Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 210–219

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Data Mining Based Techniques for Covid-19 Predictions

Rahul Rane[a], Aditya Dubey[b,*], Akhtar Rasool[c], Rajesh Wadhvani[d]

[a,b,c,d]Department of Computer Science and Engineering, Maulana Azad National Institute of Technology, Bhopal and 462003, India

Abstract

COVID-19 is a pandemic that has resulted in numerous fatalities and infections in recent years, with a rising tendency in both the number of infections and deaths and the pace of recovery. Accurate forecasting models are important for making accurate forecasts and taking relevant actions. As a result, accurate short-term forecasting of the number of new cases that are contaminated and recovered is essential for making the best use of the resources at hand and stopping or delaying the spread of such illnesses. This paper shows the various techniques for forecasting the covid-19 cases. This paper classifies the various models according to their category and shows the merits and demerits of various fore-casting techniques. The research provides insight into potential issues that may arise during the forecasting of covid-19 instances for predicting the positive, negative, and death cases in this pandemic. In this paper, numerous forecasting techniques and their categories have been studied. The goal of this work is to aggregate the findings of several forecasting techniques to aid in the fight against the pandemic.

Keywords: Stochastic Forecasting Models; Supervised ML Models; Soft Computing-based Models; Deep Learning Models

1. Introduction

In the past, different diseases have afflicted humanity over the planet [1]. The World Health Organization (WHO) and several national governments are working to com-bat the epidemic worldwide. The coronavirus COVID-19 outbreak, which was initially discovered in December 2019 in Wuhan, China, is a major Fear of modern civilization. COVID-19 is present in 213 nations and territories around the world, according to the World Health Organization. The COVID-19 virus can be transferred by direct contact with a contaminated person's body or through the air they breathe. Covid-19's intake period of at least fourteen days is critical to its promotion. For managing outbreaks and establishing effective strategies to prevent the spread of COVID-19, reliable prediction of recovered and infected

* Corresponding author. Tel.: +91-750-985-2725.
E-mail address: dubeyaditya65@gmail.com

1877-0509 © 2023 The Authors. Published by Elsevier B.V.

10.1016/j.procs.2023.01.003

COVID-19 cases is important. A key issue with this virus is that it has a 14-day isolation period during which it can spread the illness without causing any symptoms [2].

Soft computing models present a multilayer perceptron-based prediction model, a popular, supervised, and feed-forward neural network model [3], [4]. Stochastic forecasting models Susceptible, Exposed, Infection and Recover (SEIR), Susceptible, Infection and Recover (SIR), Model-based Least Square Support Vector Machine (LSSVM) growth models because of the condition of exposure to epidemic growth or extrapolation models, frequently used by various biological and social processes [5]. As is often the case with machine learning methods, time-series data is categorized into two categories such as training data and test data [6]. Section II and section III includes the details of forecasting models based on stochastic forecasting models and supervised ML models, Section IV describes the soft computing-based model, and section V describes the Deep learning models; with the help of these models, can easily predict the Covid-19 cases. Fig.1 shows the classification of various forecasting models based on their category the contribution of this paper includes-

- This paper aims to examine existing forecasting models, categorize forecasting models based on the model type, investigate symptomatic and asymptomatic factors, identify forecasting model problems, and analyze pandemic control suggestions. Several studies have used various linear and nonlinear models to forecast the epidemic's progress. These models use time-series data to make short- and long-term forecasts about a disease outbreak. Because each model is better suited to a certain problem, each computational forecasting model has its characteristics.
- This paper describes the forecasting techniques using merits and demerits.
- The research emphasis on potential problems that can occur while anticipating COVID-19 cases in order to determine the likelihood of positive, negative, and fatal cases in this pandemic.

