SN Computer Science (2022) 3:79
https://doi.org/10.1007/s42979-021-00965-2

REVIEW ARTICLE

Comparative Study Based on Analysis of Coronavirus Disease (COVID-19) Detection and Prediction Using Machine Learning Models

R. Sudha Abirami[1] · **G. Suresh Kumar**[1]

Received: 25 August 2021 / Accepted: 7 November 2021

Abstract
As the number of COVID-19 cases increases day by day, the situation and livelihood of people throughout the world deteriorates. The goal of this study is to use machine learning models to identify disease and forecast whether or not a person is infected with the virus or another common illness. More articles about COVID-19 will be released starting in 2020, but we still do not have a reliable prediction mechanism to diagnose the disease with 100% accuracy. This comparison is done to see which model is the most effective in detecting and predicting disease. Despite the fact that we have immunizations, we require a best-prediction strategy to assist all humans in surviving. Researchers claimed that the supervised learning method predicts more accurately than the unsupervised learning method in the majority of studies. Supervised learning is the process of mapping inputs to derived outputs using a set of variables and created functions. This will also help us to optimize performance criteria using experience. It is further divided into two categories: classification and regression. According to recent studies, classification models are more accurate than other models.

Keywords Coronavirus disease (COVID-19) · Machine learning · Supervised · Unsupervised · Classification · Regression

Introduction

Coronaviruses (CoV) are a large family of viruses that cause illness ranging from common cold to more severe diseases. These are positive-stranded RNA (+ssRNA) viruses with a crown-like appearance. They are mainly susceptible to mutation and recombination and are, therefore, highly diverse. They reside in bats and wild birds, and can spread to other animals and hence to humans. The virus which causes COVID-19 is thought to have originated in bats [1]. Outbreaks of COVID-19 have been reported in various types of workplaces and job categories. All workers should be protected from acquiring COVID-19 because of their work. The prevention of COVID-19 in work settings should be combined with measures for protecting physical and mental health, safety and wellbeing of workers from other occupational hazards in the operation, closures and reopening of workplaces [2]. COVID-19 is likely spread in the following ways; When the virus travels in respiratory droplets when an infected person coughs, sneezes, talks, sings or breathes near you (within six feet). This is thought to be the main way it is spread. Another way is through close contact (touching, shaking hands) with an infected person. And moreover, by touching surfaces that the virus has landed on, then touching our eyes, mouth, or nose before washing the hands [3]. Due to the huge growth of machine learning techniques, predicting and diagnosing diseases made it easy in current days. Machine learning can learn through machines and makes a proper prediction by self-learning. Several techniques of machine learning were utilized to identify the affected individuals with personalized characteristics. Machine learning is an exclusive technology that is used to allow a computer to train itself and make correct decisions and can be employed for predicting coupled information from previous records, geography, and demographics, climate as well as outbreak severity [4]. Machine learning (ML) is one of the most advanced concepts of artificial intelligence

This article is part of the topical collection "Machine Learning for Pandemic Prediction and Control" guest edited by Anand J Kulkarni, Akash Tayal, Patrick Siarry, Arun Solanki and Ali Husseinzadeh Kashan.

✉ R. Sudha Abirami
rsudhaabirami27@gmail.com

G. Suresh Kumar
mgsureshkumar@gmail.com

[1] Department of Computer Science, School of Engineering and Technology, Pondicherry University, Puducherry, India

