



Machine Learning Approaches for Tackling Novel Coronavirus (COVID-19) Pandemic

Mohammad Marufur Rahman¹ · Md. Milon Islam¹ · Md. Motaleb Hossen Manik¹ · Md. Rabiul Islam² · Mabrook S. Al-Rakhami³

Received: 5 June 2021 / Accepted: 11 July 2021 / Published online: 19 July 2021
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2021

Abstract

Novel coronavirus (COVID-19) has become a global problem in recent times due to the rapid spread of this disease. Almost all the countries of the world have been affected by this pandemic that made a major consequence on the medical system and healthcare facilities. The healthcare system is going through a critical time because of the COVID-19 pandemic. Modern technologies such as deep learning, machine learning, and data science are contributing to fight COVID-19. The paper aims to highlight the role of machine learning approaches in this pandemic situation. We searched for the latest literature regarding machine learning approaches for COVID-19 from various sources like IEEE Xplore, PubMed, Google Scholar, Research Gate, and Scopus. Then, we analyzed this literature and described them throughout the study. In this study, we noticed four different applications of machine learning methods to combat COVID-19. These applications are trying to contribute in various aspects like helping physicians to make confident decisions, policymakers to take fruitful decisions, and identifying potentially infected people. The major challenges of existing systems with possible future trends are outlined in this paper. The researchers are coming with various technologies using machine learning techniques to face the COVID-19 pandemic. These techniques are serving the healthcare system in a great deal. We recommend that machine learning can be a useful tool for proper analyzing, screening, tracking, forecasting, and predicting the characteristics and trends of COVID-19.

Keywords Novel coronavirus · COVID-19 · Machine learning · Prediction · Forecasting

Introduction

The newly found coronavirus causes an infectious disease named coronavirus disease (COVID-19). COVID-19 was first reported at Wuhan, China in December 2019 and since

then it has spread in 216 countries around the world with cases over 10.6 million and more than half a million people have lost their valuable lives [1]. These statistics are rapidly increasing every day and the crisis is growing day by day due to this pandemic. Healthcare systems [2–4] are badly in need of technological support in this pandemic situation. The medical system is expecting new technologies to overcome the crisis. The experts all over the world are coming with various techniques [5–7] that can contribute to

This article is part of the topical collection “Advances in Computational Approaches for Artificial Intelligence, Image Processing, IoT and Cloud Applications” guest edited by Bhanu Prakash K N and M. Shivakumar.

✉ Mohammad Marufur Rahman
inkmarufnayem@gmail.com

Md. Milon Islam
milonislam@cse.kuet.ac.bd

Md. Motaleb Hossen Manik
mkmanik557@gmail.com

Md. Rabiul Islam
rabiulnewemail@gmail.com

Mabrook S. Al-Rakhami
malrakhami@ksu.edu.sa

¹ Department of Computer Science and Engineering, Khulna University of Engineering and Technology, Khulna 9203, Bangladesh

² Department of Electrical and Electronic Engineering, Bangladesh Army University of Engineering and Technology, Natore 6431, Bangladesh

³ Research Chair of Pervasive and Mobile Computing, Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia

addressing the challenging situation. It has become a very challenging task because clinical information about the diseases is not properly available yet and the information is changing overtimes [8, 9]. With these limited data, many researchers came with various helpful technologies using various machine learning algorithms [10, 11].

In many developing countries, there is a shortage of medical kits for the COVID-19 test, which may cause more infections. Machine learning techniques can provide initial support to find potentially infected individuals for different types of diseases like heart disease [12], diabetes [13], liver disorder [14], breast cancer [15, 16] as well as COVID-19 [17]. Machine learning is at the core of many new technologies. It can assist in quicker identification of possible cases so that interventions can be taken as early as possible. Machine learning techniques are used in tracking the spread of COVID-19, grouping patients with high risks, and also for diagnosis purposes. Fever and cough are the most common symptoms, which are also symptoms of many other infectious diseases. Machine learning can help to differentiate between COVID-19 and other diseases. Useful information regarding COVID-19 patients can be generated by machine learning algorithms which would give clinicians extra time and confidence while treating a serious patient.

This paper focuses on the importance and impact of machine learning in the duel between humans and COVID-19. It gives a detailed review of all the technologies that came up to fight coronavirus using machine learning algorithms. This paper depicts the important role played by machine learning in this critical condition. The reader will also find various technological advances that are introduced and can be used in the time of pandemic like COVID-19. Currently, these technologies are fighting against coronavirus and will be a valuable way to combat any future pandemic. The challenges with some future trends are outlined in this paper.

The rest of the paper is structured as follows. In second section, different machine learning algorithms are demonstrated. Third section, various applications of machine learning for tackling COVID-19 is discussed. Fourth section consists of discussions and recommendations for further research. The paper is concluded in fifth section.

Machine Learning Algorithms

Machine learning (ML) is one of the well-known branches of Artificial Intelligence (AI) and its algorithms hold the credit of AI advancement and application. The ML algorithms are developed for classical statistical data analysis. The various uses of ML have directed it to different fields of science to assist human and medical science is one of them.

Recently, medical science is focusing on COVID-19 a well-known pandemic all over the world. Researchers are setting their attention towards this and coming out with their proposed solutions to handle COVID-19. Machine learning algorithms are not at a back stage. They are being used to forecast, diagnosis, screening the patients and in many more braches of application. Among them, mostly used algorithms can be mentioned by Linear Regression (LR), Support Vector Machine (SVM), Multi-Layer Perceptron (MLP) and Vector Auto-Regression.

