Utilization of machine learning algorithms in COVID-19 classification and prediction: A review

Shahad Abdelmuniem Mohamed Awadelkarim[1], Abdelrahman Elsharif Karrar[2]

[1]Department of Computer Science, National Ribat University, Sudan, [2]College of Computer Science and Engineering, Taibah University, Saudi Arabia

Abstract

Since 2019, more than 600 million people worldwide have contracted the rapidly growing COVID-19 disease. For healthy people to be protected and for patients to receive the best medical care, early COVID-19 virus detection is essential. Furthermore, due to the similarities between the symptoms of COVID and pneumonia, it was somewhat challenging to classify them. This survey reveals notable findings in several studies that combined many machine learning algorithms and the accuracy rates frequently utilized when using massive datasets of CT scans and X-ray images.

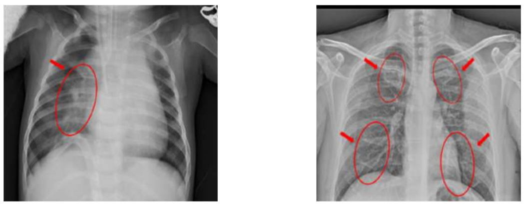
Keywords

COVID-19, Pneumonia, Classification, Prediction, Machine Learning

1. Introduction

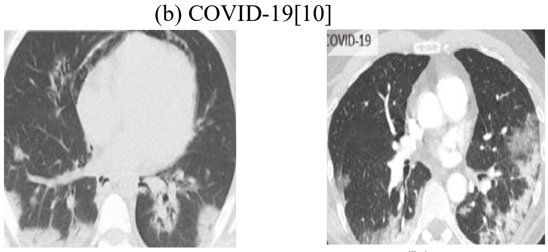
After its discovery in Wuhan, China, in 2019 [1] [2], COVID-19 overflowed to become one of the major pandemics. This aerobic disease named Coronavirus, caused by the SARS-CoV-2 virus, has spread to most countries. According to the World Health Organization (WHO), more than 600 million patients have been diagnosed globally since the pandemic [3] [4], and over 6 million deaths; were set till November 2022 [4]. This virus behaves similarly to pneumonia. However, it has a higher fatality rate if not treated immediately. WHO defines pneumonia as a swelling (inflammation) of the tissue in one or both lungs. Usually, it caused due to a bacterial infection or a virus[5]. Cough, difficulty breathing, high temperature, feeling generally unwell, sweating and shivering, loss of appetite, and rapid heartbeat are all common symptoms of pneumonia [6].

Coronavirus disease (COVID-19) is an infectious illness caused by the SARS-CoV-2 virus [7]. When an infected person coughs; sneezes; speaks; or breathes; the virus can spread through their mouth or nose in microscopic liquid particles. Most people infected with the virus will experience temperate to moderate aerobic illness and recover without requiring special treatment. Nevertheless, other people could become seriously ill and require medical attention. Older people and those with underlying medical conditions are more likely to develop serious illnesses. Anyone can get infected with COVID-19 and become seriously ill or die at any age [7] [8]. There is currently no cure for this condition. But, the average time to recover takes 5 to 6 days. Some people can take 14 days to be well. The only prevention from this virus is to get vaccinated [9]. COVID-19 prognostic procedure may use Chest X-rays and CT scans. Doctors can envisage changes in the lungs that arise because of COVID-19 pneumonia. Imaging samples of infected chests with pneumonia and COVID showed in (fig.1) and (fig.2) using X-ray and CT scan images, respectively.



(a) (b)

Fig.1: X-ray image of an infected chest with (a) Pneumonia (b) COVID-19[10]



(a) (b)

Manuscript received March 5, 2025
Manuscript revised March 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.3.10

Fig. 2: CT scan image of an infected chest with (a) Pneumonia (b) COVID-19 [11]

Manual testing of this infectious disease includes testing kits in sparse availability, too costly, and inefficient blood tests, where a blood test requires 5 to 6 hours to reveal the result. Hence, several researchers used machine learning to diagnose the disease from Chest Radiographs (CXR) or CT images [12]. In order to predict illnesses, data mining and machine learning have been extensively used in COVID-19 prediction [4]. Moreover, data mining and machine learning models helped medical researchers to identify relationships between variables and make them able to prognoses the outcome of disease using the historical datasets and helped in decision making [4, 5]. Therefore, this research aims to review machine learning and data mining techniques for pneumonia and COVID-19 classification and prediction.

This paper is structured as follows: section 2 describes machine learning and several machine learning algorithms used to predict diseases. Summaries of the recent research related to the classification and prediction of COVID-19 are in section 3. Section 4 contains a discussion of the summarized papers. Lastly, the study concluded in section 5.

