Improved Linear Factor based Grasshopper Optimization Algorithm with Ensemble Learning for Covid-19 Forecasting

P. Renukadevi[1*], Dr.A. Rajiv Kannan[2]

Abstract

Recently the COVID'19 is extensively increasing around the world with many challenges for researchers. Rigorous respiratory disease corona virus 2 show aggression to many parts of COVID'19 affected patients, together with brain and lungs. The changeableness of Corona virus with likely to infect Central Nervous System emphasize the necessity for technological development to identify, handle, and take care of brain damages in COVID'19 patients. An exact short-term predicting the quantity of newly infected and cured cases is vital for resource optimization to stop or reduce the growth of infection. The previous system designed a Linear Decreasing Inertia Weight based Cat Swarm Optimization with Half Binomial Distribution based Convolutional Neural Network (LDIWCSO-HBDCNN) approach for COVID-19 forecasting. However, the ensemble learning is required to improve the prediction outcome via integrating many approaches. This approach allows the production of better predictive performance compared to a single model. For solving this problem, the proposed system designed an Improved Linear Factor based Grasshopper Optimization Algorithm with Ensemble Learning (ILFGOA with EL) for covid-19 forecasting. Initially, the COVID-19 forecasting dataset is taken as an input. With the help of min-max approach, data normalization is done. Then the optimal features are selected by using Improved Linear Factor based Grasshopper Optimization Algorithm (ILFGOA) algorithm to improve the prediction accuracy. Based on the selected features, Ensemble Learning (EL) which includes Hyperparameter based Convolutional Neural Network (HCNN) is utilized to identify infected and demise cases across india for a period of time. The outcome of analysis shows that the introduced method attains better execution against previous system with regard to error rate, accuracy, precision, recall and f-measure.

Key Words: COVID-19, Ensemble Learning (EL), Improved Linear Factor based Grasshopper Optimization Algorithm (ILFGOA) and Min-max Normalization.

DOI Number: 10.14704/nq.2021.19.8.NQ21129 **NeuroQuantology 2021; 19(8):169-181**

Introduction

Increase in corona virus 2019 (COVID'19) has been a universal hazard so on 11th March 2020, WHO affirmed it as a worldwide virulent disease. Till 30th April 2020, total infected persons were about 3,359,055 and individuals who lost life are 238,999 due to COVID'19 across the globe. This corona-virus is widely influencing individual's life and globe's financial market (Prasad et al., 2020). Between various disease based queries, both the people and government are worried about 1) how long will the COVID'19 disease continue; 2) when this deadly infection stop increasing; 3) how many of them are finally confirmed; 4) how many of them lost their lives. These queries are major worry for India too, a country of huge populace and financial assortment (Zhang et al., 2020); (Muhammad et al., 2020). Many of the infections are coming through the direct contacts of a person, which causes one way of spreading.

Corresponding author: P. Renukadevi

Address: [1*]Assistant Professor, Paavai Engineering College, Namakkal, Tamil Nadu, India; [2]Professor, K.S.R. College of Engineering, Namakkal, Tamil Nadu, India.

[1*]E-mail: pgrenu@gmail.com

[2]E-mail: rajiv5757@yahoo.co.in

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 14 June 2021 **Accepted:** 29 July 2021

It may spread the disease by touching some contaminated surface or fabric and then touching one's mouth, nose, or eyes (Rehman et al., 2021); (Pandey et al., 2020); (Ardabili et al., 2020). For people with the infection incubation period can be from one to fourteen days.

In consonance with WHO, the indication for mild and modest cases are dehydrated cough, tiredness and running temperature whereas for severe cases breath shortness, high temperature with fatigue might happen. Individuals who already have infections similar to respiratory disorder, uneven glucose level and heart problems are easily affected by this virus and get rigorously sick (Kumari et al., 2021). There are rising facts that shows the individuals who have severe lung damage are prone to brain damage via hypoxemia and/or pro-inflammatory mediators which interlink lung with brain. Individuals are identified depending on the indication and their history of journey. Highly difficult part of this infection is an individual can have this virus for long time without any indications. Due to its spreading nature and threat, most of the nations affirmed part or complete lock-down for entire affected areas.

Conventional prediction accuracy is based on the data availability for their identification and approximate vagueness. Efficient diagnosing will identify COVID'19 rapidly and effectively to reduce the health care professionals' burden. The diagnose framework integrated with many features are developed to guess the infection risk and also to help the health care professionals all over the world in prioritizing patients with limited resources. Machine learning techniques are believed to be used widely in medical applications, so the goal is to construct a model that can identify COVID'19 (Yadav et al., 2020). ML related prediction system have established their importance in predicting preoperative results for enhancing the judgment for further action. These systems are constructed by using a collection of training data that contains all the COVID'19 features. Thus prognostic and exploratory nature of ML are vital to handle quick judgment which might assist to keep away from COVID'19 virus.

In (Chimmula & Zhang 2020), proposed a stacked auto-encoder framework that fit the vibrant spread of disease and live prediction of infected cases in China. In (Hu et al., 2020), explored a deep learning approach based on Long Short-Term Memory for predicting the COVID'19 spread in Italy, China and Canada. In (Prasanth et al., 2021) constructed a framework named a hybrid Grey Wolf Optimizer - Long Short Term Memory for diagnosing COVID'19 by optimizing the LSTM network parameters by Grey Wolf Optimizer. However, these methods does not provides satisfactory results in terms of accuracy and error rate. Hence, better feature selection and classification methods required to enhance the accuracy of forecast.