Linear Regression

LR is one of the supervised ML algorithms that carries through regression task. Basically, the goal of the model is to predict the desire value upon independent variables. The number of independent variables can be more than one at a time while the dependent variable is only one. Thus, they construct a linear relationship between independent variables (input) and dependent variable (output). The basic function that results the linear relationship is indicated as:

$$y = c_1 + c_2x \quad (1)$$

where y is the dependent variable, x is the independent variable, c_1 is the intercept and c_2 is the co-efficient. The equation tries to draw a linear line passing through maximum number of independent variables. During the training process, the model tries to find out the best values of c_1 and c_2 so that the difference between the value of y and the actual value is minimum. The measurement follows the following expression.

$$\text{minimize } \frac{1}{n} \sum_{i=0}^n (\text{predict}_i - y_i)^2 \quad (2)$$

The authors in [19, 41] prospected their model with regression algorithm for the purpose of focusing on confirmed cases, death cases and recovered cases daily. Correlation played a vital role on those models that showed an evidence on the acceptance of the outcomes for progressive days.

Support Vector Machine

Support Vector Machine is one of supervised machine learning algorithms that is powerful tool for both regression and classification, mostly used for classification. It is popular in case of categorical data. The basic idea lies in drawing hyperplanes through the data points that separate the whole dataset into different classes. The main goal is to find the maximum marginal hyperplane that divides the data points into different classes. The higher the gap between the hyperplanes the better the margin is considered. But the data points are not always separable or suitable to provide into

SVM algorithm. That is why a kernel trick is used in transforming the dimension into adequate dimension. The process can belong to any of them from linear kernel, polynomial kernel, and radial basis kernel. The kernel trick equations are as follows, respectively.

$$k(x, x_i) = \text{sum}(x * x_i) \quad (3)$$

$$k(x, x_i) = 1 + \text{sum}(x * x_i)^d \quad (4)$$

$$k(x, x_i) = e^{-(\text{gamma} * \text{sum}(x - x_i^2))} \quad (5)$$

Though SVM mechanism is simpler but it consumes higher training time thus resulting inconvenient to use for large dataset.

Multiple researchers including [27, 32, 37] applied SVM algorithm in their models to draw a conclusion on the number of days required to recover patients, to develop a circle of patients at a high risk and to generate a specific group of patients those are more likely to be recovered within a certain period. An online survey was also conducted in [37] for the collection of data with signs and symptoms of the infected patients.

Multi-Layer Perceptron

The idea of perceptron discovered from human biology through neuro-science. The way human neuron works is adopted in artificial neural network thus in Multi-Layer Perceptron (MLP). An MLP model is applied in supervised learning problems like classification, regression. An MLP model contains an input layer to receive the input data, an output layer to provide result on classification and in between these two layers, there are an arbitrary number of hidden layers that are the actual computational blocks. Throughout the training phase, the model learns the correlation between input and output by adjusting parameters called weight and bias. The basic goal is to minimize the error through a technique called backpropagation. The actual equation of a perceptron is as follows.

$$y = \phi \sum_{i=0}^n (w_i x_i + b) \quad (6)$$

Backpropagation mechanism tries to adjust the value of weight (w) and bias (b) with respect to the error that can be calculated in a various ways including root mean square error. Among two passes, in the process of forward pass, the input signal is moved towards hidden layers from input layer and finally in output layer to generate the decision with respect to ground label. On the other hand, in the backward pass, the weights and biases are propagated backward to

be adjusted via gradient-based optimization process including stochastic gradient descent. This backward process continues till the model reaches at a minimum error named convergence.

A comparative analysis between MLP and other machine learning model on the basis of infected, death and recovered patients have been drawn in [19, 24, 37]. Most of the cases MLP showed better result on decision-making in all aspects.

Vector Auto-Regression

Approximately all machine learning models provide the facility of unidirectional relationship between variables used in decision-making. Breaking the trend, Vector Auto-Regression provides the flexibility of handling the variables in bidirectional manner, i.e., one-step forward variables may have an impact on backward variables. Variables are modelled in such a manner that they influence one another equally that triggers to be named as “endogenous.” The model settles one equation for each variables. The right side of the equation contains a constant and lags of all the variables present in the whole system. The simplest equation can be shown as:

$$y_{1,t} = c_1 + \phi_{11,1} y_{1,t-1} + \phi_{12,1} y_{2,t-1} + e_{1,t} \quad (7)$$

$$y_{2,t} = c_2 + \phi_{21,1} y_{1,t-1} + \phi_{22,1} y_{2,t-1} + e_{2,t} \quad (8)$$

where $e_{1,t}$ and $e_{2,t}$ are white noise process. The coefficient $\phi_{ii,l}$ captures the impact of the l th lag of variable y_i and the coefficient $\phi_{ij,l}$ captures the impact of the l th lag of variables y_i and y_j . The forecast is generated in a recursive way for each variables in the model. The previously mentioned equations can be transformed as a new equation like mentioned below for the forecasting purpose by setting the error to zero.

$$\hat{y}_{1,T+1|T} = \hat{c}_1 + \phi_{11,1} y_{1,T} + \phi_{12,1} y_{2,T} \quad (9)$$

$$\hat{y}_{2,T+1|T} = \hat{c}_2 + \phi_{21,1} y_{1,T} + \phi_{22,1} y_{2,T} \quad (10)$$

The authors in [19] showed a time series data analysis on forecasting about the effect of COVID-19 in patients using VAR algorithm. The applied model expressed the wave of different cases throughout 69 days and finally concluding on an increase of infected, death and recovered patients.

Major Applications of Machine Learning Approaches

Researchers all over the globe are bringing various latest technologies to face a dangerous situation, the world is going through now. All the research areas that are being explored

to mitigate the COVID-19 pandemic are shown in Fig. 1. The forecasting COVID-19 outbreak, diagnosis, survey and screening, and environmental dependencies are the major research fields where machine learning is being used.