Published online: 20 November 2021

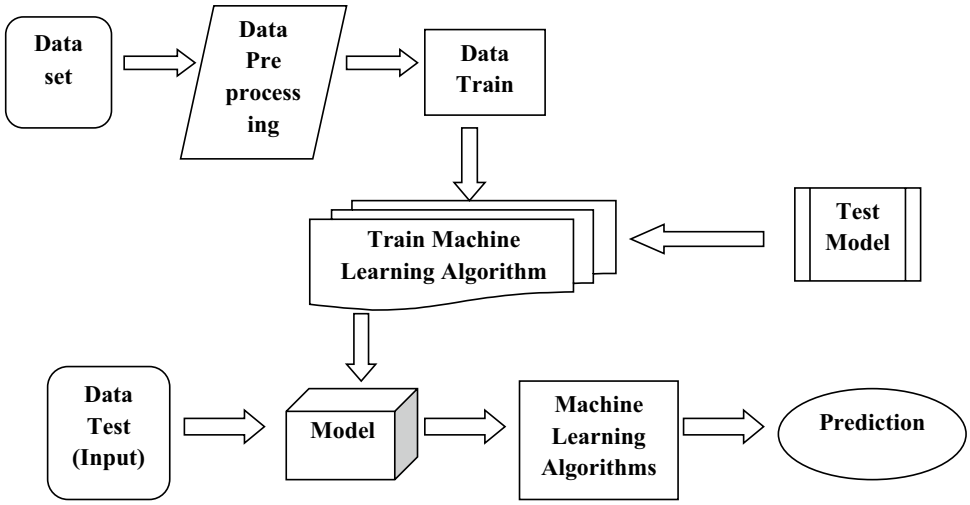

Fig. 1 Prediction using machine learning models

(AI), and provides a strategic approach to developing automated, complex and objective algorithmic techniques for multimodal and dimensional biomedical or mathematical data analysis [5]. This study presents the machine learning model as a method for predicting the transmission of COVID-19 pandemic in an easily understandable way using statistical visualization graphs e.g., normal distribution. It determines the predictive value of the technique with quality and density of collected data of WHO. The resulting predictions will reduce the population's anxiety and prepares their mentality for accepting and dealing with the next phases of the pandemic. This prediction study proposed a model for COVID-19 prediction and it has achieved a classification accuracy of 91.62%. Even more important fact is it yields a sensitivity of around 95% for COVID-positive cases [6].

From all the researches made so far stated that, the machine learning models can be used to predict and detect this disease (Fig. 1). It will help the people to reduce their panic about the symptoms caused by normal cold or fever. This pandemic will end definitely once all of us must be vaccinated and as well as ensuring the guidelines stated by WHO while in workplace, traveling, in social gathering etc., Even though we used to research on this virus alleviation and mutiny, we need the prediction about causes of spread, symptoms, death rate, recovery rate, and remedies.

The remaining paper contains the following subsections; "Related Work" for work done by other researchers after reception of this disease, "Machine Learning Methods" to describe the various machine learning models, "Case Study Table" (Table 1) to compare the methodologies used, and Conclusion.

Related Work

Due to COVID-19 pandemic, the need of disease analysis has been raised to detect whether a person is infected with virus or not. In the year 2021, Corona 2.0 obscured many people lives. To avoid this situation in future, health care providers must have some prediction techniques to save human life [1, 2]. Supervised machine learning is used to get better prediction accuracy. These are some papers given the results regarding Corona disease prediction and detection.

In recent months, different techniques using image processing are done by various researchers. In this paper, a major review on image acquisition, segmentation, diagnosis, avoidance, and management are presented. An analytical comparison of the various proposed algorithm by researchers for corona virus has been carried out. ML methods train the input data and analyze the output data statistically. The application of ML includes the detection of infected persons and the temperature of the person [7]. The machine learning models was trained to predict drug structures that could potentially treat COVID-19. The model is employed with AI neural network with multiple hidden layers and neural systems such as repetitive neural networks (RNN), convolutional neural networks (CNN), deep belief networks (DBN), and completely associated feed-forward systems that supported to investigate drug molecule design and re-positioning of drug molecules [8, 9].