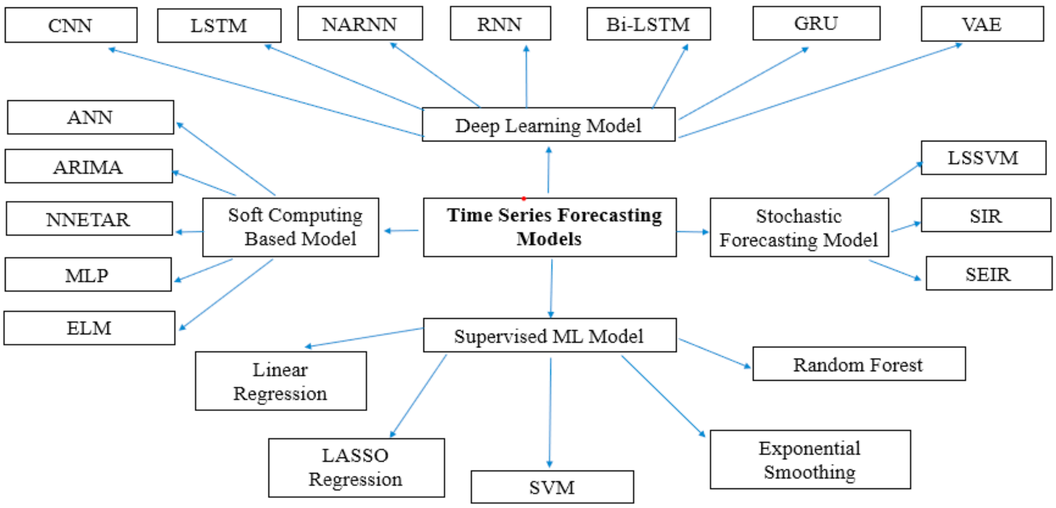

Fig. 1. Classification of various forecasting model.

2. Challenges for forecasting techniques

The first challenge is tracking an infected person and the people they come in contact with is undoubtedly a challenging task [7]. The second challenge is additional concern is that patients cannot be identified in advance since covid-19 positive individuals must be isolated for 14 days. Patients can infect anyone who comes into contact with them within the stipulated isolation period. The third challenge is that data is sometimes available in an unstructured manner. As a result, maintaining the data’s quality and quantity before entering the training stage is critical. Data accuracy is an important factor in achieving effective forecasting methods. The next challenge is that if data overfitting occurs, the model may not perform well on new data. The fifth challenge is that it is important to have clean data for analysis, but cleaning the data sometimes loses its integrity. The sixth difficulty is that data is plentiful, yet giving all of it to the model will not enhance accuracy. The seventh challenge is that if the incorrect algorithm is used, the outcome can be misleading. The same is true if the wrong attribute is chosen.

3. Stochastic Forecasting Models

In previous pandemics, stochastic models were used to predict human death and the total number of deaths during a specific time or at the end of the pandemic [8]. Compared to other forecasting methods, the stochastic approach is an experimental method that produces more accurate results. Models of this class were used for fore-casting the growth rate of patients or the number of reduced patients. The SIR Model is an example of a stochastic method. Data is collected from an official institution that gives the data for the study. The research is based on information obtained from legitimate sources such as search engines, phone records, and newspaper articles. This study used the SIR model from the compartmental models. This traditional method is very successful and gives a better prediction result. The result showed that the covid-19 disease was at its peak or before the symptom onset. For better evaluation of the disease, Stochastic Methods are used.

3.1. LSSVM

Singh et al. proposed an approach for covid -19 prediction i.e., LSSVM [9]. LSSVM has been developed from SVM, this serves as a rough approximation of the exact non-linear relationship between the input and output variables. LSSVM model aids in recognizing the essential characteristics of the propagating virus. Considering an input data vector K = $\{ d_e \} \in U^d$ and output data vector M = $\{ f_e \} \in$ U that are concatenated to produce the training dataset {(d_1 ,f_1), (d_2 ,f_2), …., ((d_w ,f_w) } with $f_e \in \{ -1 , +1 \}$ as class labels and M (K) = sign [u^TK+ y] as the linear classifier, where W is the number of samples, y ∈ bias and u ∈ weight vector. It is possible to infer that the training dataset can be divided into two classes. In the initial stage it takes the input data from covid 19 data set. In next stage Format the data in to LSSVM format. After that Data under goes LSSVM training process. In the next stage Cross validate this data set and trained the model using validated data set. After this stage, Model undergoes LSSVM forecasting process and in last stage Final forecasting of data is done using the trained model.

3.2 SIR

Malavika et al. proposed a SIR Model for covid -19 that is a compartmental model. The Susceptible population is classified into parts based on their infectious condition and the population size over time [5]. The population is divided into three categories in the model above, which are Susceptible (S), Infectious (I), and Recovered (R). Those who come in contact with contagious people are Susceptible. Infected persons are those groups of people affected by the disease and can infect susceptible groups with it. Recovered persons are the group of recovered from the disease and are not in danger of the same disease anymore. The SIR model principle describes the procedure that how the virus spreads. The SIR model is a compartmental representation and acts as a base for some other models while expanding, such as the SEIR model. Input provided to the model is the proportion of the susceptible, infected, and recovered population, and all are allowed to t (0) where t=Time.