Forecasting COVID-19 Outbreak

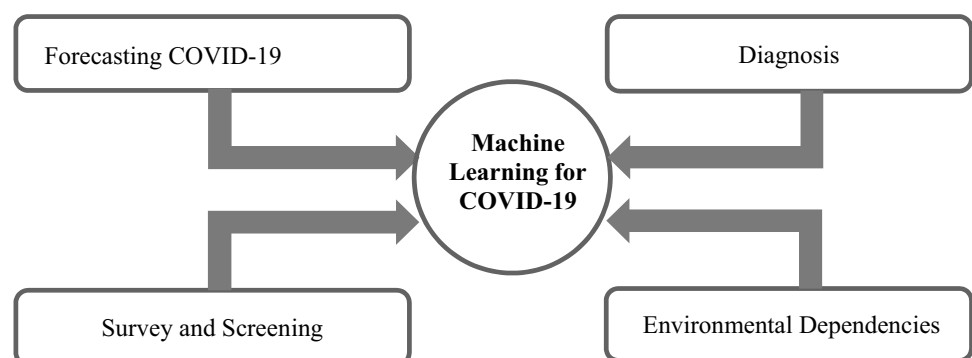
To adapt to the COVID-19 pandemic, forecasting aims to guide decisions on pandemic preparations, resource distribution, and adoption of social distance measures and other strategies. Machine learning techniques are widely used for forecasting COVID-19. Forecasting and predicting the growth of the COVID-19 pandemic using ML has become an important research topic and many researchers around the globe are contributing to this field.

Liu et al. [18] proposed a novel methodology using machine learning techniques that can forecast COVID-19 in real-time. Augmented ARGONet is used to predict COVID-19 outbreak using data from the internet, official health reports of China, news media, and daily forecast of COVID-19 activity. The developed system-generated 2 days ahead and real-time confirmed cases forecast for 32 provinces of China for the period between February 3 and February 21, 2020. The scheme clustered the provinces into several groups for training the machine learning architecture. The clustering and model training process was repeated for every prediction date. A study for forecasting the COVID-19 pandemic in India is presented by Sujath et al. [19]. The predicting model was developed using Linear Regression (LR), Multi-Layer Perceptron (MLP), and Vector Auto-Regression (VAR) learning algorithms. These were used to predict confirmed cases, death cases, and recovered cases daily. In this study, the authors claimed that the VAR model is the most suitable analysis model for forecasting. To predict the COVID-19 epidemic, an Ensemble Empirical Mode Decomposition-Artificial Neural Network (EEMD-ANN)-based hybrid architecture is proposed by Hasan [20]. The system consists of an ensemble empirical mode decomposition and an artificial neuron network for predicting the epidemic.

The statistics between January 22 and May 18, 2020, is used as the time series data for training the model. At first, EEMD decomposes the time series data and then the ANN is trained on the processed data. The proposed method was compared with some traditional statistical analysis and it outperformed all the traditional statistical approaches. Daily trends of cumulative confirm, recovery, and death cases of the global COVID-19 scenario were predicted in the study.

Punn et al. [21] proposed a machine learning-based epidemic analysis technique. The proposed scheme analyzed the increasing transmission at the beginning and a forecast was done on the possibilities of the transmission. Support vector regression, deep learning regression, and polynomial regression are used for the analysis. The system predicted the possible number of cases across the world for the next 10 days from the time of analysis. In their study, polynomial regression showed better performance on forecasting the transmission of COVID-19 than other techniques. Tiwari et al. [22] developed a forecasting technique based on time series for predicting outbreak trends of COVID-19 disease. This study tends to predict trends of the outbreak in India based on the pattern seen in China. The predictions were made on confirmed cases, recovery cases, and death cases in 22 days of the forecasting horizon. Li et al. [23] presented a model that predicts the trend of the epidemic across the world. The authors claimed that the epidemic will peak in China on February 22, 2020, and peak around the world on May 22, 2020. They also added that it will be under control in early April and late August 2020 respectively in China and around the world. Ardabili et al. [24] performed a comparative analysis of soft computing models and machine learning to predict the coronavirus outbreak. Among various machine learning models, an adaptive network-based fuzzy inference system (ANFIS) and MLP showed promising results. Evolutionary algorithms like genetic algorithm, grey wolf optimization, and particle swarm optimization are used to find appropriate parameters of the model. Though initial predictions were highly representative of the actual scenario, prediction beyond the 30 days' observation range

Fig. 1 Major applications of machine learning in COVID-19



was not realistic compared to the actual cases. The authors suggested that machine learning can be used as an effective tool for modeling the outbreak.

Ndiaye et al. [25] used the SIR model and machine learning techniques to forecast COVID-19. Classical Kermack Mckendrick SIR model is used to describe the COVID-19 transmission which is a compartmental model that is used to model how diseases spread through a population. The prophet procedure for forecasting time series data is used in this study of forecasting purposes. For time series that have several seasons of historical data and strong seasonal effects, it performed best. The system is also robust for missing data, outliers, and handles changes in the trends well. The cumulative number of confirmed and death cases are predicted in this study. The authors gave an optimistic assessment that in most of the countries the pandemic will be a culmination at the end of April. To track and predict the growth of the COVID-19 epidemic and plan policies as well as strategies, a machine learning, and cloud computing-based system, has been proposed by Tuli et al. [26]. In this study, an improved mathematical model is applied to predict and analyze the increase of the pandemic. Generalized Inverse Weibull distribution is fitted using iterative weighting to get a better fit for the prediction model. For the more precise and real-time prediction of the behavior of the epidemic, the model has been deployed on a cloud computing platform. The model is fitted on the distributions of several new cases and dead patients. The distribution included data till the start of May 2020 and the model predicted that the pandemic will come to an end at the end of October 2020 across the world. A data mining model is developed to predict the recovery of COVID-19 patients in the study of Muhammad et al. [27]. Machine learning algorithms such as Decision Tree, Naive Bayes, Support Vector Machine, Random Forest, K-Nearest Neighbor, and Logistic Regression are used to train the system. The minimum and the maximum number of days required to recover a patient from the diseases, the group of patients that have high risk, and the group of patients that are more likely to recover are predicted by the system. This study shows that the decision tree-based model performed well than other learning algorithms in predicting the recovery possibility of a patient. From this model, 99.85% prediction accuracy is achieved. A survival prediction framework is presented by Yan et al. [28]. The study used blood sample from COVID-19-infected patients to train supervised XGBoost classifiers. The model predicted the survival possibility with 90% accuracy.