While the RT-PCR is the silver bullet test for confirming the COVID-19 infection, it is limited by the lack of reagents, time-consuming, and the need for specialized labs. This study develops six predictive models for COVID-19 diagnosis using 6 different classifiers (i.e., BayesNet, Logistic, IBk, CR, PART, and J48) based on 14 clinical features. This study retrospects 114 cases from the Taizhou hospital of Zhejiang Province in China. The results showed that the

Table 1 Comparison table about method, algorithmic models used to predict and detect COVID-19

Author	Year	Method	Model	Algorithm	Result
[15]	2020	Supervised learning	Classification	Shallow Single-Layer Perceptron Neural Network (SSLPNN) and Gaussian Process Regression (GPR) model	This research used, binary classification and regression for prediction. From that binary classification model was highly accurate, with a root mean square error (RMSE) of 0.91 than regression analysis
[11]	2021	Supervised learning	Regression model and decision tree	Decision tree and linear regression	In this research, the experimental results of the proposed model showed that the overall R2 is 0.99 from the perspective of confirmed cases
[25]	2020	Supervised learning	Regression model	Linear regression, multilayer perceptron and vector auto regression	This forecasting model stated that, predicted values and matching with cases from John Hopkins University11data we can conclude that the MLP method is giving good prediction results than that of the LR and VAR method using WEKA and Orange
[26]	2021	Artificial neural network	Recurrent neural networks	Adaptive neuro-fuzzy inference system (ANFIS) and long short term memory network(LSTM)	This finding has used the ANFIS and LSTM-based prediction model to forecast the COVID-19 pandemic growth in Bangladesh. This research can say that LSTM has provided with more satisfactory results compared to ANFIS for predicting the COVID-19 pandemic
[10]	2021	Supervised learning	Classification	BayesNet, logistic, IBk, CR, PART, and J48	This study retrospects 114 cases from the Taizhou hospital of Zhejiang Province in China. The results showed that the CR (classification via regression) meta-classifier is the most accurate classifier for predicting the positive and negative COVID-19 cases with an accuracy of 84.21%
[27]	2020	Supervised learning	Classification, regression	Logistic regression and multinomial naïve Bayes classifier	Machine learning algorithms are used for classifying clinical reports into four different classes. After performing classification, it was revealed that logistic regression and multinomial naive Bayesian classifier gives excellent results by having 94% precision, and accuracy 96.2%
[14]	2021	Supervised learning	Classification, regression	Support vector machine, stacking-ensemble learning, ARIMA, CUBIST, RIDGE and RF models	In this study, the Machine Learning approaches employed various models like RF, ARIMA, SVR, CUBIST, and Gradient Boosting to precisely make predictions. It was found that the SVR and SEL were the best in accuracy terms
[28]	2020	Supervised learning	Classification	Logistic regression, decision tree, support vector machine naive Bayes, and artificial neutral network	The findings were stated that, the correlation coefficient analysis between various dependent and independent features was carried out. The result of the performance evaluation of the models showed that decision tree model has the highest accuracy of 94.99%

Table 1 (continued)

Author	Year	Method	Model	Algorithm	Result
[29]	2021	Supervised learning	Classification	Hybrid social group optimization and support vector classifier	In this work, they propose a pipeline that uses CXR images to detect COVID-19 infection. The features from the CXR images were extracted and the relevant features were then selected using Hybrid Social Group Optimization algorithm. The proposed pipeline achieves a classification accuracy of 99.65% using support vector classifier
[30]	2020	Transfer learning	Convolutional neural networks	VGG16, ResNet50, Inception v3	In this study, the transfer learning technique has been applied to clinical images. Texture feature extraction is accomplished using Haralick features which focus only to detect COVID-19 using statistical analyses
[31]	2021	Supervised learning	Classification	Random Forest, Logistic Regression, Extreme Gradient Boosting	Out of all the three methods, Random Forest gave more accuracy of 0.952. But, due to insufficient dataset, data imbalance occurs in this proposed approach. SMOTE technique was used to rectify data imbalance.

CR meta-classifier is the most accurate classifier for predicting the positive and negative COVID-19 cases with an accuracy of 84.21% [10].

Generally, a huge amount of data is accumulating regarding the COVID-19 pandemic, which makes hot research topics for machine learning researchers. This study considered machine learning approaches to predict the spread of the COVID-19 in many countries. The experimental results of the proposed model showed that the overall R2 is 0.99 from the perspective of confirmed cases. A machine learning model has been developed to predict the estimation of the spread of the COVID-19 infection in many countries and the expected period after which the virus can be stopped [11, 12].