3.3 SEIR

The SEIR model expands the SIR model [10], which has an extra parameter E, which shows the fraction of asymptomatic infected people (individuals not having general symptoms). As per Gois et al., Since the SIS and SIR model allows the cases to be treated without an isolation time, that is not the possibility for many ranges of communicable infections. A model was suggested named the SEIR model, where after a specific period, the susceptible people can get an infection. COVID-19, for example, has a 14-day isolation period. It is beginning with the addition of the new equation, which depicts the number of people exposed to the virus. The isolation rate γ, which is the rate at which hidden or susceptible individuals are becoming infectious (typical time of isolation 1/γ), is included in the model. Where β is the average number of people that came in contact, C number of Susceptible cases, K number of infectious instances, P number of exposed instances, and H number of death instances.

4. Supervised ML Models

Over the past years, ML has become an important research field [11]. It solves various complex problems in real-time. In supervised learning, a model is trained with the labeled dataset by the learning algorithms. A trained model is provided with input, and based on the training; it predicts the result.ML algorithms have many applications like disease prediction, Prediction of missing values in health care datasets, temperature prediction, etc. Linear regression is one of the important mathematical analyses of machine learning. In this COVID-19 prediction study, In this paper, five regression models are Support Vector Machine (SVM), Random Forest (RF), Exponential smoothing (ETS), Linear Regression, and LASSO Regression. Some of the important ML areas are expected. These methods can better understand the virus's characteristics and predict future pandemic threats. The importance of ML in resolving the COVID-19 pandemic crisis is discussed in this study.

4.1 SVM

Alghazzawi et al. proposed an approach for covid -19 prediction analysis, i.e., SVM [12]. SVM is an ML algorithm that is applied a repeated number of times for predictive analysis. The data set utilized in the application of the SVM model is generally labeled. It creates a sequence of feature details and labeled input-output mapping functions. In SVM, input data is split into three classes, positive cases, negative cases, and deaths. The model is trained using the data to forecast the threat of a covid-19 pandemic. Application of SVM is not possible directly to the multiclass problem because it is a binary classification tool. To split multiclass instances, they used binary clusters grouping by class. The construction of three binary SVMs (three represents the counting of classes) was carried out in this research, with each binary classifier being one level different from the rest of the classes. An SVM model is simple to implement because it is one of the high-quality machine learning methods [13]. SVM uses a method of increasing accuracy by reducing the boundary distance between the hyperplane and training data instead of conventional methods of minimizing observational errors in testing.

4.2 RF

RF works on regression and classification and is a well-known unsupervised learning method [14]. RF is an ensemble learning method. The classifiers display a decision tree, M output by M decision tree acquired by this approach. It uses a voting technique for decisions. RF is a simple model for using parallel processing. Yesilkanat proposed that it is a machine learning algorithm with multiple decision trees. RF is the collection of bagging and random subspaces methods. In RF, the data set is initially arbitrarily split into two parts: training data for learning and validation data for testing the learning level. After that many decision trees are created randomly from that data set. The branch of every tree is identified by randomly chosen prediction at the node point. The final approximate result is the mean of all the results that are produced from the trees. In addition, the RF algorithm manages the overfitting level as training is done at various randomly chosen sub-datasets by boot-strap sampling. Data is split into processes; the RF algorithm is applied to the training data set in the next phase. After that, the model is formed, and it is tested with the testing sub-data set, and the results are shown; in the final stage, the result can be classified with the use of step 3.

4.3 ETS

Shoaib et al. proposed a technique called ETS [15]. ETS is a prediction technique for forecasting a time series for one variable. It is a well-organized prediction method it can be used as an alternative to the most common technique Box–Jenkins. To build smoothing time series, ETS is a powerful format. ETS are of three different types simple, Holt's linear method, and Holt-Winters. The simple ETS forecasting method is used for forecasting data with no clear trend. Time series are in a complex form then; to study them, they needed to develop a model. It is seen that ETS is relatively faster than the ARIMA method. The method is clear for seasonality complex time series. In this study, the advantage of ETS is versatile and extensible; this paper extends the survey to apply the ARIME model to ETS. The study of intervention with ETS does not depends on stationary, so it is simple to identify which opponent is most suitable for a given time series by using detailed criteria for the selection of model.