Remaining part of the article is as indicated: Section 2 presents the prediction model of previous research efforts. Section 3 provides the proposed technique. Section 4 provides experimental outcomes with performance comparison. Finally, Section 5, provides the conclusion and describes the future research opportunities.

Literature Review

Rustam et al., (2020) explains the capacity of machine learning techniques in predicting the amount of future individuals infected by COVID'19 that is currently regarded as impending danger for human. Specifically, four typical predicting approaches; like LR, LASSO, SVM, ES is utilized to predict the COVID'19 forbidding factors. Forecasting were of three forms like amount of freshly affected cases, demise number, amount of recovered cases for subsequent 10 days. Outcome of the study indicated that it is hopeful to utilize these techniques for present situation of COVID'19. It also indicated that ES provides better result among other methods along with LR, LASSO that provides better prediction in identifying freshly affected cases, demise rate and recovered cases whereas SVM provides reduced performance in prediction. However it does not suitable for large database.

Tamang et al., (2020) uses Artificial Neural Network (ANN) based curve fitting approach to forecast the amount of growing COVID'19 cases and demise rates in India, france, UK, USA in view of extreme trends in south korea, china. Here, three cases are measured to estimate the COVID'19 eruption namely 1) predicting the current trend of growing cases of various countries; 2) predicting a week follow up along with development in accordance with china, south korea; 3) predicting whether followed trends grow in accordance with south korea and china a week ahead. Outcome of analysis indicated that ANN effectively predicts the upcoming COVID'19 cases of any nation. It also indicated that affected cases of USA, India, France, UK are 12,00,000 to 17,00,000, 50,000 to 1,60,000, 2,40,000 to 2,50,000, 1,40,000 to 1,50,000 correspondingly and possibly will take 2 to 10 months depending on growth rate in south korea and china. Likewise, demise rate for all the nations prior controlling will be 1600 to 4000 - India, 1,00,000 to 1,35,000 - USA, 40,000 to 55,000 - France, 35,000 to 47,000 – UK at the similar time of research. However, the duration of the Artificial Neural Network (ANN) is unknown.

Shetty (2020) exhibits the capacity of Multilayer Perceptron (MLP), ANN system for predicting the amount of affected cases in Karnataka, India. a quick training algorithm is employed for training such as Extreme Learning machine (ELM) to decrease the training period. Partial Auto Correlation Function (PACF) is utilized for choosing parameters that are needed for predicting system, which is a traditional technique and the execution is correlated against CS approach chosen parameters, a well-liked nature-inspired optimization procedure. testing of prediction system is performed and contrast among two parameter selection approach is done. CS method provides good prediction depending on MAPE (Mean Absolute Percentage Error) value 6.62 % on training and 7.03% on testing data. additionally, efficiency of the system is evaluated for COVID'19 cases of Hungary starting March 4th 2020 till April 19th2020. It provides MAPE of 1.55%, thus provides strength of proposed ANN model for predicting COVID'19 cases in Karnataka. However, it has issue with high computation time.

Ahamad et al., (2020) proposed a system which uses supervised ML techniques to predict the presentation features of COVID'19 infection with improved accuracy. The features analyzed contains the information about patients like gender, age, fever observation, travel history, medical information like cough severity, lung disease occurrence. Many ML techniques were used to gather information and identified that XGBoost algorithm performs better with prediction accuracy >85% and chosen features appropriately show COVID-19 position for entire age groups. Experimental analysis indicated that main important indications are fever with 41.1%, cough with 30.3%, lung infection with 13.1%, cold with 8.43%. whereas 54.4% individuals analyzed does not possess any indications utilized for prediction. Forecasting system can considerably advance the forecast of COVID-19 position, considering premature phase of disease. However, it has issue with classification accuracy.

Jia et al., (2020) designed a progress trend examination of incremental affected cases, incremental demise, incremental recovered case depending on information from Wuhan, Hubei Province, China starting 23rd January 2020 till 6thapril 2020 through Elman neural network, LSTM, SVM. SVM through fuzzy granulation is employed to forecast the progress of affected new cases, new demise and recovered cases. Experimental analysis indicated that Elman neural network, SVM forecast the growth of incremental affected numbers, demise and recovered numbers but LSTM is highly appropriate for forecasting incremental affected numbers. SVM with fuzzy granulation can effectively forecast the augmentation affected fresh cases along with fresh recovered cases, even though average predicted values are high. However, better classification methods are required to improve the true positive rate.

Proposed Methodology

The coronavirus various parts of COVID'19 affected individuals, together with brain and lungs (Claassen et al., 2019); (Kaseda & Levine 2020). An individual with COVID'19 with more high white matter damage and mainly obvious lactate elevation on Magnetic Resonance Spectroscopy is another symptom for brain injury for oxygen deficiency. Hence different approaches have been proposed to detect the spread of infectious diseases to take proper medical assistance. The proposed system designed an Improved Linear Factor based Grasshopper Optimization Algorithm with Ensemble Learning (ILFGOA with EL) for effective covid-19 forecasting. Figure 1 depicts the flow of the introduced system.