2. Machine Learning Algorithms

Machine learning is an autonomous learning technique; used to educate models on how to treat data more effectively [13]. Machine learning algorithms gain knowledge from previous datasets, which are collections of a large amount of data. The machine learning model learns from data by evaluating it, and then, using the learning model, it generates a prediction [14]. The following machine learning algorithms are commonly used for COVID categorization and prediction:

A. Support Vector Machine (SVM)

It is a supervised learning technique employed in problems involving classification and regression. It consists of theoretical and numeric functions to solve most regression problems. It is a powerful machine-learning technique based on 3D and 2D modeling [8, 15, 16].

B. K- Nearest Neighbor (KNN)

The k-nearest neighbors (KNN) algorithm is a simple, supervised machine-learning technique for solving classification and regression problems. KNN is simple to use and comprehend. However, it has an issue of becoming noticeably slower as the size of that dataset increase [17, 18].

C. Decision Tree Algorithms

Decision tree algorithms are successful machine learning classification methods. They are the supervised learning techniques that use information gained and pruned to enhance results. Furthermore, decision tree algorithms are commonly used for classification in numerous research; for example, in the medical fields and health care issues. Decision tree algorithms come in several forms, including ID3, C4.5, and J48 [8, 13].

D. Naïve Bayes

Naive Bayes classifiers are a set of classification algorithms based on Bayes' Theorem. Naive Bayes is a classification algorithm for binary (two-class) and multi-class classification problems. It is a family of algorithms that all share a mutual standard in which every pair of classified features is independent of each other. The technique uses the Bayesian method to compute probability or likelihood. It accurately determines the frequency of noisy data supplied as input [14, 15].

E. Random Forests

Random forests are groups of classification and regression trees, which are simple models using binary splits on predictor variables to locate outcome predictions. Many classification and regression trees are formed in the random forest setting using randomly chosen training datasets and random selections of predictor variables for modeling outcomes. The scores from each tree are combined to create a prediction. The random forest has a significant advantage in prediction modeling; it can handle datasets with multiple predictor variables [16].

F. Neural Networks

An Artificial Neural Network (ANN) is a model of facts based on how the human brain processes information. ANN consists of interconnected layers called input, hidden, and output layers. The neural networks receive the data at the input layer, processing the data by a hidden layer and then providing the training results to the output layer. The network learns by setting the weights in the learning phase to predict the correct class label of the input [13]. There are many different kinds of neural networks, including convolutional neural networks (CNNs), recurrent neural networks (RNNs), probabilistic neural networks (PNNs), and more [13, 14].

G. Deep Learning

Deep learning is a subfield of machine learning that gradually separates higher-level features from the input data using many layers. Deep learning models are sometimes known as deep neural networks because most deep learning techniques use neural network architecture [15]. The expression (deep) refers to the volume of hidden layers in the neural network. Deep neural networks can have up to 150 hidden layers, compared to typical neural networks' 2-3 hidden layers. Training deep learning models use vast amounts of labeled data and neural network architecture that learn attributes from the data, eliminating the need for manual feature extraction [16, 17].

H. Logistic Regression

This method predicts a categorical dependent variable's output. Therefore, the result must be a categorist or discrete value. True or false, 0 or 1, or Yes or No are all possible outcomes. It gives the probabilistic values which lie between 0 and 1. Except for how they are applied, it is similar to linear regression. Linear regression is utilized for solving Regression problems, while Logistic regression is for solving classification problems [18].

I. Exponential Smoothing

Exponential smoothing is a time series prediction method for urinary data. All Exponential smoothing forecasting types are similar in that a prediction is a weighted sum of past observations, except the model explicitly uses an exponentially decreasing weight for observed data. Exponential smoothing has three types: Single, Double, and Triple Exponential smoothing [19] [20].

3. Literature Review

The review aim was to focus on the recent papers from 2018 until now that contained the classification and prediction of COVID-19 using the most known Machine Learning techniques. Based on the keywords (COVID-19, pneumonia classification, and machine learning), publications from IEEE, ELSEVIER, Springer, Hindawi, Science Report, and Research Gate are gathered.

This research uses data mining models to predict the recovery time for patients from COVID-19. The researchers used several algorithms on epidemiological datasets of COVID-19 patients; the algorithms were: The decision tree, support vector machine, naive Bayes, logistic regression, random forest, and K-nearest neighbor algorithm. They were applied directly to the dataset using python programming language to develop the models. The model predicted a minimum and a maximum number of days for COVID-19 patients to recover from the virus. The prediction was according to the age group of patients who are at high risk of not recovering from COVID-19, likely to recover, and might recover quickly from the disease. According to the study, the Decision Tree method created the most effective model, with a 99.85% accuracy rate [8].

Continuing with COVID-19, the researchers of this research are concerned with predicting the incidence of COVID-19 in Iran. They used linear regression and long short-term memory (LSTM) models to estimate the number of positive COVID-19 cases. Root Mean Square Error (RMSE) evaluated the model. The RMSE was 27.187, which is an acceptable rate [13]. The authors developed a hybrid 2D/3D CNN model based on convolutional neural networks in this study. The proposed model's testing trials revealed that it performed reasonably well when tested against a dataset that included three classes to predict: COVID-19, Pneumonia, and Normal. The model achieved a sensitivity of 98.33%, a specificity of 98.68%, and overall accuracy of 96.91% [14].