4.4 Linear Regression

Another one of the well-known prediction methods is linear regression [16]. It finds a better variable set for prediction and then the exact variable from the set for the outcome of the prediction. It depends on the sign and beta estimation; this regression explains the relationship between dependent variables and independent variables. The linear regression equation is shown below, where X is the dependent variable and e is an independent variable, C_0 is intercept and C_1, C_2 are coefficient and n show the number of observations, Linear regression methods are easy to solve prediction problems, if it is having one input variable then it is called simple linear regression, if it is having more than one variable then it is known as multi regression. Linear regression depends on two values, one is dependent and the other is independent. Linear regression has the relation between the independent and dependent variables. The difference between the predicted value and actual value must be minimum and to check the minimization of the problem following equation is used

$$X=C_0+C_1e_1+C_2e_2+\dots\dots\dots+C_ne_n \quad (1)$$

4.5 LASSO Regression

Rustam et al. proposed the ML technique, the LASSO Regression model [17]. LASSO is a type of linear regression model which uses shrinkage. In LASSO, the error magnitude is decreased by using the process of shrinking. The LASSO regression model is a suitable approach for multi-linear regression. Hence, the LASSO regression model is smoother in the amount of function it uses. Since the LASSO regression model does not represent the importance of zero when the new function might not enhance the penalty terms related to that function, it only makes one attempt at a time. This LASSO Regression minimizes the coefficient, which is called the square residual. LASSO gives better results in approximating and confirming the death rate. As a result, LASSO reduces the number of characteristics used in the regression. It utilizes a regularization method to penalize the excess features automatically. That is, characteristics that are not helpful to the regression findings can be assigned to a very low value, possibly zero.

5. Soft Computing-based Models

Soft computing-based models are based on neural networks and fuzzy logic, and these models have been used for forecasting Covid-19. Soft computing solves the problem based on the human mind approach, not the computer approach. With the help of soft computing, those can be solved manually. This section briefly describes the basic principle of five soft computing-based models that are using for COVID-19 time-series forecasting, namely Artificial Neural Networks (ANN), Autoregressive Integrated Moving Average (ARIMA), Multilayer perceptron (MLP), Extreme learning machines (ELM) and Neural Network Auto Regression Model (NNETAR). In this section, this paper describes the MLP and ANN model for predicting the number of positive cases; using ARIMA models can predict the daily number of positive, negative, and recovered cases. NNETAR is the two-phase model. NNETAR and ELM these models are used for predicting the number of positive and death cases individually.

5.1 ANN

Shoaib et al. proposed the usage of ANN for Covid-19 cases [15]. ANN is influenced by the human brain and supports many non-linear structures. It gives us a more accurate result due to the simultaneous processing of data across the neural network and the connection weights. There are three levels in the ANN structure. Input neurons are carried by the first layer, which is connected to the hidden layer, and the hidden layer is connected to the output layer. In a general model of ANN, the data are divided into three groups for validation, testing, and training purposes. The structure of ANN shows the information flow in an onward direction from the input to the output layer and the backflow of modeling error from the output layer to the input layer. While training a network, the value of data is used. The next step after the training operation is the attestation step; the trained ANN prediction is judged in testing. When the sample size is small, ANN meets at local minima, resulting in poor generalization. This issue is resolved by splitting the data into three samples 70% training,15% validation, and 15% testing, with the attestation process as the generalization method to test the built network.

5.2 ARIMA

Sahai et al. introduced the ARIMA model, which is a combination of both moving average (MA) and autoregressive (AR) models for covid -19 prediction analysis [18]. ARMA model has been utilized for univariate time series modeling. In an Auto-Regressive model, the variable's future value is considered to be determined by a linear combination of a random error term and past observations. Only a univariate stationary time series was used for AR and MA. When the series is not found to be stationary, it must be differentiated through time to make it so. It demonstrates a systematic way of developing future forecasting models for moving and non-moving time series based on observed values with equal time intervals. ARIMA forecasting relies on the autoregressive characteristic of time series and necessary incremental corrective adjustments.

5.3 MLP

Talkhi et al. proposed a technique, MLP, a perceptron model type [2]. In MLP, there are at least three layers. MLP architecture consists of biases, weights, inputs, and an activation function that gives the output. The input given to the neuron is multiplied by an adaptive coefficient known as weight and with a function such as sigmoid, hyperbolic tangent, etc., which are nonlinear activation functions. It can be declared from a statistical point of view that nonlinear regression runs on MLPs. In an MLP network, the output from the neuron is a number of input i, b is the bias, and weight is associated with each. Depending on the error function calculation, the network weight is adjusted in the training phase. In further steps, weight updation mostly depends on an error in every step and learning rate. For the last step, each step is to be repeated up to it comes the number of epochs.

5.4 ELM

Shetty and Pai proposed a technique is ELM [4]. This method is simple and non-iterative. one iteration of learning helps to reduce the learning time considerably and in the emerging area finds a major application area such as pattern recognition, image analysis, bioinformatics, and other prediction and forecasting purposes. The ELM algorithm finds the arbitrary distinct training data, input, target vector, and hidden nodes; if an activation function is used, then the model's output is found. This model overcomes the drawback of a traditional algorithm. As compared to traditional learning algorithms, it shows the smallest training error. The input weight and biases of hidden layers ELM are randomly determined while training of only the output layer is done. In the other step, the activation function of the network is selected. For a unique solution, least squares are used, and this solution has a minimum error, and the values of the weight matrix are calculated by this method.