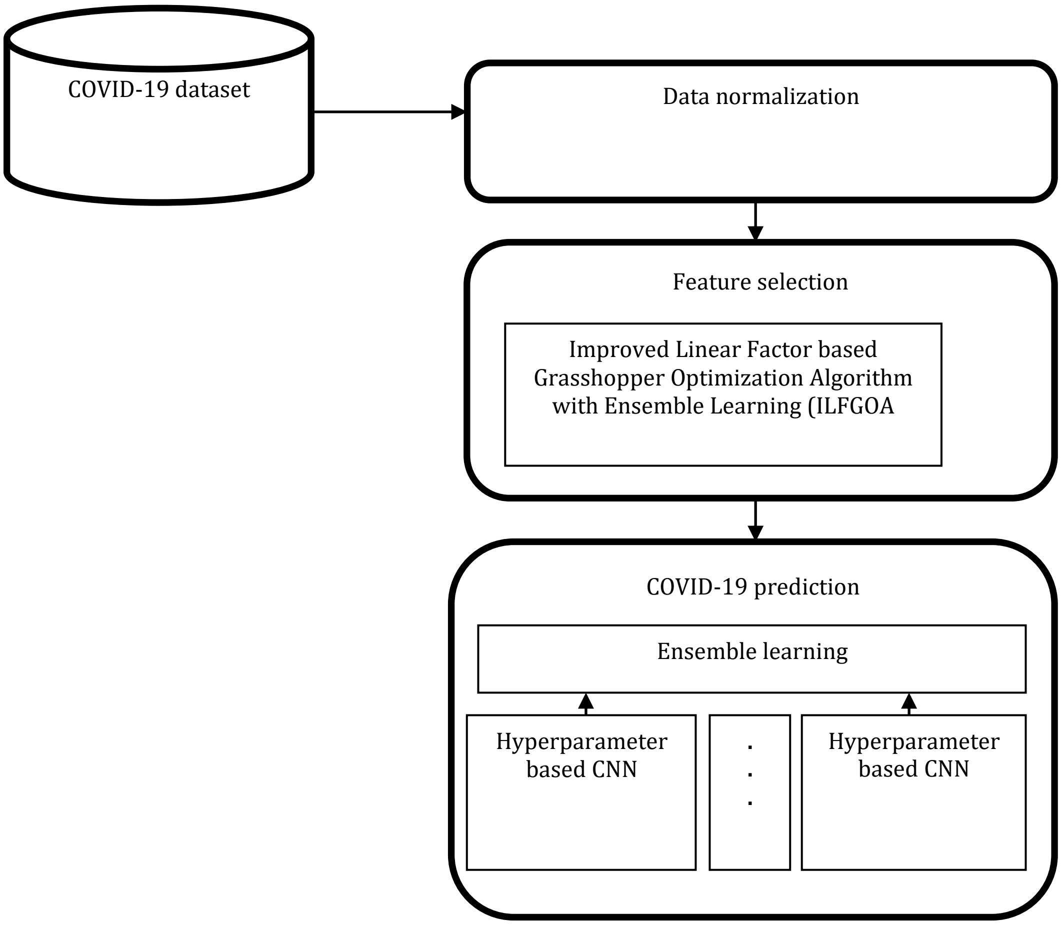

Figure 1. Flow diagram of the proposed model

Input

COVID-19 forecasting database is taken as an input. It is collected from https://www.kaggle.com/davidbnn92/weather-data-for-covid19-data-analysis. In this database the information, such as id, state, region, lat, long, date, confirmed cases, fatalities, temperature, fog and etc are added. Weather data is imported from the NOAA GSOD dataset, continuously updated to include recent measurements.

Data Normalization Using Min-max Normalization

Normalization process converts numerical data to a new form by utilizing mathematical function. Here, min-max normalization is utilized to normalize covid-19 dataset. Min-max normalization is one of the most common ways to normalize data. The values in the dataset are normalized within the given range and every data are altered based on the given equation.

$$v' = \frac{v - min_A}{max_A - min_A}(new\_max_A - new\_min_A) + new\_min_A \quad (1)$$

Where,

A is the Attribute value

Min(A), Max(A) –A's least and highest absolute value correspondingly

v'- new value for every data entry

v - old value for every data entry

new_max(A), new_min(A) –high and low value in the range correspondingly

Feature Selection Using Improved Linear Factor based Grasshopper Optimization Algorithm (ILFGOA)

Feature selection is done by using Improved Linear Factor based Grasshopper Optimization Algorithm. The Grasshopper optimization algorithm simulates the grasshoppers behavior. Normally, grasshopper swarm travel for a long distance to locate a new environment with food. Here, communication between grasshoppers manipulate each other within the swarm. Wind and gravity power outside swarm have impact on grasshoppers' route. Food intention is also a major deciding factor.

Metaheuristic algorithms reasonably separate the searching process into two stages namely, exploitation, exploration. During exploration phase, grasshoppers are motivated to travel quickly in search of highly likely place. During exploitation, grasshoppers travel in local in search of fine target place. These two travelling stage in search of food source are accomplished through grasshopper obviously. It acts as an abstract of optimization problem. grasshopper swarm is extracted from swarm of search agents.