Similar to the previous study, this research also classified chest X-ray images according to the same three labels from previous research. The difference between them is that the image classification basis on Dense Convolution Neural Networks (DNN) and transfer learning helps to avoid the problem of DNN, which tends to overfit when trained on small datasets and many inputs. Transfer learning involves retraining (or fine-tuning) a network to solve a different problem on another dataset. By doing this, representations learned by the DNN in the first database can help the model generalization in the second. It is helpful when the first dataset is much bigger than the second one. According to this method, the accuracy rate reached 100% on the test dataset [15].

On the other hand, this research aimed to extract the most relevant variables to enhance patient risk prediction. It also presents a machine learning system that assesses risk prediction in COVID-19 patients. The proposed model was able to process the clinical, radiological, and laboratory data to predict the disease. Researchers used several models combined in a 10-fold cross-validation scheme. The models were: Boruta, Random Forest (RF), and Associative Trees (ATs). Although, the research had some limitations in referring to the difference in robustness between RF and ATs. The few numbers of samples in the dataset were another issue where the proposed system had only 300 data to use because no larger sample sizes were available yet [13]. Random Forest, Multi-Layer Perceptron, and Support Vector Machines Regression assessed clinical factors to predict COVID-19 and screened patients by hospitalization units. Random Forest test results showed an accuracy rate above 96% [14].

Likewise, this research used Linear Regression (LR) model to predict the confirmed, recovered, and death cases in India caused by Corona virus during the pandemic. Many datasets from different cities tested this model. Using the R-square score metric, LR presented 92% as an accuracy rate [15]. Researchers in this study created a novel method depending on Complete Blood Count (CBC) biomarkers. They used the Random Forest method for feature selection from 375 patients from Wuhan Hospital in China and 103 from Dhaka Medical College Hospital in Bangladesh, respectively. The testing phase was processed based on two categories: survival and death. Results presented that 46.4% died and 53.6% survived in Wuhan hospital. The model where 59.2% of patients survived and 40.8 % of patients died in Dhaka hospital. Logistic regression classified data among several testing classifiers and had the highest accuracy rate of 88% [16]. The authors of this study used several models to predict COVID-19 from chest X-ray images. First, they used two techniques for feature extraction; techniques were CheXNet and a handcrafted one. Then they used Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) methods to pick the most critical features. Finally, researchers tested the dataset using six machine learning models: K-Nearest Neighbor (KNN), Random forest (RF), Extra Tree (ET), XGboost, Bagging, and Support Vector Machine (SVM). Testing results showed that the most accurate model was SVM of 90% rate [17].

A study to predict COVID-19 was published in 2021 [13]. The research objective was to use wearable technology, such as smartwatches, to identify the disease and then use machine learning algorithms to analyze the data. Six models, including XGBoost, k-nearest neighbor (KNN), support vector machine (SVM), logistic regression, decision tree, and random forest, were tested on a dataset of 54 samples. According to the testing results, the most accurate learning model was the KNN rate reached 78%. On the other hand, this research is concerned with predicting mortality at different time windows in COVID-19 patients within the intensive care unit. Researchers created a novel model based on a heterogeneous graph model (HGM). Experiments were performed on 1269 patients using data from their electronic health records (EHRs). Research results showed a considerable increase of 30 to 40% above other models [14]. In this research [15], researchers used a combination of two algorithms to predict COVID-19. The proposed prediction model used the Harris hawks' optimization (HHO) to optimize the Fuzzy K-nearest neighbor (FKNN), called HHO-FKNN. This model was used to recognize the severity level of COVID-19. Based on testing results, HHO-FKNN had the highest results compared with several machine learning models like Random Forest and Support Vector Machine (SVM).