There is no data at all at the beginning of the outbreak of the epidemic and making predictions becomes widely doubtful as time passes. The predictive models with a very small forecasting horizon provided some insightful predictions that may help to estimate the pandemic situation a little ahead of time. Now it is clear that all the long-term predictions on the

pandemic have completely failed. Machine learning algorithms highly depends on data samples to learn interesting features, relations, and dependencies to generalize unseen data. Poor data and past evidence hamper the learning of the algorithm. Forecasting requires a sufficient amount of historical data and at the same time, there is no guarantee that the future will repeat itself in the same way as the past. Since the COVID-19 is a novel pandemic, poor historical data at hand are not sufficient for a robust and long-term forecasting model. Also, psychological considerations play an important role in understanding and reaction of individuals to the risk of the disease and the anxiety that it may affect them directly. Since people's psychology is hard to predict, long-term forecasting of the epidemiology model becomes very hard and error-prone.

Some organizations have centralized facilities to predict the pandemic condition. The Centers for Disease Control and Prevention (CDC) is a national public health institute of the US that brings out forecasts of COVID-19 deaths, hospitalization, and cases per week for the next 4 weeks [29]. Approximately, 45 different models are used for forecasting using various types of data, e.g., COVID-19 data, mobility data, demographic data, etc. All the independent forecasts are aggregated into one forecast using the ensemble technique. The predictive model revealed accurate short-term projections while the accuracy starts to deteriorate at longer prediction horizons up to 4 weeks [30].

Though long-term forecast of the epidemic has failed in almost all the studies, for better understand the current situation and plan for the future, forecasting is invaluable. Forecasts of the pandemic scenario can be a valuable part of the decision-making process, especially in a high-risk situation.

Diagnosis

Diagnosis is a very important phase in identifying COVID-19 patients. The physicians are facing a hard time in this phase. Machine learning is a dependable tool for them while taking important decisions quickly and more confidently.

Yan et al. [31] proposed a machine learning-based model for the prognosis of COVID-19 disease. The system developed the prediction model using a supervised XGBoost classifier. The proposed scheme was trained on epidemiological, clinical, demographic, medication, laboratory, and nursing records which are extracted from electronic records of 2779 suspected COVID-19 patients between the 10th of January and 18th of February, 2020 at Tongji Hospital in Wuhan, China. The parameters of the Multi-Tree XGBoost were max depth with 4, the number of estimators was 150 and the learning rate was set to 0.2. Single-tree XGBoost was trained for final prediction because of its explainability. The framework identified three key features from the patient's clinical reports which are lactic dehydrogenase (LDH), lymphocyte,

and high sensitivity C-reaction protein (hs-CRP) to precisely and quickly assess the risk of death. The study found that the most frequent initial symptom was fever (49.9%), followed by cough (13.9%), nausea (3.7%), and dyspnea (2.1%). The developed model successfully predicted 100% death cases with precision 0.89 and 90% survival cases with recall 0.83. F1-score for death and survival case are respectively 0.94 and 0.91 from the experiment. Batista et al. [32] developed a machine learning approach for the diagnosis of COVID-19 emergency care patients. The scheme predicted the risk of being positively diagnosed. The data were collected from 235 adult patients from a hospital in Brazil, between March 17 and March 30, 2020. The model was trained by support vector machine, neural network, gradient boosting trees, random forests, and logistic regression separately. The support vector machine performed better than other learning algorithms with an AUC of 0.85, the sensitivity of 0.68, specificity of 0.85, and the Brier Score of 0.16. For training the algorithms, a total of 15 variables were used that include gender, age, hemoglobin, red blood cells, platelets, etc. The number of lymphocytes, eosinophils, and leukocytes was the most three important variables for the predictive performance of the algorithm.

The diagnosis of COVID-19 using chest X-ray images by the machine learning method is proposed by Elaziz et al. [33]. The X-ray images are classified into COVID-19 patients and non-COVID-19 patients into separate classes. Fractional Multichannel Exponent Moments (FrMEMs) is used for extracting necessary features from the chest X-ray images. To accelerate the computational process, a parallel multi-core computational framework is used. Then, a differential evolution-based modified Manta-Ray Foraging Optimization is used to select the most important features. Two X-ray datasets of COVID-19 patients are used to evaluate the system. The accuracy achieved from the experiment was 96.09% and 98.09% from the datasets. A classifier model to distinguish COVID-19 from various forms of pneumonia using chest X-ray images is presented by Khuzani et al. [34]. To extract important features from the X-ray images, dimensionality reduction techniques are used. The model obtained 94% accuracy in distinguishing COVID-19 patients from others. Wu et al. [35] presented a machine learning-based COVID-19 infection prediction system. The framework was trained on 11 blood properties from patients. The developed system successfully identified infected patients with similar symptoms, with an accuracy of 96.97% on test data. The scheme also performed better in case of operational data from overseas patients with an accuracy of 91.67%.