5.5 NNETAR

Talkhi et al. proposed a method called NNETAR, a neural network model, and for forecasting issues, a parametric method that is not linear is used [2]. The forecasting of this model is done in 2 parts. The arrangement of the AR model is in the initial part for requires time series; in the next part, it trains the neural network by using the trained data set. The sequence of autoregressive decides the input nodes of a neural network. The fitted model with a no-sessional pattern comprises two components, S and R; S is the count of inputs, and R is the count of neurons hidden so that the resulting model can be represented in NNAR (S, R) form. It is similar to ARIMA with a nonlinear function. In NNETAR models, the input variables might be gone up and acquired model by the hidden node and first input lag.it is a statistical model which is used in forecasting; it plays an important part in determining the future tendency of the disease.

6. Deep Learning Models

In this phase, deep learning models show the rates of progress in various applications [19]. This study uses information from time-series data to forecast the best time-series data acquired by COVID-19, including Convolutional neural network (CNN), Long Short-Term Memory (LSTM), recurrent neural network (RNN), Nonlinear Autoregression Neural Network (NARNN), Gated Recurrent Unit (GRU), Variational autoencoders (VAE), Bi-directional LSTM(Bi-LSTM). Deep learning models show that language training and image processing are both effective. Verma et al. suggested a method for monitoring and diagnosing healthcare using deep learning [20]. The suggested architecture FETCH provides a very helpful framework for actual health care services, like heart disease and more, by integrating with edge smart devices to work on deep learning technologies and automated monitoring.

6.1 CNN

CNN is a technique suggested by Nabi et al. for the covid-19 prediction strategy [21]. Every convolutional layer and one neuron are only incorporated into neurons in the top layer in a small rectangle. The network can focus on lower-level elements in the hidden layer, then collect them into higher-level characteristics in the next hidden layer, and so on. 4 convolutional layers are integrated to discover complicated features and patterns in a time series using the CNN model in this prediction. One flat layer and two thick layers are employed in the final stage for the required outcome—the ability of CNN to extract features from large amounts of data. Indeed, depending on the data format, CNN model training incorporates features taken from input such as pictures, 2Dimension signal, or 1Dimension sensor data. CNN is a hierarchical feature extractor in which the first layers extract low-level characteristics such as edges or lines, while the subsequent levels extract more complicated shapes, textures, or object pieces [22].

6.2 RNN

Zeroual proposed RNNs to predict Covid-19 (RNN) [19]. Regular feed-forward neural networks have been effectively applied in many different fields. In such networks, data is moved in one way via a hidden layer, and the output is only affected by the current condition. RNNs have been created to control time-dependent learning difficulties to overcome this constraint. RNNs are designed to take into account the influence of previous data when generating results. To do this, cells representing gates that influence the outcome utilizing past observations are included in the result generation. LSTM and GRU are two powerful RNN models which are cost-effective for time-dependent time series data. Deep learning models outperform other traditional time series models in modeling and forecasting, and traditional networks have shown that they can beat other time series models in many application domains.

6.3 LSTM

LSTM is an alternative to traditional RNNs that effectively address long-term dependency [21]. Each LSTM block in the network module comprises of a predetermined set of vectorized operations between the old and new input, as well as the appropriate application of mathematical functions like sigmoid and tanh. The LSTM block, a memory cell, and three operating gates, each with its own weights and bias vectors, are all used to serially transmit data. A nonlinear sigmoid function is utilized to process a mix of current block input and old activation values. To begin, a nonlinear sigmoid function is used to choose which values to remember and which to discard, utilizing a mix of current block input and prior activation values. This layer is called the forget gate. The in-put/Update Gate is then utilized to find the identical activation combination from the previous layer and the current block input. The updated input A tanh (hyperbolic tangent) function processes the main input group, generating new unique values for the current block.

6.4 GRU

Nabi et al. proposed a technique for predicting covid-19 that is GRU; it might be viewed as a variation of LSTM and has several similarities to LSTM [21]. GRU helps the network understand long-term dependencies by resolving the "Vanishing Gradient" problem that is present in conventional RNNs. The sigmoid and tanh functions are also employed by GRU blocks to calculate required data, but this kind of block does not have a separate Forget gate; instead, the Update gate controls the data transmission. These two fundamental changes result in LSTM having lesser parameters and a simpler design, making it more economical and straightforward for training. Inside, the sigmoid and tanh functions are utilized to calculate the required data. GRU blocks feature a Reset gate in addition to the Update gate.