In the same context, authors in this research proposed a dynamic fusion-based federated method for analyzing CT scan images to detect COVID-19 infection. This method had two parties: the client and the server, with the client participating with the server to select data for inclusion in the model. This process helped to surpass the patient privacy restrictions. The testing results revealed that the proposed method was functional and performed better than the default setting of a federated learning [16]. In the early stages of the pandemic, the authors of this study used a modeling framework based on Capsule Networks, referred to as the COVID-CAPS. This model evaluated a limited dataset of X-ray pictures. The testing process revealed great accurate results above 95% [17]. In [18], authors used a method based on transfer learning to screen COVID-19 in X-Ray images, which they called the method Domain Extension Transfer Learning (DETL). CNN classified 1277 images among four classes: Normal, other diseases, pneumonia, and Covid−19. This model had three layers of VGGNet, AlexNet, and ResNet. The model achieved the highest accuracy, at 90.13%, in VGGNet. The research focused on improving feature selection and classification of COVID-19 by using CNN, Guided Whale Optimization Algorithm (Guided WOA), and aggregating classifiers to choose the most voted class [19]. Classifiers included Support Vector Machine (SVM), Neural Networks (NN), k-Nearest Neighbor (KNN), and Decision Trees (DT). The study evaluated a dataset of active COVID-19 data on the illness and others with passive COVID-19 data. All data were CT images. Testing experiments showed significant results. In [20], a web-based tool predicting COVID-19 from chest radiographs (CXR). As part of the classification process, this tool utilized RestNet50, a CNN with 50 layers [21]. Results were significant in testing the dataset. About 99% of the testing data was accurate. Furthermore, RestNet50 also had been used in this study plus (ResNet101, ResNet152, InceptionV3, and Inception-ResNetV2) models which are all CNN models. These models had been tested on three different datasets that consist of CXR images. testing results showed that RestNet50 had the highest accuracy results on the 3 datasets at 96.1%, 99.5%, and 99.7% respectively [13]. This research also tested CXR images as a dataset using a method called CoroNet which is a convolutional neural network designed to identify COVID-19 cases using CXR images. Results were great with an 89.6% as accuracy rate. The only issue was the few numbers of samples on the dataset with less than 800 images in it [14].

Moreover, the authors in this research used a combination of two methods: multi-level Thresholding and Support Vector Machine (SVM) to classify COVID-19 from CXR images as well. classification results presented a high accuracy rate of 97.48% [15]. In this research, a novel deep learning model called Deep-COVIDNet had been proposed to rate the range of increase in the number of infected cases on a particular day and to predict the growth of infected cases in the future. The proposed model tested several heterogeneous features and used the projection interval of the model which was 7 days. This means the model can forecast the range of increases in cases for 7 days into the future. According to the experiments, the aim of the study has been achieved, which was to inform policymakers and researchers to make response strategies and help fight the spread of COVID-19 [16].

On the other hand, this research explained the ability of several machine learning models to predict COVID-19 based on multiple factors: the number of newly infected cases, the number of deaths, and the number of recoveries. The authors used Linear Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine (SVM), and Exponential Smoothing (ES) models. Experimental results showed that the most accurate model was ES among the other models in all factors [17]. Researchers used a combination model that consists of weakly supervised COVID-19 lesion localization combined with the activation regions produced by the deep classification network (DeCoVNet) and the unsupervised lung segmentation method. The proposed model tested a dataset of 630 3D CT volumes. Results revealed that the model had the highest accuracy rate compared to other tested models like SVM and Random forest [18]. Deep and Shallow Convolution Neural Networks predicted pneumonia and COVID-19 [19]. The authors used 3487 samples as a dataset which contains several samples of COVID-19 and pneumonia. Testing results showed that Deep CNN was more accurate in classifying and predicting with 99.7%. In this research [20], the authors used 2685 samples to classify and predict COVID. They developed and used Random Forests for classification with 87.9%.

This paper [21] tested 326 CT images to predict COVID-19 using the Support Vector Machine (SVM) model. The model achieved a 95% accuracy score for radiomics features. The authors' goal in this research was to develop a computer-aided system to detect positive COVID-19 cases. The system used CNN to extract the features from CXR images. Classification experiments used models such as Decision trees, Random Forest, Neural Networks (NN), Naive Bayes, Logistic Regression, and k-nearest neighbor. According to testing results, the most accurate model was NN at 97.24% [22]. In this study [23], the proposed model has been tested on two different datasets based on X-ray images; both had more than 2000 samples in each dataset. The proposed model was called Fusion and normalization features based on RNNLSTM (F-RNN-LSTM). SVM, ANN, ensemble, and KNN were the models used for classification. According to the testing results, accuracy rates were high in achievements above 70% in each model. However, F-RNN-LSTM had the most significant accuracy scores in both datasets at 89.36% and 87.82%.

Moreover, this research aimed to predict COVID-19 based on portable CXR images. Classification methods included XGB-L, XGB-Tree, Decision Tree, KNN, and Naive Bayes. A dataset of 558 samples was classified using the mentioned methods. The highest accuracy rate was 80.49% for Naïve Bayes [17]. The following study had a dataset of 5840 balanced chest X-ray images using the synthetic minority oversampling technique (SMOTE) to create an equal number of samples for each class. Then for classification, Logistic Regression, KNN, Decision Trees, Random Forest, Adaptive Boosting (AdaBoost), Naïve Bayes, and XG Boost(XGB) have been used to classify the COVID-19, Normal, and Pneumonia classes. According to the results, the most accurate classifier was XG Boost at 97.7% [18]. In this research, authors used chest CT images as a dataset of 100 samples to classify COVID-19 from pneumonia. In this research [19], Multiple processes extracted features from 100 CT images. Ensemble of Bagged Tree (EBT), SVM, Linear Regression, KNN, and Decision tree were the methods used for classification. The most accurate model was EBT at an 86.9% rate. Similar to the previous research, this one used CT scan images as dataset samples for classification. Several machine learning methods classified COVID: KNN, Random Forest, CNN, iCNN, and two transfer learning models. iCNN had the highest accuracy rate among other models scoring 99.3% [20].