At the time of a global pandemic like COVID-19, the healthcare system faces a hard time for diagnosing patients with limited medical equipment and manpower. Machine learning-based diagnosing techniques can help to overcome this crisis a great deal. Especially, medical imaging

techniques are the most suitable alternative for the clinical diagnosis of the patients. Because other techniques require examining various components of the blood that require some time. On the other hand, medical devices like an X-ray machine and CT scanner are very common in almost all Medicare that can provide accurate data very fast for the machine learning model. Thus ML-based approaches provide the fastest diagnosis with reasonable accuracy.

Survey and Screening

The identification of potential areas with highly infected people and the separation of COVID-19 patients from others has become a vital issue in the time of the pandemic.

Srinivasa Rao and Vazquez [36] proposed a smartphone-based survey to group individuals into no risk, minimal risk, moderate risk, and high-risk categories. The developed system is used to collect common symptoms and signs along with basic travel history. While a possible cause is identified, an alert is sent to the nearest health center. The proposed algorithm categorized people into three categories based on the severity of their condition. A study for predicting potential COVID-19 patients using various machine learning techniques such as logistic regression, support vector machine, and multilayer perceptron is proposed by Fayyoubi et al. [37]. The data were collected from the Institutional Review Board at the Hashemite University (HU-IRB: 2020//7/1) and an online survey where the data was the signs and symptoms of the infected patients. Multi-layer perceptron showed 91.62% accuracy which was better compared to other approaches and support vector machine obtained the highest precision 91.67%.

Whitelaw et al. [38] proposed an aggressive contact tracking model that was mainly applied in South Korea and Germany. The machine learning algorithm incorporated facial recognition technology and the Global Positioning System (GPS) to track people's lives. Furthermore, they have observed a smart-watch-based model that collects the pulse, temperature, and sleeping pattern of people as input data and calculates the likelihood of COVID-19 incidences across the nation. These models were enough fruitful in diminishing the effect of coronavirus throughout the nation. Ferretti et al. [39] introduced their mathematical model that used proximity contacts and notified close contacts regarding the positivity of coronavirus infection. They proposed two interventions to isolate individuals with symptoms and quarantining the infected one. In their model, no direct success rate was reported but addressed that the success of these interventions was positively growing with the exponential growth of the epidemic. The actual validation of these models is tough enough since a novel pandemic releases little evidence regarding fighting against it. Nevertheless, authors validated their model by applying them in a time-series, fair

index using and mathematical validation like sensitivity, and specificity.

Since the ground truth is unknown, validating surveying and contact tracking algorithm is a hard task and in many cases, the models may be misleading. But at the time of global pandemic where governments do not know where to search for and which area is likely in the more serious condition, these techniques can provide an initial estimate for identifying potentially affected persons. These techniques are very helpful at the beginning of the pandemic for tracking down COVID-19 and also applying limited resources of a nation in high-risk regions.

Environmental Dependencies

The experts are trying to find out the dependencies between environmental variables with the spread of coronavirus.

Ogundokun and Awotunde [40] introduced a machine learning-based system to find the equivalence between COVID-19 and environmental variables. In the study, the developed platform found that ecological variables such as water and air play a significant role in the outbreak of COVID-19. These variables had a positive equivalence with the occurrence of many cases. Malki et al. [41] proposed a model based on machine learning to indicate that there is a relation between weather conditions and the spread of coronavirus. They had used temperature and humidity as input where the cases were collected from the official case reported by different countries and the weather data were collected from a historical weather database. Finally, the data were supplied to different linear models and ensemble learning-based models which showed a clear indication that the weather has an impact on the spread of coronavirus. Gupta et al. [42] performed a similar kind of study on the relevance of weather and the spread of coronavirus where they had used the daily data collected in India. Support Vector Machine was their main machine learning model through which they had drawn a conclusion on the direct relationship between temperature and spread of coronavirus but confusion on the relation between humidity and outbreak of coronavirus. A study to examine the effect of weather on COVID-19 spread in South Asian countries is presented by Hossain et al. [43]. The dataset collected from the first day of COVID-19 confirmed cases to August 31, 2020, includes weather parameters and confirmed cases. Using the Autoregressive Integrated Moving Average with Explanatory Variables (ARIMAX) model, collected data were analyzed distinctly for each country. The study found that wind speed, air pollutants, rainfall, and temperature have a significant impact on COVID transmission. A similar study [44] performed in the capital city of Norway found that temperature and precipitation levels are significantly correlated with the spread of COVID-19. A different claim was made by Pan

et al. [45]. In their study, they claimed that the weather, humidity, wind speed, and UV radiation had not any significant relation with the reproductive number of COVID-19 patients. The study was conducted on data collected from 202 locations in 8 different countries. Time frequency-based approach was followed to determine the COVID-19 dependencies on the environmental components.

Discussions and Recommendation for Future Work

This section describes a comparative analysis of the machine learning-based developed systems to combat the COVID-19. The major challenges with potential future trends are also recommended here.