The increasing interest in developing artificial intelligence (AI) applications has addressed several medical problems. Some of the AI applications in healthcare are AI-enabled clinical support system, machine learning for drug discovery [13], diagnosing tissue samples, and AI-assisted image analysis for radiology etc. However, such applications remain insufficient given the high potential threat posed by this virus to global public health.

This systematic review addresses automated AI applications based on data mining and machine learning (ML) algorithms for detecting and diagnosing COVID-19. We aimed to obtain an overview of various deep learning techniques included 3D and 2D analysis of the chest CT images a Hybrid-Covid network, an IRRCNN model with NABLA-N network and Covid-Net and CheXNet network architecture-based detection. The accuracy of the CheXNet architecture-based detection had the highest accuracy. [14].

Sensing the threatening impacts of COVID-19, researchers of computer science have started using various techniques and approaches of machine learning and deep learning to detect the presence of the disease using X-rays and CT images. Artificial intelligence-centered tools can be designed and developed quickly for adapting the existing AI models and for leveraging the ability to modify and associating them with the preliminary clinical understanding to address the new group of COVID-19 and the novel challenges associated with it [15].

Many research done so far have emphasized that we need a more effective technique to detect and forecast infectious people based on all of these facts. The following section contains a diagram (Fig. 2) that depicts various machine learning models to enhance the research in prediction and detection of crucial COVID-19 in early days from the basic symptoms.

Machine Learning Models

Machine learning is typically a system of automated data processing algorithms that help to make decisions more natural and enhance performance based on the results. The "Learning" here means that the algorithm can garner new information and insights without being explicitly programmed. There are several models of machine learning present. From those, some primary paradigms are described as follows [16]:

Supervised machine learning: It is defined as, use of labeled datasets to train algorithms that can classify data or predict outcomes accurately. As input data are fed into the model, the weights are adjusted until the model is properly fitted, which happens during the cross-validation phase. Organizations can use supervised learning to tackle a range of real-world problems at scale, such as spam classification in a distinct folder from our email.

Unsupervised machine learning: These algorithms can be used to find patterns as the inputs are unlabeled. Unlabeled datasets are analyzed and clustered using machine learning methods. Without the need for human intervention, these algorithms uncover hidden patterns or data groupings. *It is the best solution for exploratory data analysis, cross-selling techniques, consumer segmentation, and image identification* because of its capacity to detect similarities and differences in information.

Reinforcement machine learning: Similar to supervised ML, but in this case, instead of a labeled output, there are rewards and the algorithm's goal is to maximize rewards. Reinforcement learning is a machine learning training strategy that rewards desirable behaviors while penalizing undesirable ones. A reinforcement learning agent can perceive and comprehend its surroundings, act, and learn through trial and error in general.

Semi-supervised machine learning: Input data are a mixture of labeled and unlabelled, so the model must learn structures to organize as well as make predictions. Semi-supervised learning is a type of machine learning that tries to solve issues involving both labeled and unlabeled data. Semi-supervised learning makes use of mathematical concepts such as clustering and classification method features.

From these types, supervised learning is intensively helpful in finding solution for real-world computational problems. The algorithms predict the outcome results from unlabeled data. Different types of supervised learning are regression, classification, naïve Bayesian Model, decision trees, random forest model, neural networks, support vector machines.

The diagram (Fig. 2) below showed the major types of machine learning models.