Also, using CNN, the authors of this study tested this method on a dataset of 20000 chest X-ray images. Results were good at a 95% accuracy rate [21]. CT scan images also have been used in the dataset of this study [22]. The authors developed a novel method based on ResNet architecture and Grad-Cam. The accuracy of the testing results was 91%. Using 84 samples of Chest X-ray images in the dataset of this study [23], this dataset has been tested using a novel machine learning model based on a dimensionality reduction approach to generate an optimized set of features. The proposed model achieved a high accuracy rate of 94%.

Moreover, this study used the SVM method on 1100 chest X-ray images. Testing results were high, and the accuracy rate was above 97% [17]. Authors in this research developed a computer-aided detection tool for predicting COVID-19. The proposed model developed the SVM algorithm. This model tested a dataset of chest X-ray images and achieved a 98.8% accuracy rate [18]. On a dataset of 219 CT images, the authors of this study [19] developed a machine-learning classifier using the Extra Trees algorithm. Testing results showed that the classifier was above 95% accuracy rate. Authors in this research [20] developed a machine learning model based on a convolutional neural network (CNN). The model trained and tested about 4300 chest X-ray images. The proposed model achieved a significant rate of 95.5%. In [21], an Artificial Neural Network-based proposed model classified a dataset containing X-ray and CT scan images. The model presented good scores of accuracy with 97.19%.

4. Discussion

This research summarizes multiple works of literature that utilized different machine-learning algorithms to classify and predict COVID-19. A summary of the papers considered in this research is in Table.1. The observation is in the classification of COVID-19 and pneumonia; the accurate models were the neural networks with the least accuracy rate above 91%. Also, results were better when using a combination of algorithms for prediction better than a single algorithm. Moreover, several papers used X-ray images for predicting COVID-19 since they are the most readily available information. However, we should keep in mind that each study varies from the other according to the type of datasets, the processes done on them, and the number of samples used in the proposed models.

Table 1: Summary of reviewed papers on COVID-19

Ref #	Research Title	Algorithm	Data Type	Accuracy Rate
[8]	Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery	Decision Tree	Patient report	99.85%
[13]	Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks	CNN (ResNet50)	X-Ray images	99.7%
[14]	CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images	Novel model (CoroNet)	Chest Radiographs (CXR)	89.6%
[15]	Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine	Combination of Multi-level Thresholding and (SVM)	Chest Radiographs (CXR)	97.48%
[16]	Deepcovidnet: An interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions	Novel model (Deep-COVIDNet)	Influencing factors	-
[17]	Predicting COVID-19 incidence through analysis of Google trends data in Iran: data mining and deep learning pilot study	Linear regression and long short-term memory (LSTM)	Google search data (Corona virus disease)	Above 90%
[18]	a novel hybrid 2D/3D CNN based on cross-domain adaptation approach for COVID-19 screening from chest X-ray images	Novel model (hybrid 2D/3D CNN)	X-Ray images	96.91%
[19]	A deep convolutional neural network for COVID-19 detection using chest X-rays	DNN and transfer learning	X-Ray images	100%
[20]	Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments	Random Forest	Patient report + CXR image	74%
[21]	Machine learning applied in SARS-CoV-2	Random	Clinical	Above

	COVID-19 screening using clinical analysis parameters	Forest	parameters	96%
[22]	Analysis and predictions of spread, recovery, and death caused by COVID-19 in India	Linear Regression (LR)	Reported cases	92%
[23]	Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters	logistic regression	Complete Blood Count (CBC) biomarkers	88%
[24]	Applying different machine learning techniques for prediction of COVID-19 severity	SVM	X-Ray images	90%
[25]	COVID-19 Diagnosis at Early Stage Based on Smartwatches and Machine Learning Techniques	KNN	Heart rate and number of steps	78%
[26]	Relational Learning Improves Prediction of Mortality in COVID-19 in the Intensive Care Unit	Novel model (Heterogeneous Graph Model (HGM))	Lead time window (time utilized before outcome)	87.5%
[27]	Diagnosing coronavirus disease 2019 (COVID-19): Efficient Harris Hawks-inspired fuzzy K-nearest neighbor prediction methods	(HHO-FKNN)	Patient features (binary data)	94%
[28]	Dynamic-fusion-based federated learning for COVID-19 detection	Dynamic Fusion-Based Federated	CT scan images	-
[29]	COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images	Capsule Network	X-Ray images	Above 95%
[30]	Deep learning for screening COVID-19 using chest X-ray images	Novel model (Domain Extension Transfer Learning (DETL))	X-Ray images	90.13%
[31]	Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images	PSO-GuidedWOA	CT scan images	95%
[32]	A web-based Diagnostic Tool for COVID-19 Using Machine Learning on Chest Radiographs (CXR)	CNN (ResNet50)	Chest Radiographs (CXR)	99%
[33]	COVID-19 future forecasting using supervised machine learning models	Exponential Smoothing (ES)	Daily case report	98%
[34]	A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT	deep classification network (DeCoVNet)	3D CT images	90%
[35]	Can AI Help in Screening Viral and COVID-19 Pneumonia?	Deep CNN	Chest Radiographs (CXR)	99.7%
[36]	Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification	Random Forest	CT images	87.9%