Discussion

In this paper, the review has been described in accordance with some features namely forecasting, diagnosis, environmental dependencies, survey, and screening. These are the applications of machine learning techniques that are considered during this COVID-19 epidemic. Recently, developed systems based on machine learning techniques on COVID-19 have a different point of view to be described where some used medical images and some used blood samples, some works are on single country dataset whereas some on multiple country dataset, some used their dataset whereas some used real-time dataset. Among the described systems earlier, X-ray images have been used as samples in [33, 34, 37]. The systems developed in [19, 20, 22–24, 27] used statistical dataset consists of infected, death, and recovered patients. Blood samples have been used for identification and survival prediction in [31, 35]. The real-time dataset has been used in [20, 22, 23] where the schemes considered a frame of days. Since datasets are available for research purpose, the systems developed in [19, 23, 24, 27, 31, 33–35, 37] have collected dataset from open research forums. The same dataset from Jons Hopkins University has been used in [19, 20]. The forecasting models on COVID-19 have been developed in [19, 20, 22–24, 27, 31, 37] in accordance with the peak of the epidemic, amount of infected, death and recovered patients. The diagnosis-based models are proposed to diagnose and identify infected patients accurately developed in [33–35].

Since COVID-19 is a recent pandemic, all the systems on this topic are recently developed. The age period of this disease is not so long. Without a proper dataset, the developed systems may cause inaccurate predictions. The dataset consisting of few samples has less variety of X-ray images. The systems developed on X-ray images [33, 34, 37] may fall down due to the shortage of large datasets. The real-time dataset in [20, 22, 23] has neglected the weather's

correlation with COVID-19 in the south Asian region which may result in inaccurate forecasting. In addition, due to the inconsideration of several factors, i.e., economy, educational condition, medical facilities, and religious beliefs may cause the wrong prediction on the spreading of COVID-19. The predicted models suggested the diminishing of COVID-19 around April–October which has been proven failed. Global pandemic like COVID-19 does not have any trend or seasonality. The scenario of confirmed cases, death cases, and recovered cases are all different in different countries. Hence, the lack of any ground truth of the scenario makes it harder to predict. Furthermore, low availability of initial data and lower accurate data may direct the prediction to another path where actual prediction may be a challenging task. From the review, medical imaging-based models provide inadequate performance due to some factors such as using generalized data, acquisition of sufficient data, and low quality of images. Some statistical data-based models are doubtful since the data are collected from questionnaires. The data collected from questionnaires may have faulty information. They have limited to specific features rather than a substantial attribute. Blood sample-based techniques are limited to consider compact symptoms while the outbreak of COVID-19 may discover newborn symptoms.

In these studies, the works [18–30] used machine learning algorithms for predicting COVID-19 outbreak. All of them predicted the end of COVID-19 pandemic would be in different month of 2020 but all of them had been proved wrong. The studies performed to diagnose COVID-19 patients provided promising results. In [31], the authors proposed a model that predicted death and survival probability with 100% and 90% accuracy, respectively. In the study [33], a machine learning model was applied on two different datasets. The model achieved 96.09% and 98.09% accuracy for first and second dataset, respectively. The study [34] was performed using chest X-ray data that classified normal, pneumonia and COVID-19 patients 94% accurately. A study [35] conducted in China that used data from overseas patients achieved 91.67% accuracy.

Among the applied approaches, mostly used approaches are Linear Regression, Support Vector Machine, Multi-Layer Perceptron, Vector Auto Regression, Decision Tree, Random Forest and XBoost. These approaches mainly focused on real-time-infected patient detection, forecast on recovery rate or recovered patients, death count and cluster creation with infected patients. Most of the authors applied the approaches simultaneously rather than applying a single method. These multi approach triggers the acceptability of the outcome of the research. Since the researches were carried out on initially available datasets, methods like Linear Regression, Logistic Regression, Support Vector Machine and Decision Tree provided better accuracy on prediction in [19, 27, 32, 37]. As the time passed, the datasets started to be

more acceptable. The methods like Multi-Layer Perceptron, Random Forest and XBoost classifier obtained better accuracy on enriched dataset in [28, 31, 32, 34, 35].

Researchers, all over the world are working on developing proper technologies to tackle COVID-19. A lot of models are being proposed but a specific model cannot be certified as a standard model in tackling COVID-19 since these models are developed using different datasets from different distributions. There is also a possibility of overfitting of the models since various models are trying to learn the outlier samples. As a result generalization ability of the model as a whole is hampered resulting in the problem of multiple testing. These problems can only be solved if a gold standard dataset containing a large number of samples with no noise and incorrect labeling for COVID-19 is created.

Recommendation for Future Work

As COVID-19 is the latest epidemic, there is a scope of further development in recently proposed systems against COVID-19. In [19], the claim of the model could be enlarged using deep learning methods on time series data. Weather conditions could be inspected in [20] by applying other optimization algorithms before training the ANN method for better prediction accuracy. Applying the proposed model on a single country dataset does not provide durability. Furthermore, a small number of feature leads to the generation of an imprecise result. Hence, the current dataset with more features and dataset of other countries on COVID-19 might be used in [22–24, 37] for the enlargement of the system. Moreover, increasing the number of features could lead to a better system. A data-driven system designed in [28, 35] could be replaced by a model-driven system to disperse the burden of the proposed system. The accuracy is not the best metric for the imbalanced class label. Currently, the available dataset incorporates less COVID-19 positive cases than negative cases. Therefore in [33], sensitivity, specificity could have mentioned which would make it robust. The single method used in [34] could have been replaced by other machine learning methods namely support vector machine, decision tree, and random forest. Finally, a comparative study could have established among neural network-based classifiers and machine learning-based classifiers which could draw appropriate results. The mortality rate is essential since it specifies the number of patients and required beds in the Intensive Care Unit (ICU). The mortality rate could be modeled in all the systems which could lead to acceleration for the planning of a nation.

Almost all the systems reviewed here provide a guide against COVID-19. However, no single system can provide all the requirements needed to fight against COVID-19. Some model includes forecasting, some diagnosis, and some identification. Thus in the development of systems against

COVID-19 considering previously mentioned issues may lead to new research for providing better aid to the infected patients. Since the initial barrier in the progress of research regarding the novel pandemic is the generation of authentic data, medical centers should have a dedicated section to generate authentic data for the technical research community. Researchers can combine the same types of data from different sources to generate a large gold standard dataset.