In this proposed research work, features in the dataset are taken as an input. Here, every feature in a dataset is given a position and fitness value. Position represents a point corresponding to a possible feature set. The classification accuracy is considered as fitness function.

Fitness function =Max (Accuracy) (2)

S. Saremi offered a numerical model for immigration of grasshopper swarm. Equation for the simulation is given as:

$$X_i = S_i + G_i + A_i \quad (3)$$

Where, X_i is ith feature position, S_i indicate social interaction strength, G_i indicate ith feature manipulating factor of significance force, and A_i indicate wind's impact factor. S_i is described as:

$$S_i = \sum_{j=1, j\neq i}^{N} s(d_{ij})\hat{d}_{ij} \quad (4)$$

Where d_{ij} describe Euclidean distance among ith and jth feature in dataset, and is computed as $d_{ij} = |x_j - x_i|$, $\hat{d}_{ij}$ is unit vector from ith to jth attribute, described like $\hat{d}_{ij} = \frac{|x_j - x_i|,}{d_{ij}}$, s is a function that indicates social relationship affection in the feature and is described as:

$$s(r) = fe^{\frac{-r}{l}} - e^{-r} \quad (5)$$

Here, e indicates Natural Logarithm, f indicates concentration of attraction and l explains attractive length scale.

While using for handling numerical optimization problem, minor variation factors need to be added in equation 3 for numerical module optimization. G_i and A_i indicate outside control substituted by food target parameter. hence equation is reframed as:

$$x_i = c\left(\sum_{j=1, j\neq i}^{N} c\frac{u-l}{2} s(|x_j - x_i|)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d \quad (6)$$

here u and l indicate search space's upper and lower boundaries correspondingly, $\hat{T}_d$ is location of food target that indicates finest optimal location that entire search features can identify all-time in numerical module. moreover, c is the comfort zone parameter modifying to equalize exploitation and exploration process that is computed as:

$$c = c_{max} - iter\frac{c_{max} - c_{min}}{Max_{iter}} \quad (7)$$

Where, c_{max} and c_{min} are maximum and minimum value of c correspondingly, iter indicate the present iteration, and Max_{iter} indicate maximum iterations. Equation (6) needed to be repeated again for optimal solution evolution and it is terminated when the condition is satisfied. Generally, when the predetermined maximum iterations reached. After evolution process terminates, the procedure obtain rough finest fitness and its subsequent target location.

- *Improved Linear Factor based Grasshopper Optimization Algorithm*

The conventional Grasshopper Optimization Algorithm (GOA) has minor change due to deficiency of arbitrary factors. parameter c in parenthesis helps reduce the attraction and repulsion between grasshoppers. If the parameter drops too fast, the convergence in the early stage of GOA algorithm will be insufficient. And also make a balance between exploration and exploitation is crucial issues. To overcome these problems, the comfort zone parameter is updated using Improved Linear Factor (ILF).

$$c = (c_{min} - c_{max})\left(\frac{T_{max} - t}{T_{max}}\right) + c_{max} \quad (8)$$

Where,

c_{max}- maximum value of c

c_{min} - minimum value of c

T_{max}- maximum number of iterations

Algorithm 1: ILFGOA

Input: Number of features in the dataset

Output: Optimal features

1. Initialize number of features X_i (i = 1, 2, · · ·, n)
2. Set c_{max}, c_{min}, maximum iterations
3. compute classification accuracy for all features
4. Target = best search agent
5. while (1<Maximum iterations)
6. revise c using eq (8)
7. Forevery search agent
8. standardize the distanceamong grasshoppers in [1,4]
9. revise the current featureposition by the eq (6)
10. get the current search agent if it outside the boundaries
11. End for
12. revise target if better solution
13. l=l+1
14. end while
15. Return features