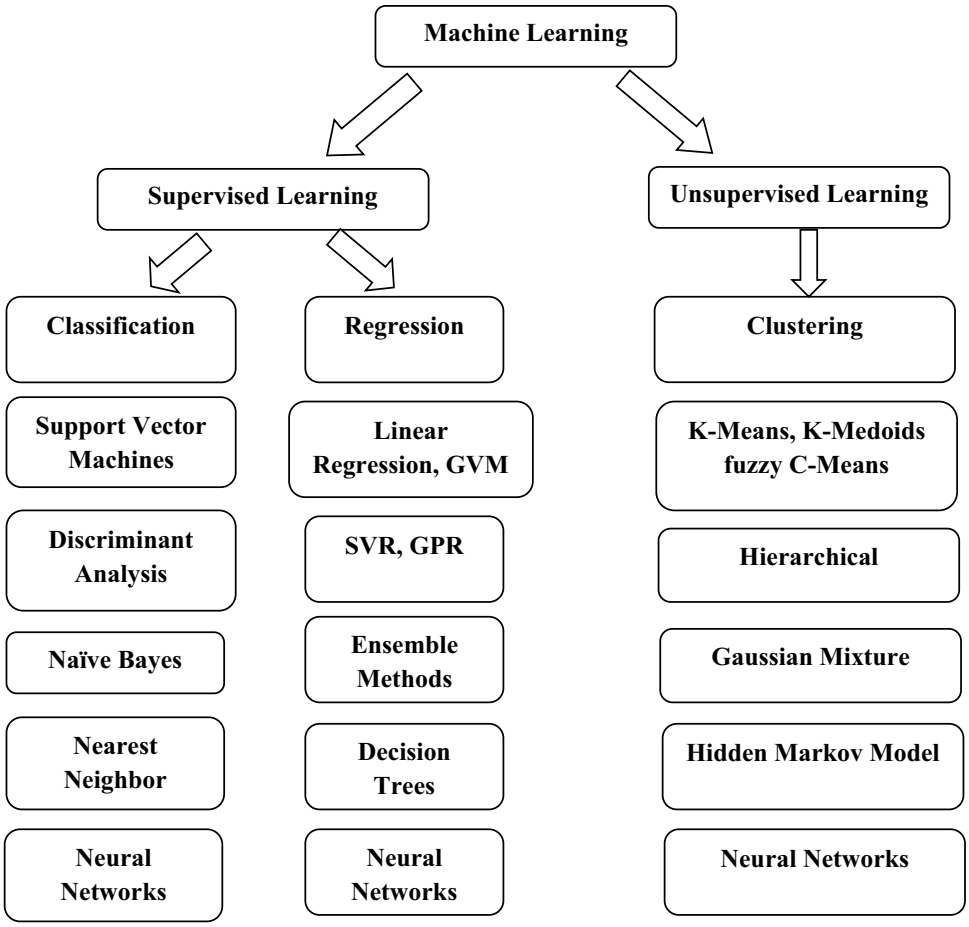

Fig. 2 Machine learning models overview

Discussion

The purpose of this study is to recognize which machine learning method works well in predicting and detecting diseases with highest accuracy. The entire world faces a big crisis due to COVID-19. During this pandemic, more researches have been made to identify the infectious persons in a proper way at right time. From those researches, this comparison table illustrates that the supervised learning plays a vital role in prediction using symptoms, clinical features and chest X-rays. Some researchers used classification model with different algorithms for example, shallow single-layer perceptron neural network, Gaussian process regression, Bayes Net, logistic, J48, multinomial Naive Bayes classifier, support vector machine etc., [17–24]

Some researchers used regression, recurrent neural network and convolutional neural network. From all the experiments and studies, it is noted that supervised learning techniques are alone used to obtain more accurate results. To develop an effective algorithm for detection of corona virus presence in human beings, the above table (Table 1) would be helpful. The results were compared and it is observed that classification model is best among all the others. It yields higher accuracy, more specificity when compared with regression and convolutional neural networks. Furthermore, this study would be useful to develop a new algorithm for detecting this deadly disease in efficient manner. The research may be fruitful by concentrating on new classification model.