6.5 NARNN

Kırbas et al. proposed a Deep learning technique for predicting covid-19 that is NARNN is a popular method for predicting time series [22]. This artificial neural net-work uses a portion of the time series. Training data and multiplier weights are obtained in the artificial neural network—ten neurons in the NARNN 2-delay model. This neural network forecasts the future value by looking at two past data sets. The values predicted by the neural network are matched with previously known findings, and the difference is examined to measure the model's performance. The difference value should be as near to 0 as possible for best results. One input layer, one or more hidden layers, and one output layer make up the NARNN design [23]. The NARNN is a feedback-connected dynamic and recurrent network. The NARNN may anticipate time series one step ahead or multistep ahead.

6.6 Bi-LSTM

Zeroual et al. suggested a bi-directional approach for covid-19 prediction [19]. The Bi-LSTM algorithm is a more advanced variation of the LSTM procedure. As an earlier state, the current state of the LSTM can only be recreated using the backward context. However, the forward context is linked to the actual state; the LSTM model does not consider this. Only when the time-series data is on a scale of a specified range can Bi-LSTM models operate successfully [24]. The Bi-LSTM method has been utilized to overcome this limitation and improve state reconstruction accuracy by joining the favorable Characteristics of the bidirectional RNN with those of the LSTM [19]. This was done by combining two hidden states that allow the data to flow from the backward to the forward layers.

Table 1. Comparison table of various forecasting techniques.

Sr. No.	Techniques	Advantage	Disadvantage
1.	ARIMA [9,25]	For stationary time series, this model is utilized. It is used for future predictions of a fixed time series.	Recognition of model is difficult. Not suitable for long term prediction.
2.	LSTM [22,26]	Each weight adjustment's complexity is reduced to O (1).	Training is done for a long duration of time. It is possible to stack a few LSTM layers, which leads to more overfitting.
3.	MLP [2,27]	For long term prediction, it gives good result. MLP can distinguish data that cannot be separated linearly.	For representing complex data, Networks should learn.
4.	ANN [8,28]	Based on the output variable, its categories as supervised or unsupervised learning. Could access various training algorithms.	Nature of being a black box, overtraining. Depend on weight value initialization.
5.	SEIR [29]	An iterative model that uses a change in relative average recently.	It declares the output only, not describing the process.
6.	BI-LSTM [19,30]	PM2.5 concentration prediction is used in Bi-LSTMs. The situations that requiring context input the Bi-LSTM is useful.	It needs long data to pull out relevancy in time-series data
7.	NNETAR [2]	It uses in machine learning problems because it is a statistical model in a neural network.	It will take more time because it is a two-phase forecasting process.
8.	ELM [2]	ELM has high speed in learning because it gives better generalization performance and lesser the training time.	non-iterative method because it has one pass learning.
9.	CNN [21]	Due to its feature learning capability that made CNN the best forecasting model.	CNN design causes over-fitting in the model.
10.	SIR [8,10]	SIR modeling is done local(county) level.	SIR model was not applied where the data size is not large.
11.	NARNN [22]	This model makes a non-linear regression by the neural network.	The NARNN model made optimistic estimates that were lower than the actual values.
12.	RNN [3,19]	The problems that are time-dependent learning are handled. temporal information learning is efficient in RNN.	RNN is difficult to train because of the gradient vanishing/exploding problem.

13.	LSSVM [9]	It reduces the big task of solving complex quadratic programs to a set of the equation that is linear.	LSSVM is applied to the signal processing, pattern recognition, and non-linear regression only.
14.	GRU [3,21]	The GRU model uses fewer training parameters and different gating units.	GRU block separate memory cell is absent.
15.	VAE [19]	It goes pass the Overfitting problem. The loss function is minimized due to the key insight behind VAE.	Loss function has a large cost because of inadequate reconstruction.
16.	LINEAR Regression [16]	For tackling Forecasting problems, linear regression models are more approachable and practical.	In linear regression, only a single input variable is there.
17.	LASSO Regression [17]	It minimizes the extreme data sample into central value by the use of the shrinking process.	For small dataset, LASSO Regression results in poor forecasting.
18.	SVM [8]	It can define a convex optimization problem and ignore overfitting. generalization error can be minimized from the training error.	It is difficult to choose kernel as well as speed and size of training and testing sets. transparency of results is less.
19.	ETS [11]	By looking at the large-scale dataset, ETS works best in the considering forecast domain.	The model does not based on statistical distribution because it does not include model selection criteria.
20.	RF [14]	This method gives good results in both regression and classification problems.	Particular tree cause RF estimation at certain weights. Each tree is not checked individually.