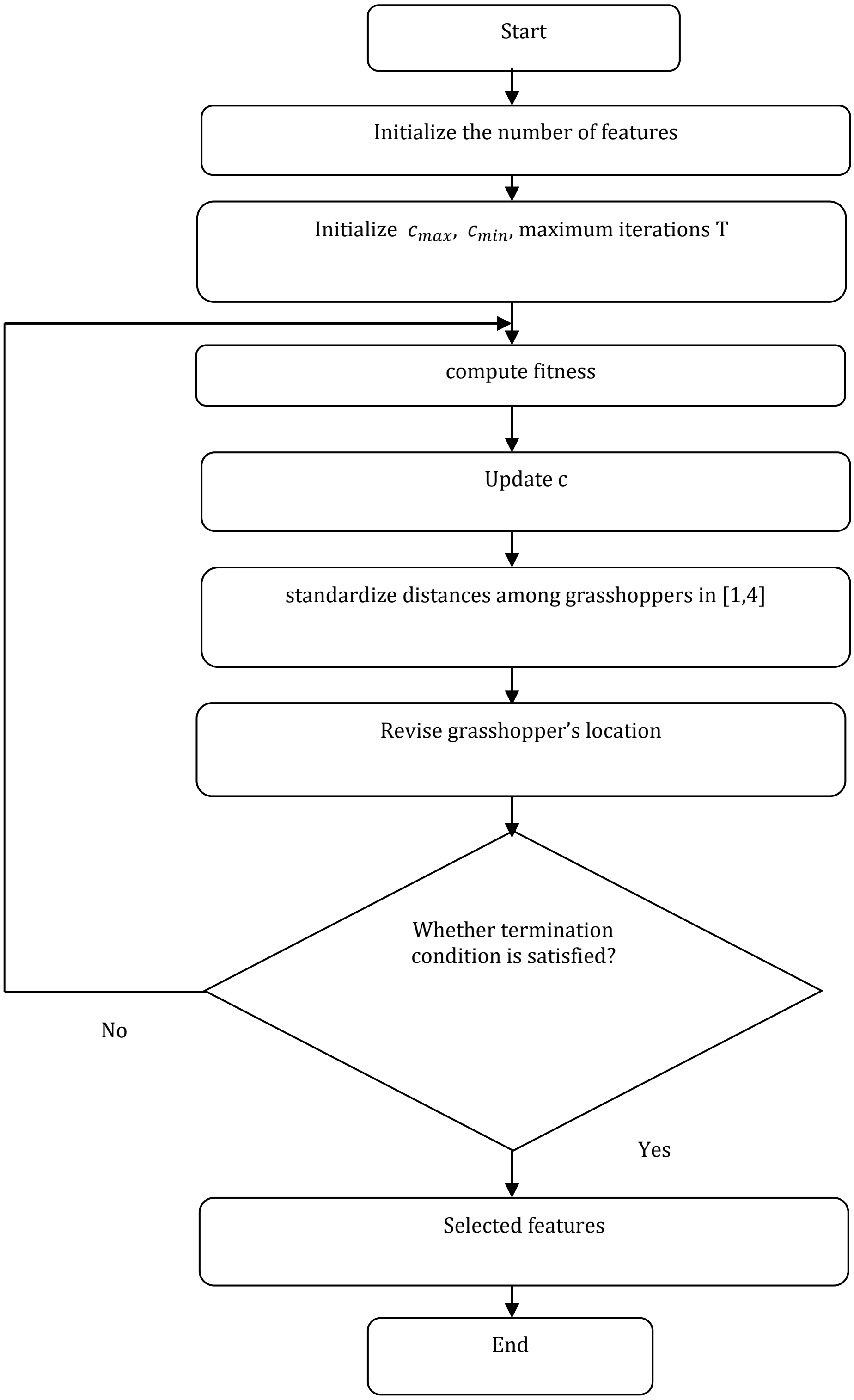

Figure 2. Operation flowchart of ILFGOA

Classification Using Ensemble Learning (EL)

In this proposed research work introduced Ensemble Learning (EL) which includes multiple Hyperparameter based Convolutional Neural Network (HCNN) for covid 19 prediction which is utilized for forecasting. Amongst the most potent deep networks is the CNN which contain many hidden layers doing convolution and sub sampling to extort low to high levels of features from input. Basically, CNN contains three layers: convolution, sub sampling or pooling, full connected layers which is shown in figure 3.

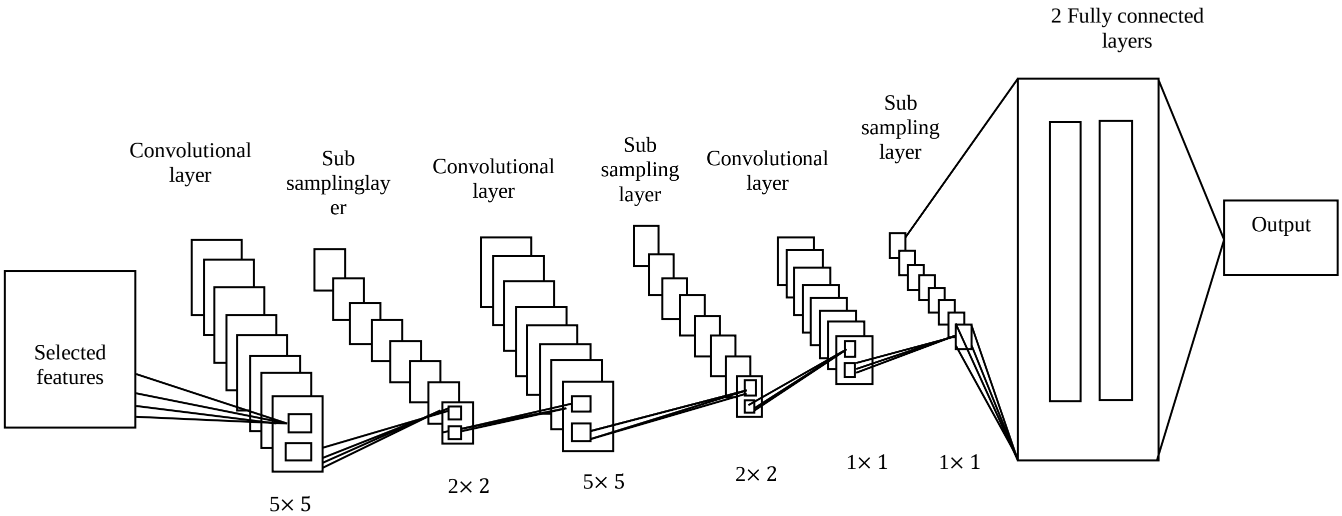

Figure 3. Convolutional Neural Network

- *Convolution Layer*

In this proposed work, the selected features are taken as an input. In this convolution layer, an input features are convolved with a kernel (filter). Outcome of convolution of input feature and kernel is employed to create n output features. usually, a kernel of the convolution matrix is termed to be a filter whereas the output features acquired by convolving kernel and input are termed as feature maps with size i*i.