Conclusion

Thus, this comparative case study ensures that the machine learning models are efficient in a way to predict and detect the infectious persons. In that, supervised learning algorithms are more powerful when compared to other algorithms. This study focused to show that this prediction will be helpful to end this epidemic in future. Our conclusion is machine learning applications in the field of disease diagnostics showed promising results with higher accuracy, and specificity. In particular, supervised machine learning takes a leading role in disease diagnosis. This machine learning technique yield accurate results when compared to other techniques. The researchers have used so many models such as, linear regression, naïve Bayes classifier, support vector

machine, decision tree and ResNet50 etc., to train and test the data set for predicting and detecting Coronavirus disease in an effective manner. This study will be expanded into developing a new algorithm based on the models available in terms of supervised machine learning for earlier diagnosis of diseases especially COVID-19.

Funding Not applicable.

Data availability Introduction about COVID-19 used in this study was collected from official website of WHO (World Health Organization).

Code availability Not applicable.

Declarations

Conflict of interest There are no conflicts of interest.

References

1. https://www.cebm.net/covid-19/coronaviruses-a-general-introduction. Accessed 22 June 2021.
2. https://www.who.int/publications/i/item/WHO-2019-nCoV-workplace-actions-policy-brief-2021-1. Accessed 22 June 2021.
3. https://my.clevelandclinic.org/health/diseases/21214-coronavirus-covid-19. Accessed 22 June 2021.
4. Nayak J, Naik B, Dinesh P, Vakula K, Rao BK, Ding W, Pelusi D. Intelligent system for COVID-19 prognosis: a state-of-the-art survey. Appl Intell. 2021. https://doi.org/10.1007/s10489-020-02102-7.
5. Sajda P, Columbia University. Machine learning for detection and diagnosis of disease. Annu Rev Biomed Eng. 2006;8(1):537–65. https://doi.org/10.1146/annurev.bioeng.8.061505.095802.
6. Malki Z, Atlam E-S, Ewis A, Dagnew G, Ghoneim OA, Mohamed AA, Abdel-Daim MM, Gad I. The COVID-19 pandemic: prediction study based on machine learning models. Environ Sci Pollut Res. 2021. https://doi.org/10.1007/s11356-021-13824-7.
7. Bhargava A, Bansal A. Novel corona virus (COVID-19) diagnosis using computer vision and artificial intelligence techniques: a review. Multimed Tools Appl. 2021. https://doi.org/10.1007/s11042-021-10714-5.
8. Das AK, Ghosh S, Thunder S, Dutta R, Agarwal S, Chakrabarti A. Automatic COVID-19 detection from X-ray images using ensemble learning with convolutional neural network. Pattern Anal Appl. 2021. https://doi.org/10.1007/s10044-021-00970-4.
9. Keerthika K, Suganya A (2020) A survey on machine learning applications to tackle covid-19 viral pandemic. Int J Adv Res Eng Technol (IJARET) 11(11): 9–14. Available online at http://iaeme.com/Home/issue/IJARET?Volume=11&Issue=11. Doi: https://doi.org/10.34218/IJARET.11.11.2020.002. Accessed 22 June 2021.
10. Arpaci I, Huang S, Al-Emran M, Al-Kabi MN, Peng M. Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. Multimed Tools Appl. 2021;7:1–15. https://doi.org/10.1007/s11042-020-10340-7.
11. Albahri S, AlAmoodi AH, Albahri AS, Hamid RA, Alwan JK, Al-qays ZT, Zaidan A, Zaidan B, Jamal Mawlood Khlaf AO, Almahdi EM, Thabet E, Hadi SM, Mohammed KI, Alsalem MA, Al-Obaidi JR, Madhloom HT. Role of biological data mining and machine learning techniques in detecting and diagnosing the novel coronavirus (COVID-19): a systematic review. J Med Syst. 2020;44(7):122. https://doi.org/10.1007/s10916-020-01582-x.