Conclusion

This article presents recently developed machine learning technologies that are proposed against the ongoing COVID-19 pandemic. The motive of the developed systems is to provide early detection, accurate forecasting, and proper diagnosis of patients, and measuring dependencies between various environmental variables using different machine learning techniques. The reviewed types of systems have been already developed on different case studies with different datasets (i.e., pneumonia). Now, the COVID-19 is a new epidemic without any authentic medicine which has opened a path of research to technologists all over the world. In this review, the systems that can be incorporated into different frameworks (mobile phones) have been provided. Further development in this field can be achieved by ensuring research on new symptoms, enlarging data, and collecting them from authentic sources. However, the main goal is accomplished when researchers and medical technologists jointly develop a system to aid the infected cases. Hopefully, this review would assist researchers wishing to develop a system to ensure aid against COVID-19.

Funding None.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Coronavirus disease (COVID-19). <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed 06 July 2020.
2. Ullah SMA, Islam M, Mahmud S, et al. Scalable telehealth services to combat novel coronavirus (COVID-19) pandemic. *SN Comput Sci.* 2021;2:18. <https://doi.org/10.1007/s42979-020-00401-x>.
3. Rahaman A, Islam M, Islam M, Sadi M, Nooruddin S. Developing IoT based smart health monitoring systems: a review. *Rev d'Intell Artif.* 2019;33(6):435–40. <https://doi.org/10.18280/ria.330605>.
4. Islam MM, Rahaman A, Islam MR. Development of smart health-care monitoring system in IoT environment. *SN Comput Sci.* 2020;1(3):185. <https://doi.org/10.1007/s42979-020-00195-y>.
5. Islam MM, Karray F, Alhaji R, Zeng J. A review on deep learning techniques for the diagnosis of novel coronavirus (COVID-19). *IEEE Access.* 2021;9:30551–72. <https://doi.org/10.1109/ACCESS.2021.3058537>.
6. Islam M, Mahmud S, Muhammad LJ, et al. Wearable technology to assist the patients infected with novel coronavirus (COVID-19). *SN Comput Sci.* 2020;1:320. <https://doi.org/10.1007/s42979-020-00335-4>.
7. Saha P, Sadi MS, Islam MM. EMCNet: automated COVID-19 diagnosis from X-ray images using convolutional neural network and ensemble of machine learning classifiers. *Inf Med Unlocked.* 2021;22:100505.
8. Guan W, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med.* 2020;382(18):1708–20. <https://doi.org/10.1056/NEJMoa2002032>.
9. Jiang F, Deng L, Zhang L, Cai Y, Cheung CW, Xia Z. Review of the clinical characteristics of coronavirus disease 2019 (COVID-19). *J Gen Intern Med.* 2020;35(5):1545–9. <https://doi.org/10.1007/s11606-020-05762-w>.
10. Asraf A, Islam MZ, Haque MR, et al. Deep learning applications to combat novel coronavirus (COVID-19) pandemic. *SN Comput Sci.* 2020;1:363. <https://doi.org/10.1007/s42979-020-00383-w>.
11. Rahman MM, Manik MMH, Islam MM, Mahmud S, Kim J-H. An automated system to limit COVID-19 using facial mask detection in smart city network. In: 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2020, pp. 1–5. <https://doi.org/10.1109/IEMTRONICS51293.2020.9216386>.
12. Ayon SI, Islam MM, Hossain MR. Coronary artery heart disease prediction: a comparative study of computational intelligence techniques. *IETE J Res* 2020:1–20. <https://doi.org/10.1080/03772063.2020.1713916>.
13. Islam Ayon S, Milon Islam M. Diabetes prediction: a deep learning approach. *Int J Inf Eng Electron Bus.* 2019;11(2):21–7. <https://doi.org/10.5815/ijieeb.2019.02.03>.
14. Haque MR, Islam MM, Iqbal H, Reza MS, Hasan MK. Performance evaluation of random forests and artificial neural networks for the classification of liver disorder. In: 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), 2018, p. 1–5. <https://doi.org/10.1109/IC4ME2.2018.8465658>.
15. Islam MM, Iqbal H, Haque MR, Hasan MK. Prediction of breast cancer using support vector machine and K-Nearest neighbors. In: 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), 2017, vol. 2018, January, p. 226–229. <https://doi.org/10.1109/R10-HTC.2017.8288944>.
16. Hasan MK, Islam MM, Hashem MMA. Mathematical model development to detect breast cancer using multigene genetic programming. In 2016 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016, p. 574–579. <https://doi.org/10.1109/ICIEV.2016.7760068>.
17. Rahman MM, Nooruddin S, Hasan KMA, et al. HOG + CNN Net: Diagnosing COVID-19 and pneumonia by deep neural network from chest X-Ray images. *SN Comput Sci.* 2021;2:371. <https://doi.org/10.1007/s42979-021-00762-x>.
18. Liu D, et al. A machine learning methodology for real-time forecasting of the 2019–2020 COVID-19 outbreak using Internet searches, news alerts, and estimates from mechanistic models, 2020 [Online]. Available from: <http://arxiv.org/abs/2004.04019>.
19. Sujath R, Chatterjee JM, Hassanien AE. A machine learning forecasting model for COVID-19 pandemic in India. *Stoch Environ Res Risk Assess.* 2020;34(7):959–72. <https://doi.org/10.1007/s00477-020-01827-8>.
20. Hasan N. A methodological approach for predicting COVID-19 epidemic using EEMD-ANN hybrid model. *Internet Things.* 2020;11:100228. <https://doi.org/10.1016/j.iot.2020.100228>.