Convolutional neural network contains many convolutional layers, the inputs and outputs. A cluster of n filters is there in every convolution layer and filters are convolved with input, and depth of generated feature maps (n*) is equal to the amount of filters useable in the convolution operation. To be noted every filter map is believed to be a particular feature at absolute input location.

$C_i^{(l)}$ denotes the lth convolutional layer that contains feature maps and is calculated like

$$C_i^{(l)} = B_i^{(l)} + \sum_{j=1}^{a_i^{(l-1)}} K_{i,j}^{(l-1)} * C_j^{(l)} \qquad (9)$$

Where, $B_i^{(l)}$is bias matrix,$K_{i,j}^{(l-1)}$is convolution filter/ kernel with size a*a that links j-th feature map in layer $(l-1)$ with i-th feature map in similar layer. $C_i^{(l)}$ layer output contains feature maps. For equation (10), initial convolutional layer $C_i^{(l-1)}$is input space, i.e., $C_i^{(0)} = X_i$. kernel produces feature map. Following the convolution layer, the activation function is used for nonlinear transformation of convolutional layer outputs

$$Y_i^{(l)} = Y(C_i^{(l)}) \qquad (10)$$

here,

$Y_i^{(l)}$ - output of activation function

$C_i^{(l)}$- received input.

Usually employed activation functions are sigmoid, tanh, ReLUs. here, we have introduced ReLUs that is signified as

$$Y_i^{(l)} = \max\left(0, Y_i^{(l)}\right) \qquad (11)$$

This function is popularly used in deep learning models due to its help in reducing the interaction and nonlinear effects. ReLU change the output to 0 for negative input, whereas it provides same input value for positive. Benefit of this activation function above other functions is the quicker training since it is error derivative, that is reduced to very small in saturating region; hence, weight updating nearly become extinct. It is termed as vanishing gradient problem.

- *Sub Sampling or Pooling Layer*

Goal of this layer is to contiguously decrease the feature map dimensionality extorted from the preceding convolution layer. The sub sampling operation performed among the mask and feature maps. In order to perform, a mask with size b∗b is chosen and sub sampling operation among the mask and attribute maps is carried out.

$$X_j^l = f(\beta_j^l \text{down}\,(X_j^{l-1}) + b_j^l) \qquad (12)$$

here, down (·) indicates sub-sampling function. usually this function will sum over every separate n-by-n attributes in input dataset such that output is n-times lesser along both spatial dimensions. Every output map will have their multiplicative bias $\beta$and an additive bias b.

- *Hyperparameter based Pooling*

one input feature map of a pooling layer as an assemblage X_i.In an arbitrary pooling region X_i, denote the -picked activation as act_l, where lϵ $[1, k]$.

$$act_l = \max(X_i \ominus \sum_{j=1}^{l-1} act_j) \quad (13)$$

here, $\ominus$indicatesdeleting elements from assemblage. summation symbol in (13)symbolize a small set of elements, that consists of top1~(l-1) activation however not accumulation up all the activation numerically.

once picking the top-k activation, it is not added together as an output or calculated the average of them. In this work introduced a hyper parameter γ as a constraint factor to multiply the sum of the top-k activation. The last output refers to

$$\text{Output} = \gamma * \sum_{j=1}^{k} act_j \quad (14)$$

Where, the summation symbol indicates sum operation, $\gamma \in (0,1)$. Mainly, if $\gamma = \frac{1}{k}$, the output is the average value. Constraint factor,γ, is employed to regulate the output. Hyperparameter based pooling scheme improve the strong sparse features representation capability to achieve higher classification accuracy.

- *Fully Connected Layer*

The last layer of CNN is a traditional feed forward network with more than one hidden layers. It utilizes Softmax activation function:

$$Y_i^{(l)} = f(z_i^{(l)}) \quad (15)$$

$$\text{Where } z_i^{(l)} = \sum_{i=1}^{m_i^{(l-1)}} w_{i,j}^{(l)} y_i^{(l-1)} \quad (16)$$

Where, $w_{i,j}^{(l)}$ are weights supposed to be tuned by entire fully connected layer so as to form ever class representation, f is transfer function that indicates nonlinearity. To be noted that nonlinearity in fully connected layer is constructed within its neurons, and not as individual layers like convolutions and pooling layers. The output of Hyperparameter based Convolutional Neural Network (HCNN) is introduced to predict confirmed and death cases across india in particular period.