12. Rastogi V, Jain AK (2020) COVID-19: nature, outbreak and role of machine learning. Int J Adv Res Eng Technol (IJARET). 11(12): 2387. http://iaeme.com/Home/issue/IJARET?
13. Gupta JP. A computer-based disease prediction and medicine recommendation system using machine learning approach. Int J Adv Res Eng Technol (IJARET). 2021;12(3):673–83. https://doi.org/10.34218/IJARET.12.3.2021.062.
14. Aishwarya T, Ravi Kumar V. Machine learning and deep learning approaches to analyze and detect COVID-19: a review. SN Comput Sci. 2021;2(3):226. https://doi.org/10.1007/s42979-021-00605-9.
15. Almuayqil SN, Humayun M, University A-J, Naseem S, Khan WA. Prediction of COVID-19 cases using machine learning for effective public health management. Comput Mater Contin. 2020. https://doi.org/10.32604/cmc.2021.013067.
16. https://www.ibm.com/cloud/learn/supervised-learning. Accessed 26 Oct 2021.
17. Pourhomayoun M, Shakibi M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. Smart Health. 2021;20:100178. https://doi.org/10.1016/j.smhl.2020.100178.
18. Dey A. Machine learning algorithms: a review. Int J Comput Sci Inf Technol. 2016;7(3):1174–9.
19. Solanki A, Singh T. COVID-19 epidemic analysis and prediction using machine learning algorithms. Emerg Technol Battling Covid-19. 2021. https://doi.org/10.1007/978-3-030-60039-6_3.
20. Assaf D, Gutman Y, Neuman Y, Segal G, Amit S, Gefen-Halevi S, Shilo N, Epstein A, Mor-Cohen R, Biber A, Rahav G, Levy I, Tirosh A. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. Intern Emerg Med. 2020;15:1435–43.
21. Brunesea L, Martinellib F, Mercaldoab F, Santonec A. Machine learning for coronavirus covid-19 detection from chest x-rays. Elsevier Proc Comput Sci. 2020;176(2020):2212–21. https://doi.org/10.1016/j.procs.2020.09.258.
22. Kassania SH, Kassanib PH, Wesolowskic MJ, Schneidera KA, Detersa R. Automatic detection of coronavirus disease (COVID-19) in X-ray and CT images: a machine learning-based approach. Biocybern Biomed Eng. 2020;41(3):867–79.
23. Kwekha-Rashid AS, Abduljabbar HN, Alhayani B. Coronavirus disease (COVID-19) cases analysis using machine-learning applications. Appl Nanosci. 2021. https://doi.org/10.1007/s13204-021-01868-7.
24. Nassef M, Shorrab AA, Buheji M, Buheji AR, Abosamak MF. Alleviation of antimicrobial therapy in ICU during COVID-19 second wave—a review paper. Int J Manag. 2020. https://doi.org/10.34218/IJM.11.12.2020.204.
25. Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. Stoch Environ Res Risk Assess. 2020;34:959–72.
26. Chowdhury AA, Hasan KT, Hoque KKS. Analysis and prediction of COVID-19 pandemic in Bangladesh by using ANFIS and LSTM network. Cognit Comput. 2021;12:1–10. https://doi.org/10.1007/s12559-021-09859-0.
27. Khanday AM, Rabani ST, Khan QR, Rouf N, Din MM. Machine learning based approaches for detecting COVID-19 using clinical text data. Int J Inf Technol. 2020;30:1–9. https://doi.org/10.1007/s41870-020-00495-9.
28. Muhammad LJ, Algehyne EA, Usman SS, Ahmad A, Chakraborty C, Mohammed IA. Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset. SN Comput Sci. 2020;2:1–3.

29. Singh AK, Kumar A, Mahmud M, Kaiser MS, Kishore A. COVID-19 infection detection from chest X-ray images using hybrid social group optimization and support vector classifier. Cognit Comput. 2021. https://doi.org/10.1007/s12559-021-09848-3.
30. Varalakshmi P, Narayanan V, Rajasekar SJS. Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features. Appl Intell. 2021. https://doi.org/10.1007/s10489-020-01831-z.
31. Aljameel SS, Khan IU, Aslam N, Aljabri M, Alsulmi ES. Machine learning-based model to predict the disease severity and outcome in COVID-19 patients. Hindawi Sci Program. 2021. https://doi.org/10.1155/2021/5587188.