[37]	Machine learning-based CT radiomics model distinguishes COVID-19 from non-COVID-19 pneumonia	SVM	CT images	95%
[38]	Automatic COVID-19 pneumonia diagnosis from x-ray lung image: A Deep Feature and Machine Learning Solution	Neural Network (NN)	Chest Radiographs (CXR)	97.24%
[39]	Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques	F-RNN-LSTM	Chest Radiographs (CXR)	98.36%
[40]	Machine learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection	Naïve Bayes	Chest Radiographs (CXR)	80.49%
[41]	Accurate prediction of COVID-19 using chest X-ray images through deep feature learning model with SMOTE and machine learning classifiers	XG Boost	Chest Radiographs (CXR)	97.7%
[42]	Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning	EBT	CT images	86.49%
[43]	Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs	iCNN	CT images	99.3%
[44]	Pneumonia and COVID-19 Detection using Convolutional Neural Networks	CNN	Chest Radiographs (CXR)	95%
[45]	Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients	Novel model (based on ResNet and Grad-Cam)	CT images	91.%
[46]	COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images	Novel model (Dimensionality Reduction)	Chest Radiographs (CXR)	94%
[47]	Machine Learning Model Applied on Chest X-Ray Images Enables Automatic Detection of COVID-19 Cases with High Accuracy	SVM	Chest Radiographs (CXR)	Above 97%
[48]	Development of a computer-aided tool for detection of COVID-19 pneumonia from CHEST RADIOGRAPHS (CXR) images using machine learning algorithm	SVM	Chest Radiographs (CXR)	98.8%
[49]	The study of automatic machine learning base on radiomics of non-focus area in the first chest CT of different clinical types of COVID-19 pneumonia	Extra Trees	CT images	Above 95%
[50]	CovXR: Automated Detection of COVID-19 Pneumonia in Chest X-Rays through Machine Learning	Novel model (based on CNN)	Chest Radiographs (CXR)	95.5%
[51]	Identification of pneumonia disease applying an intelligent computational framework based on deep learning and machine learning techniques	Novel model (based on ANN)	CXR+CT images	97.19%

5. Conclusion

The deadly COVID-19 pandemic has a massive impact on people's lives across the globe, and it can lead to death if the diagnosis is not at the early stages. This paper summarized the recent studies in the classification and prediction of COVID-19 from several publishers. The discussion showed that neural network models were the most used and performed the best among other models in the reviewed studies. Also, the observation showed that a good and large dataset provides better prediction accuracy. Furthermore, X-ray images had the most share selection as dataset samples due to cost-effectiveness and availability compared to CT scan images. According to these researches, data mining, and machine learning algorithms may prospectively help in the treatment followed by extracting information and knowledge from suitable datasets.

References

[1] Y. Chen *et al.*, "New-onset autoimmune phenomena post-COVID-19 vaccination," vol. 165, no. 4, pp. 386-401, 27-12-2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1111/imm.13443

[2] A. M. Maatuk, E. K. Elberkawi, S. Aljawarneh, H. Rashaideh, and H. Alharbi, "The COVID-19 pandemic and E-learning: challenges and opportunities from the perspective of students and instructors," *Journal of Computing in Higher Education,* vol. 34, no. 1, pp. 21-38, 2022.

[3] "COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE)." Johns Hopkins University (JHU). https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 (accessed 2-10-2022, 2022).

[4] W. H. ORGNIZATION. "WHO Coronavirus (COVID-19) Dashboard." https://covid19.who.int/ (accessed 29-11-2022, 2022).

[5] W. H. ORGNIZATION. "Pneumonia." https://www.who.int/health-topics/pneumonia#tab=tab_1 (accessed 20-11-2022, 2022).

[6] N. H. SERVICE. "Overview Pneumonia." https://www.nhs.uk/conditions/pneumonia/ (accessed 25-11-2022, 2022).

[7] W. H. Orgnization. "Coronavirus disease (COVID-19)." https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed 3\9\2022, 2022).

[8] L. Muhammad, M. Islam, S. S. Usman, and S. I. Ayon, "Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery," *SN Computer Science,* vol. 1, no. 4, pp. 1-7, 2020.

[9] T. T. Le, Z. Andreadakis, A. Kumar, R. G. Román, S. T. Saville, and S. Mayhew, "The COVID-19 vaccine development landscape," *Nature Reviews | Drug DISCOVERY,* vol. 19, 2020, doi: 10.1038/d41573-020-00073-5.