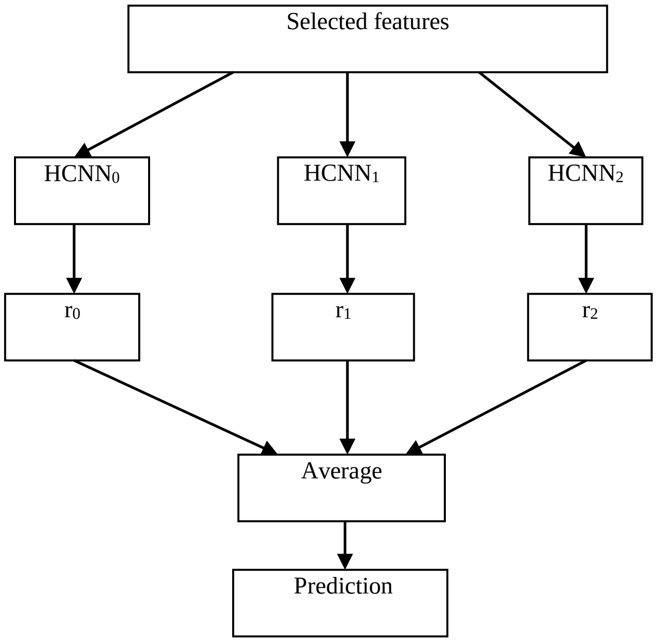

Figure 4. Ensemble Convolutional Neural Networks

For a given selected features, the output probabilities from Hyperparameter based Convolutional Neural Network (HCNN) are averaged before making a decision. For output i, the average output S_i is given by:

$$S_i = \frac{1}{n} \sum_{j=1}^{n} r_j(i) \quad (17)$$

here, $r_j(i)$ is the output i of jth network for a given input features.

Experimental Results

The experimental assessment is performed using matlab for present and introduced research techniques. In this research work, COVID-19 forecasting data is collected from https://www.kaggle.com/davidbnn92/weather-data-for-covid19-data-analysis. Performance of the proposed Ensemble Learning (EL) is compared with the previous Half Binomial Distribution Convolutional Neural Network (HBDCNN) and CNN and LSTM networks with respect to recall, accuracy, error rate, precision, f-measure. Additionally, the obtained outcome result is correct (i.e.true) or incorrect (i.e.false).

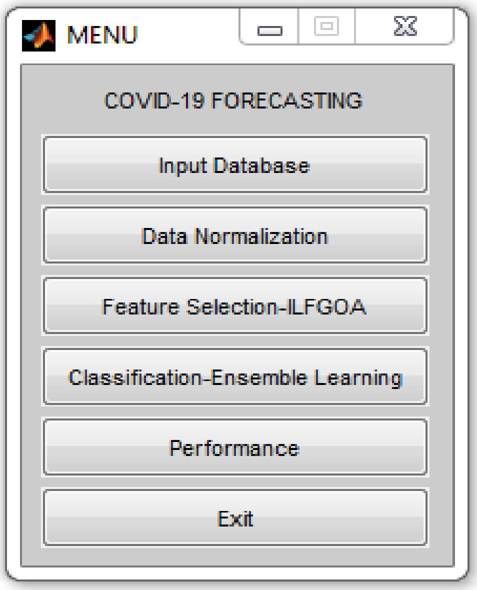

Figure 5. Improved Linear Factor based Grasshopper Optimization Algorithm with Ensemble Learning approach

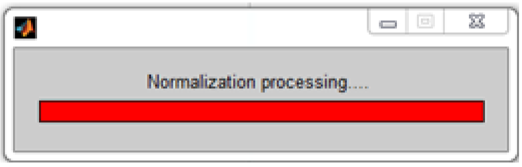

Figure 6. Data normalization

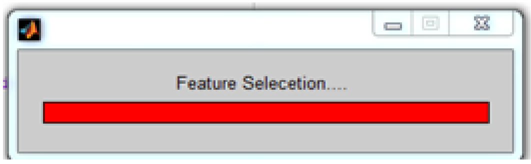

Figure 7. Feature selection

The overall process of the proposed Improved Linear Factor based Grasshopper Optimization Algorithm with Ensemble Learning approach is shown in figure 5. Figure 6 represents the data normalization process. In this proposed work, Improved Linear Factor based Grasshopper Optimization Algorithm (ILFGOA) is utilized for optimal feature selection which is shown in figure 7.

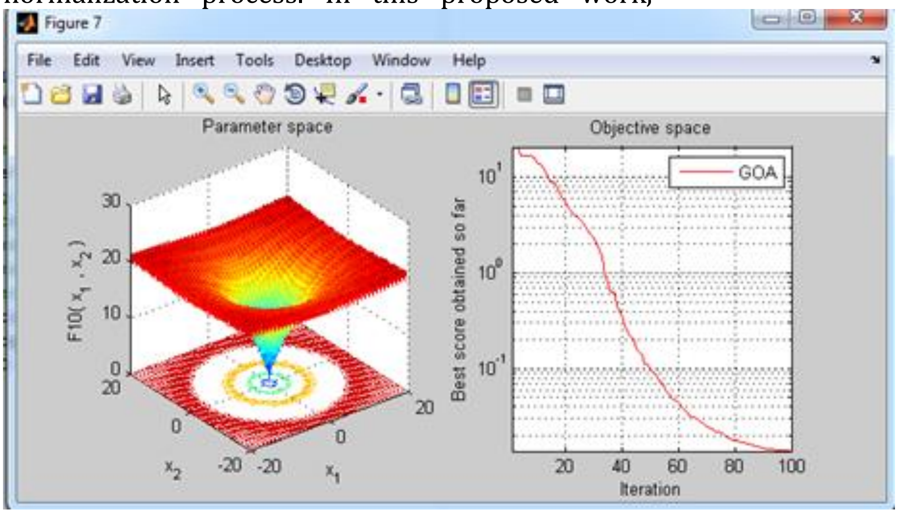

Figure 8. Improved Linear Factor based Grasshopper Optimization Algorithm

The objective space of the proposed Improved Linear Factor based Grasshopper Optimization Algorithm is represented in figure 8.

