms: False Positive (FP), True Negative (TN), True Positive (TP), and False Negative (FN).In this case, positive denotes that the patient's test result was positive, whereas negative denotes that the patient's test result was negative. Consequently, it is evident that the word FP designates a patient who is not ill but whose test results are positive, whereas TP designates a patient who is ill and whose results are negative. Similar to how FP designates patients with negative reports who do not have any diseases whereas FN designates patients with negative reports who have:

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

In order to identify the dangerous disorders, high sensitivity is necessary. If the sensitivity is 100%, the patients with the condition are appropriately identified. Here is how the specificity is defined:

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

Procedures for specificity and sensitivity testing are not viewed as the cutoff point. Statistics on false positives and false negatives are affected. If the cut-off values are high, the test is less sensitive but very specific, and if they are low, it is more sensitive but less specific. High cut-off values also boost false negatives. Accuracy is the most often used performance statistic for classifiers, and it is defined as follows:

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

Recall and accuracy are also often employed in classifier performance assessments. The precision only computes positive situations, and its formula is:

$$Precision = \frac{TP}{TP + TF}$$

Statistics are utilized to assess the classifier's performance using the F-score. It calculates the value ranging from 0 to 1 for classifier indicatives and needs to know the classifiers' accuracy and recall values. In accordance with this, classifier performance is ranked from lowest to highest. F1-score may be calculated using:

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

In order to evaluate the proposed model, three scenarios that include several classes are employed.

In scenario 1, the two classes employed for the proposed model's training and testing are Normal and COVID-19.

In scenario 2, the training and testing process uses three classes: Normal, COVID-19, and Viral Pneumonia.

In scenario 3, a total of four classes—Normal, COVID-19, Viral Pneumonia, and Bacterial Pneumonia—are used to train and test the proposed model.

Fig 2: Performance Analysis of CNN

Fig 2: Illustrates the proposed method of CNN performance analysis for categorizing and forecasting COVID-19 infections. It can be shown that greater than 90% accuracy is reached for the three scenarios—that is, for the 2 classes, 3 classes, and 4 classes as well. The increased Precision and F- score values were also reached, as seen in the figure.

Table 1: Classifier Performance Metrics analysis

Model	Classes	Accuracy	Precision	F-Score
Proposed CNN	2	0.9802	0.9851	0.9812
	3	0.9123	0.9306	0.9735
	4	0.8975	0.9123	0.9328
VGG16	2	0.9768	0.9806	0.9793
	3	0.8895	0.9216	0.9523
	4	0.8751	0.8924	0.9131
AlexNet	2	0.6892	0.7126	0.7654
	3	0.8726	0.8968	0.9234
	4	0.8274	0.8862	0.9156

The comparative results are shown in Table 1 above for the three distinct models, VGG16, AlexNet, and CNN, in three independent classes, two, three, and four. The results show that the proposed model performs better than the two other classes in terms of accuracy, precision, and F1-score metrics.

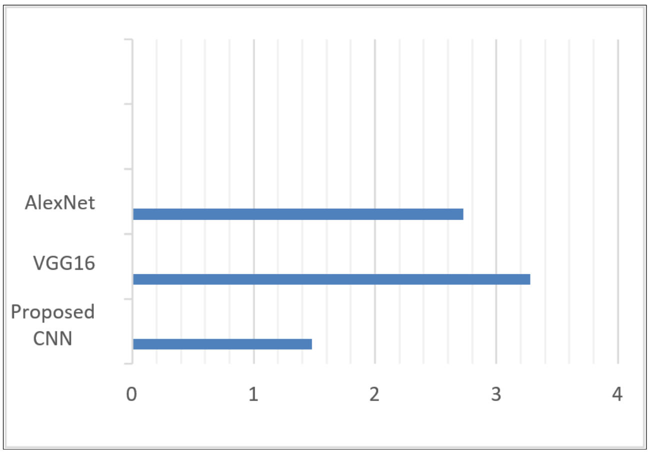

Fig 3: Different Approaches of Training Time

Figure 3 compares the training times for three distinct models, including AlexNet, VGG16, and the proposed CNN. Both the AlexNet and VGG16 models require longer training time than the suggested CNN technique because of their complex architectural designs. As a result, the suggested method performs best at forecasting a variety of illnesses like the corona virus and several bacterial infections.