7. Conclusion and future work

For maximum technique execution, evaluations of correct parameters and operational framework for the optimum implementation of the various techniques must be carried out. Certain advancements can exploit the full functional capabilities of the previously proposed algorithms. While the forecasting methods discussed in this research paper give good results, better techniques for reducing time complexity are still required. The study explores forecasting methods from the traditional approach to the current one. This review study can keep researchers up to date on the various techniques of forecasting used by other researchers. This study will provide an overview of each approach to determine the benefits and drawbacks of future studies in this area. This review has gone through many forecasting approaches based on various criteria, but all have the same goal of decreasing complexity and conveying data more understandably. Several new studies have examined the efficacy of the existing forecasting algorithm. However, because various factors can affect forecasting efficiency, there is no explicit consent for every dataset and suitable approach. Many of these strategies are reviewed in depth in this study. The correctness of the forecasting result is a critical stage in estimating any algorithm's performance. This study aims to assist readers in learning about recent advances in this subject and to stimulate the development of new algorithms.

References

[1] Aditya Dubey, and Akhtar Rasool (2022) "Usage of Clustering and Weighted Nearest Neighbors for Efficient Missing Data Imputation of Microarray Gene Expression Dataset." *Advanced Theory and Simulations* **220460**: 1-12.

[2] Talkhi, Nasrin, Fatemi, Narges Akhavan, Ataei, Zahra, Nooghabi, and Mehdi Jabbari. (2021) "Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods." *Biomed Signal Process Control*: 1-8.

[3] Aditya Dubey, and Akhtar Rasool (2021) "Efficient technique of microarray missing data imputation using clustering and weighted nearest neighbour." *Scientific Reports* **11**: 1-12.

[4] Rashmi P. Shetty, and P. Srinivasa Pai. (2021) "Forecasting of COVID 19 Cases in Karnataka State using Artificial Neural Network (ANN)." *Journal of The Institution of Engineers (India): Series B*: 1201-1211.

[5] B Malavika, S Marimuthu, Melvin Joy, Ambily Nadaraj, Edwin Sam Asirvatham, and L Jeyaseelan. (2021). "Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models." *Clin Epidemiol Glob Health:* 26-33.

[6] Aditya Dubey, and Akhtar Rasool (2022) "Recent advances and applications of deep learning technique." *Journal of Theoretical and Applied Information Technology* **100 (13)**: 5000-5009.

[7] Parth Vyas, Falak Sharma, Akhtar Rasool, and Aditya Dubey, "Supervised Multimodal Emotion Analysis of Violence on Doctors Tweets." *International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC)*: 962-967.

[8] Arne Stoll. (2021) “A Survey of Forecasting Techniques for COVID-19.”: 1-17.

[9] Sarbjit Singh, Kulwinder Singh Parmar, Sidhu Jitendra Singh Makkhan, Jatinder Kaur, Shruti Peshoria, and Jatinder Kumar. (2020) “Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries.” *Chaos, Solitons & Fractals*: 1-9.

[10] Gois, Francisco Nauber Bernardo, Lima, Alex, Santos, Khennedy, Oliveira, Ramses, Santiago, Valdir, Melo, Saulo, Costa, Rafael, Oliveira, Marcelo, Henrique, Francisco das Chagas Douglas Marques, Neto1, José Xavier, Sobrinho, Carlos Roberto Martins Rodrigues, Marques, and João Alexandre Lôbo. (2021) “Predictive models to the COVID-19.” *Data Science for COVID-19*: 1-24.

[11] Ramesh Kumar Mojjada, Arvind Yadav, A.V. Prabhu, and Yuvaraj Natarajan. (2020) “Machine Learning Models for covid-19 future forecasting.” *published online ahead of print*: 1-5.

[12] Alghazzawi, Daniyal, Atika Qazi, Javaria Qazi, Khulla Naseer, Muhammad Zeeshan, Mohamed E.M. Abo, Najmul Hasan, Shiza Qazi, Kiran Naz, Samrat K. Dey, and Shuiqing Yang. (2021) “Prediction of the Infectious Outbreak COVID-19 and Prevalence of Anxiety: Global Evidence.” *Sustainability*: 1-16.

[13] Manish Kumar Sharma, Prince Kumar, Akhtar Rasool, Aditya Dubey, and Vishal Kumar Mahto (2021) “Classification of Actual and Fake News in Pandemic.” *International Conference on IoT in Social, Mobile, Analytics and Cloud (I-SMAC)*: 1168-1174.