[10] J. E. Luján-García, M. A. Moreno-Ibarra, Y. Villuendas-Rey, and C. Yáñez-Márquez, "Fast COVID-19 and pneumonia classification using chest X-ray images," *Mathematics,* vol. 8, no. 9, p. 1423, 2020.

[11] G. Meirelles *et al.*, "Padrão típico para pneumonia viral," ed. COVID-19 ASPECTOS DE IMAGEM: Grupo Fleury, 2020.

[12] C. Fan, Z. Zeng, L. Xiao, and X. Qu, "GFNet: Automatic segmentation of COVID-19 lung infection regions using CT images based on boundary features," *Pattern Recognition,* vol. 132, p. 108963, 2022.

[13] A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks," *Pattern Analysis and Applications,* vol. 24, no. 3, pp. 1207-1220, 2021.

[14] A. I. Khan, J. L. Shah, and M. M. Bhat, "CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images," *Computer methods and programs in biomedicine,* vol. 196, p. 105581, 2020.

[15] L. N. Mahdy, K. A. Ezzat, H. H. Elmousalami, H. A. Ella, and A. E. Hassanien, "Automatic X-ray COVID-19 Lung Image Classification System based on Multi-Level Thresholding and Support Vector Machine," *MedRxiv,* 2020.

[16] A. Ramchandani, C. Fan, and A. Mostafavi, "Deepcovidnet: An interpretable deep learning model for predictive surveillance of covid-19 using heterogeneous features and their interactions," *Ieee Access,* vol. 8, pp. 159915-159930, 2020.

[17] H. Z. S. M. Ayyoubzadeh, M. Ahmadi, and S. R. N. Kalhori, "Predicting COVID-19 incidence through analysis of google trends data in Iran: data mining and deep learning pilot study," vol. 6. [Online]. Available: https://publichealth.jmir.org/2020/2/e18828

[18] K. Bayoudh, F. Hamdaoui, and A. Mtibaa, "Hybrid‑COVID: a novel hybrid 2D/3D CNN based on cross‑domain adaptation approach for COVID‑19

Manuscript received March 5, 2025
Manuscript revised March 20, 2025
https://doi.org/10.22937/IJCSNS.2025.25.3.10

screening from chest X‑ray images," Scientific paper 2020, doi: 43:1415–1431. Physical and Engineering Sciences in Medicine (2020).
[19] P. R. Bassi and R. Attux, "A deep convolutional neural network for COVID-19 detection using chest X-rays," *Research on Biomedical Engineering,* vol. 38, no. 1, pp. 139-148, 2022.
[20] E. Casiraghi *et al.*, "Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments," *Ieee Access,* vol. 8, pp. 196299-196325, 2020.
[21] R. F. A. P. de Oliveira, C. J. A. Bastos Filho, A. C. A. de Medeiros, P. J. B. L. dos Santos, and D. L. Freire, "Machine learning applied in SARS-CoV-2 COVID 19 screening using clinical analysis parameters," *IEEE Latin America Transactions,* vol. 19, no. 6, pp. 978-985, 2021.
[22] R. Kumari *et al.*, "Analysis and predictions of spread, recovery, and death caused by COVID-19 in India," *Big Data Mining and Analytics,* vol. 4, no. 2, pp. 65-75, 2021.
[23] T. Rahman *et al.*, "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters," *Ieee Access,* vol. 9, pp. 120422-120441, 2021.
[24] S. A.-F. Sayed, A. M. Elkorany, and S. S. Mohammad, "Applying different machine learning techniques for prediction of COVID-19 severity," *Ieee Access,* vol. 9, pp. 135697-135707, 2021.
[25] J. Skibinska, R. Burget, A. Channa, N. Popescu, and Y. Koucheryavy, "COVID-19 Diagnosis at Early Stage Based on Smartwatches and Machine Learning Techniques," *IEEE Access,* vol. 9, pp. 119476-119491, 2021.
[26] T. Wanyan *et al.*, "Relational Learning Improves Prediction of Mortality in COVID-19 in the Intensive Care Unit," *IEEE transactions on big data,* vol. 7, no. 1, pp. 38-44, 2020.
[27] H. Ye *et al.*, "Diagnosing coronavirus disease 2019 (COVID-19): Efficient Harris Hawks-inspired fuzzy K-nearest neighbor prediction methods," *Ieee Access,* vol. 9, pp. 17787-17802, 2021.
[28] W. Zhang *et al.*, "Dynamic-fusion-based federated learning for COVID-19 detection," *IEEE Internet of Things Journal,* vol. 8, no. 21, pp. 15884-15891, 2021.
[29] P. Afshar, S. Heidarian, F. Naderkhani, A. Oikonomou, K. N. Plataniotis, and A. Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," *ELSEVIER,* 2020.
[30] S. Basu, S. Mitra, and N. Saha, "Deep learning for screening COVID-19 using chest X-ray images," *ELSEVIER,* 2020.
[31] E.-S. M. El-Kenawy, A. Ibrahim, S. Mirjalili, M. M. Eid, and S. E. Hussein, "Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images," *IEEE access,* vol. 8, pp. 179317-179335, 2020.
[32] E. B. G. Kana, M. G. Z. Kana, A. F. D. Kana, and R. H. A. Kenfack, "A web-based Diagnostic Tool for COVID-19 Using Machine Learning on