21. Punn NS, Sonbhadra SK, Agarwal S. COVID-19 epidemic analysis using machine learning and deep learning algorithms. medRxiv 2020:2020.04.08.20057679. <https://doi.org/10.1101/2020.04.08.20057679>.
22. Tiwari S, Kumar S, Guleria K. Outbreak trends of corona virus (COVID-19) in India: a prediction. Disaster Med Public Health Prep. 2020:1–6. <https://doi.org/10.1017/dmp.2020.115>.
23. Li M, et al. Predicting the epidemic trend of COVID-19 in China and across the world using the machine learning approach. medRxiv 2020. <https://doi.org/10.1101/2020.03.18.20038117>.
24. Ardabili SF, et al. COVID-19 outbreak prediction with machine learning. SSRN Electron J. 2020. <https://doi.org/10.2139/ssrn.3580188>.
25. Ndiaye BM, Tendeng L, Seck D. Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting. 2020. [Online]. Available from: <http://arxiv.org/abs/2004.01574>.
26. Tuli S, Tuli S, Tuli R, Gill SS. Predicting the growth and trend of COVID-19 pandemic using machine learning and cloud computing. Internet Things. 2020;11: 100222. <https://doi.org/10.1016/j.iot.2020.100222>.
27. Muhammad LJ, Islam MM, Usman SS, Ayon SI. Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. SN Comput Sci. 2020;1:206. <https://doi.org/10.1007/s42979-020-00216-w>.
28. Yan L, et al. A machine learning-based model for survival prediction in patients with severe COVID-19 infection. medRxiv 2020. <https://doi.org/10.1101/2020.02.27.20028027>.
29. "COVID-19 Mathematical Modeling | COVID-19 | CDC." <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/mathematical-modeling.html>. Accessed 29 Oct. 2020.
30. Ray EL, et al. Ensemble forecasts of coronavirus disease 2019 (COVID-19) in the U.S. <https://doi.org/10.1101/2020.08.19.20177493>.
31. Yan L, et al. Prediction of criticality in patients with severe covid-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan. medRxiv 2020. <https://doi.org/10.1101/2020.02.27.20028027>.
32. Batista AFM, Miraglia JL, Donato THR, Filho ADPC. COVID-19 diagnosis prediction in emergency care patients: a machine learning approach. medRxiv 2020. <https://doi.org/10.1101/2020.04.04.20052092>.
33. Elaziz MA, Id KMH, Salah A, Darwish MM, Lu S, Sahlol AT. New machine learning method for image-based diagnosis of COVID-19. 2020. <https://doi.org/10.1371/journal.pone.0235187>.
34. Khuzani AZ, Heidari M, Shariati A. COVID-classifier: an efficient machine learning model to assist in the diagnosis of COVID-19 infection in chest x-ray images. medRxiv 2020. <https://doi.org/10.1101/2020.05.09.20096560>.
35. Wu J, et al. Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results. medRxiv 2020. <https://doi.org/10.1101/2020.04.02.20051136>.
36. Srinivasa Rao ASR, Vazquez JA. Identification of COVID-19 can be quicker through artificial intelligence framework using a mobile phone-based survey when cities and towns are under quarantine. Infect Control Hosp Epidemiol. 2020;41(7):826–30. <https://doi.org/10.1017/ice.2020.61>.
37. Fayyoumi E, Idwan S, Aboshindi H. Machine learning and statistical modelling for prediction of novel COVID-19 patients case study: Jordan. Int J Adv Comput Sci Appl. 2020;11(5):122–6. <https://doi.org/10.14569/IJACSA.2020.0110518>.
38. Whitelaw S, Mamas MA, Topol E, Van Spall HGC. Applications of digital technology in COVID-19 pandemic planning and response. Lancet Digit Heal. 2020;2(8):e435–40. [https://doi.org/10.1016/S2589-7500\(20\)30142-4](https://doi.org/10.1016/S2589-7500(20)30142-4).
39. Ferretti L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020;368(6491):eabb6936. <https://doi.org/10.1126/science.abb6936>.
40. Ogundokun RO, Awotunde JB. Machine learning prediction for Covid 19 pandemic in India. medRxiv 2020. <https://doi.org/10.1101/2020.05.20.20107847>.
41. Malki Z, Atlam ES, Hassanien AE, Dagnew G, Elhosseini MA, Gad I. Association between weather data and COVID-19 pandemic predicting mortality rate: machine learning approaches. Chaos Solitons Fract. 2020;138: 110137. <https://doi.org/10.1016/j.chaos.2020.110137>.
42. Gupta A, Pradhan B, Maulud KNA. Estimating the impact of daily weather on the temporal pattern of COVID-19 outbreak in India. Earth Syst Environ. 2020;4(3):523–34. <https://doi.org/10.1007/s41748-020-00179-1>.
43. Hossain MS, Ahmed S, Uddin MJ. Impact of weather on COVID-19 transmission in south Asian countries: an application of the ARIMAX model. Sci Total Environ 2020: 143315. <https://doi.org/10.1016/j.scitotenv.2020.143315>.
44. Menebo MM. Temperature and precipitation associate with Covid-19 new daily cases: a correlation study between weather and Covid-19 pandemic in Oslo, Norway. Sci Total Environ. 2020;737: 139659. <https://doi.org/10.1016/j.scitotenv.2020.139659>.
45. Pan J, et al. Warmer weather unlikely to reduce the COVID-19 transmission: an ecological study in 202 locations in 8 countries. Sci Total Environ. 2021;753: 142272. <https://doi.org/10.1016/j.scitotenv.2020.142272>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.