[14] Cafer Mert Yeşilkanat. (2020) “Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm.” *Chaos, Solitons & Fractals* **140**: 1-8.

[15] Muhammad Shoaib, Hamza Salahudin, Muhammad Hammad, Shakil Ahmad, Alamgir Akhtar Khan, Mudasser Muneer Khan, Muhammad Azhar Inam Baig, Fiaz Ahmad, and Muhammad Kaleem Ullah. (2021) “Performance Evaluation of Soft Computing Approaches for Forecasting COVID-19 Pandemic Cases.” *SN Comput Sci:* 1-13.

[16] Dimple Tiwari, Bhoopesh Singh Bhati, Fadi Al-Turjman, and Bharti Nagpal. (2021) “Pandemic coronavirus disease (Covid-19): World effects analysis and prediction using machine-learning techniques.” *Expert Syst*: 1-20.

[17] Rustam, Furqan, Reshi, Aijaz Ahmad, Mehmood, Arif, Ullah, Saleem, On, Byung-Won, Aslam, Waqar, Choi, and Gyu Sang. (2020) “COVID-19 Future Forecasting Using Supervised Machine Learning Models.” *IEEE Access* **8**: 101489-101499.

[18] Sahai, Alok Kumar, Rath, Namita, Sood, Vishal, Singh, and Manvendra Pratap. (2020) “ARIMA modelling & forecasting of COVID-19 in top five affected countries,Diabetes & Metabolic Syndrome.” *Clinical Research & Reviews* **14**(**5**): 1419-1427.

[19] Abdelhafid Zeroual, Fouzi Harrou, Abdel kader Dairid, and YingSunc. (2020) “Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study.” *Chaos, Solitons & Fractals* **140**: 1-12.

[20] Parag Verma, Rajeev Tiwari, Wei-Chiang Hong, Shuchi Upadhyay, and Yi-Hsuan Yeh. (2022). “FETCH: A Deep Learning-Based Fog Computing and IoT Integrated Environment for Healthcare Monitoring and Diagnosis.” *IEEE Access* **10**: 12548-12563.

[21] Nabi, Khondoker Nazmoon, Tahmid, Md Toki, Rafi, Abdur, Kader, Muhammad Ehsanul, Haider, and Md. Asif. (2021) “Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks.” *Results in Physics* **24**: 1-13.

[22] Kırbaş, İsmail, Sözen, Adnan, Tuncer, Azim Doğuş, Kazancıoğlu, and Fikret Şinasi. (2020) “Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches.” *Chaos, Solitons & Fractals* **138**: 1-7.

[23] Patricia Melin, Julio Cesar Monica, Daniela Sanchez, and Oscar Castillo. (2020) “A new prediction approach of the COVID-19 virus pandemic behavior with a hybrid ensemble modular nonlinear autoregressive neural network.” *Soft comput*: 1-10.

[24] Xiaolei Liu and Zi Lin. (2021) “Impact of Covid-19 pandemic on electricity demand in the UK based on multivariate time series forecasting with Bidirectional Long Short Term Memory.” *Elsevier Ltd*: 1-11.

[25] Paulo Renato Alves Firmino, Jair Paulino de Sales, Jucier Gonçalves Júnior, and Taciana Araújo da Silva. (2020) “A non-central beta model to forecast and evaluate pandemics time series.” *Chaos, Solitons & Fractals* **140**: 1-11.

[26] Vinay Kumar Reddy Chimmula, and Lei Zhang. (2020) “Time series forecasting of COVID-19 transmission in Canada using LSTM networks.Chaos.” *Solitons & Fractals* **135**: 1-6.

[27] R Sujath, Jyotir Moy Chatterjee, and Aboul Ella Hassanien . (2020) “A machine learning forecasting model for COVID-19 pandemic in India.” *Stoch Environ Res Risk Assess 34:* 959–972.

[28] Gülhan Toğa, Berrin Atalay, and M Duran Toksari. (2021) “COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey.” *Journal of Infection and Public Health* **14** (**7**): 811-816.

[29] Iman Rahimi, Fang Chen, and Amir H. Gandomi. (2021) “A review on COVID-19 forecasting models.” *Neural Comput & Applic*: 1-11.

[30] Rahul Ahirwar, Akhtar Rasool, Aanya Chouhan, Aditya Dubey, Sachin Mehra, and Ankit Kumar. (2022) “Covid-19 Detection based on Transfer Learning & LSTM Network using X-ray Images.” *IEEE World Conference on Applied Intelligence and Computing (AIC)*: 300-306.