Chest Radiographs (CXR)," *MedRxiv,* 2020.
[33] F. Rustam *et al.*, "COVID-19 future forecasting using supervised machine learning models," *IEEE access,* vol. 8, pp. 101489-101499, 2020.
[34] X. Wang *et al.*, "A weakly-supervised framework for COVID-19 classification and lesion localization from chest CT," *IEEE transactions on medical imaging,* vol. 39, no. 8, pp. 2615-2625, 2020.
[35] M. E. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?," *IEEE Access,* vol. 8, pp. 132665-132676, 2020.
[36] F. Shi *et al.*, "Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification," *Physics in medicine & Biology,* vol. 66, no. 6, p. 065031, 2021.
[37] H. J. Chen *et al.*, "Machine learning-based CT radiomics model distinguishes COVID-19 from non-COVID-19 pneumonia," *BMC infectious diseases,* vol. 21, no. 1, pp. 1-13, 2021.
[38] M. Ahmed, Z. Al-qaysi, M. L. Shuwandy, M. M. Salih, and M. H. Ali, "Automatic COVID-19 pneumonia diagnosis from x-ray lung image: A Deep Feature and Machine Learning Solution," in *Journal of Physics: Conference Series*, 2021, vol. 1963, no. 1: IOP Publishing, p. 012099.
[39] S. Goyal and R. Singh, "Detection and classification of lung diseases for pneumonia and Covid-19 using machine and deep learning techniques," *Journal of Ambient Intelligence and Humanized Computing,* pp. 1-21, 2021.
[40] L. Hussain *et al.*, "Machine-learning classification of texture features of portable chest X-ray accurately classifies COVID-19 lung infection," *BioMedical Engineering OnLine,* vol. 19, no. 1, pp. 1-18, 2020.
[41] R. Kumar *et al.*, "Accurate prediction of COVID-19 using chest X-ray images through deep feature learning model with SMOTE and machine learning classifiers," *MedRxiv,* 2020.
[42] C. Liu, X. Wang, C. Liu, Q. Sun, and W. Peng, "Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning,"

Biomedical engineering online, vol. 19, no. 1, pp. 1-14, 2020.

[43] L. Saba *et al.*, "Six artificial intelligence paradigms for tissue characterisation and classification of non-COVID-19 pneumonia against COVID-19 pneumonia in computed tomography lungs," *International journal of computer assisted radiology and surgery,* vol. 16, no. 3, pp. 423-434, 2021.

[44] S. V. Militante, N. V. Dionisio, and B. G. Sibbaluca, "Pneumonia and COVID-19 Detection using Convolutional Neural Networks," *2020 the third International Conference on Vocational Education and Electrical Engineering (ICVEE),* 2020.

[45] S. Sharma, "Drawing insights from COVID-19-infected patients using CT scan images and machine learning techniques: a study on 200 patients," *Environmental Science and Pollution Research,* vol. 27, no. 29, pp. 37155-37163, 2020.

[46] A. Zargari Khuzani, M. Heidari, and S. A. Shariati, "COVID-Classifier: An automated machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images," *Scientific Reports,* vol. 11, no. 1, pp. 1-6, 2021.

[47] Y. Erdaw and E. Tachbele, "Machine Learning Model Applied on Chest X-Ray Images Enables Automatic Detection of COVID-19 Cases with High Accuracy," *International Journal of General Medicine,* vol. 14, 2021.

[48] N. Absar *et al.*, "Development of a computer-aided tool for detection of COVID-19 pneumonia from CXR images using machine learning algorithm," *ELSEVIER,* Journal of Radiation Research and Applied Sciences 15 (2022) pp. 32-43, 2022.

[49] H.-B. Tan *et al.*, "The study of automatic machine learning base on radiomics of non-focus area in the first chest CT of different clinical types of COVID-19 pneumonia," *Scientific reports,* vol. 10, no. 1, pp. 1-10, 2020.

[50] V. Shenoy and S. Malik, "CovXR: Automated Detection of COVID-19 Pneumonia in Chest X-Rays through Machine Learning," in *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2021: IEEE, pp. 1-6.

[51] Y. Muhammad, M. D. Alshehri, W. M. Alenazy, T. Vinh Hoang, and R. Alturki, "Identification of pneumonia disease applying an intelligent computational framework based on deep learning and machine learning techniques," *Mobile Information Systems,* vol. 2021, 2021.