Conclusion

Using the X-ray pictures of the chest In this article, deep learning and convolutional neural network-based classification are used to distinguish COVID-19 patients from healthy individuals. The COVID-19 data set is quite tiny compared to the enormous dataset used by deep learning models, making it impossible to calculate the generality and toughness of these models. This issue might be cleared up by concatenating the CNN network with enormous X- ray images using the Keras Image Data Generator class. CNN's effectiveness is measured using a variety of classes (as 2, 3, & 4), including healthy humans, COVID-19 sufferers, people with bacteremia, and people with bacterial meningitis. With modeling techniques like AlexNet, VGG16, and the suggested CNN model, many metrics, such as precision, accuracy, and F-score values, are utilized to examine performance.

References

1. Gao S, Wu Y, Fan D, Mei J, Xu J, Cheng M, *et al*. JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation, IEEE Trans. on Image Processing; c2021, 30.
2. Gomez-Rios A, Tabik S, Sevillano-García I, Martín-Rodriguez JL, Charte Guirado E, Rey-Area M, *et al*. COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images, IEEE Journal of Biomedical and Health Informatics; c2020, 24(12).
3. Xiao Y, Liu Y. Analysis and Prediction of COVID- 19 in Xinjiang Based on Machine Learning, 2020 5th International Conf. on Information Sci., Computer Technology and Transportation (ISCT); c2020.
4. Hossain Mondal MR, Prajoy P. Machine Learning to Predict COVID-19 and ICU Requirement, 2020 11th International Conf. on Electrical and Computer Engg. (ICECE); c2020.
5. Choi GS, Reshi A, Rustam F, Saleem U, Arif M, Aslam On W. COVID-19 Future Forecasting Using Supervised ML Models, IEEE Access; c2020, 8.
6. Durga T, Kumari P. Real-time estimation of COVID-19 cases using machine learning and mathematical models - The case of India, 2020 IEEE 15th International Conf. on Industrial and Information Sys. (ICIIS); c2020.
7. Shimpi R, Varun P, Mohit M. Development and evaluation of an AI System for early detection of Covid-19 pneumonia using X-ray (Student Consortium), 2020 IEEE Sixth International Conf. on Multimedia Big Data (BigMM); c2020.
8. AYDIN I, Sevi M. COVID-19 Detection Using Deep Learning Methods, 2020 International Conf. on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI); c2020.
9. Ali Nasseri M, Abouzar E, Hasan Sarhan M, Maier M, Daniel Z, Nassir N, *et al*. Machine Learning Techniques for Ophthalmic Data Processing: A Review, IEEE Journal of Biomedical and Health Infor; c2020, 24(12).
10. Alaei M, Ghorvei M, Rezaeijo SM. A machine learning method based on lesion segmentation for quantitative analysis of CT radiomics to detect COVID-19, 2020 6th Iranian Conf. on Signal Proce. and Inte. Sys. (ICSPIS); c2020.
11. Anandan C, Ashok Kumar D. Sentiment and Emotion in Social Media COVID-19 Conversations: SAB-LSTM Approach, 020 9th International Conf. Sys. Modeling and Advancement in Research Trends (SMART); c2020.
12. Camelia D, John Milne R, Corina I, Cotfas L. Evaluation of Boarding Methods Adapted for Social Distancing When Using Apron Buses, IEEE Access; c2020, 8.
13. Fareed NY, Mustafa HI. COVID-19 Cases in Iraq; Forecasting Incidents Using Box - Jenkins ARIMA Model", 2020 2nd Al-Noor International Conf. for Sci. and Tech. (NICST); c2020.
14. Bader Alsabek M, Ismail S, Abdelfattah H. COVID-19 Detection System using Recurrent Neural Networks, 2020 International Conf. on Comm., Computing, Cyber security, and Infor. (CCCI); c2020.
15. Usman M, Siddique L, Iqbal W, Manzoor S, Tyson G, Jon C, *et al*. Leveraging Data Science to Combat COVID-19: A Comprehensive Review, IEEE Trans. on Artificial Intelligence; c2020, 1(1).