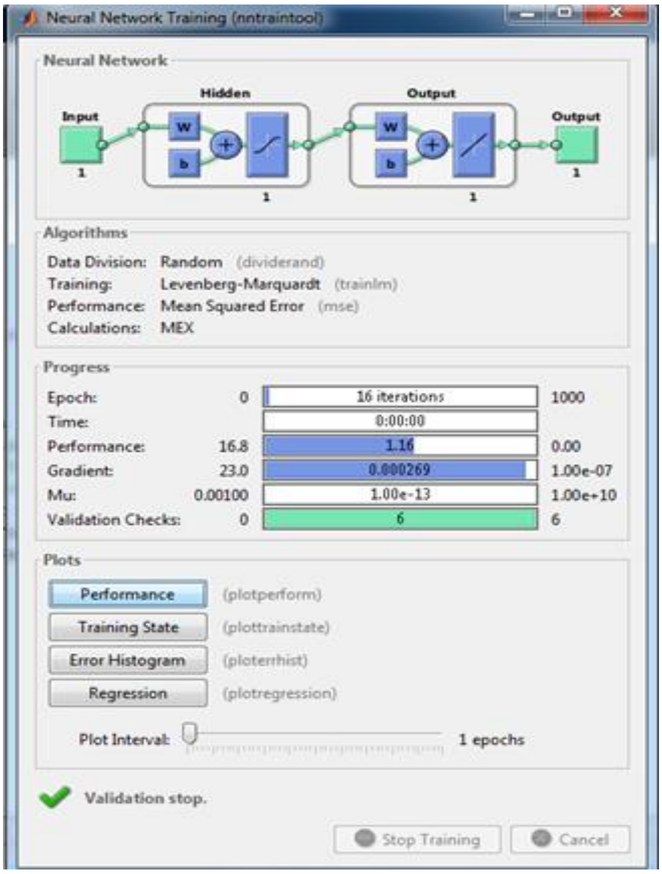

Figure 9. Neural network training

Figure 9 shows the training process of the Hyperparameter basedCNN (HCNN) for predicting the confirmed and death cases across India. Table 1 represents the introduced system's performance evaluated against previous methods in terms of accuracy, precision, recall, f-measure, error rate.

Table 1. Performance comparison

Metrics	Methods			
	LSTM	**CNN**	**HBDCNN**	**EL**
Accuracy	87.9220	90.7445	92.5777	94.8706
Precision	87.7028	90.5500	92.4192	95.3245
Recall	87.8875	90.7531	92.5654	94.1894
F-measure	87.7950	90.6514	92.4922	94.7535
Error Rate	12.0780	9.2555	7.4223	5.8294

1. Accuracy

It is the main sensitive production metric which is a proportion of appropriately identified observation to entire observations.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \qquad (18)$$

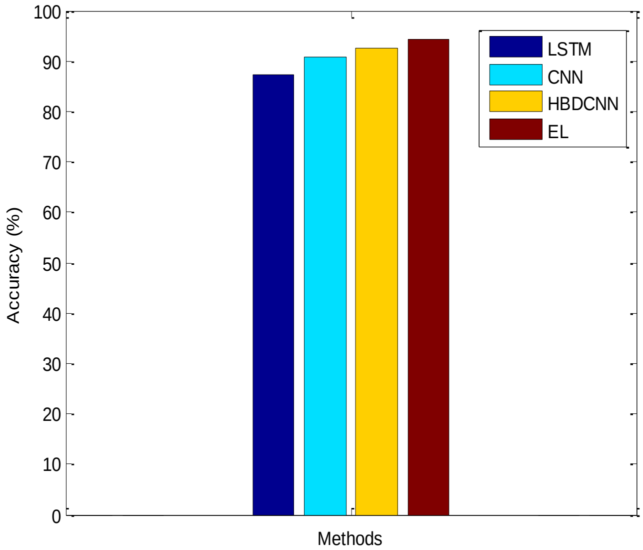

Figure 10. Accuracy comparison

From the above Figure 10, the introduced Ensemble Learning (EL) is contrasted against previous LSTM, CNN, HBDCNN interms of accuracy. The techniques are considered in X-axis and the accuracy in y-axis. In this research work, optimal features are selected by Improved Linear Factor based Grasshopper Optimization Algorithm (ILFGOA). According to the selected features, prediction is performed by using ensemble learning to achieve higher accuracy. The outcome of execution proved that the introducedalgorithmobtains accuracy of 94.87%. The investigational outcome indicate that the introduced algorithm provides greater accuracy value than existing methods.

2. Precision

Precision is the proportion of appropriately identified positive observations to entire identified positive observations and is calculated using the following equation,

$$Precision = \frac{TP}{TP+FP} \quad (19)$$

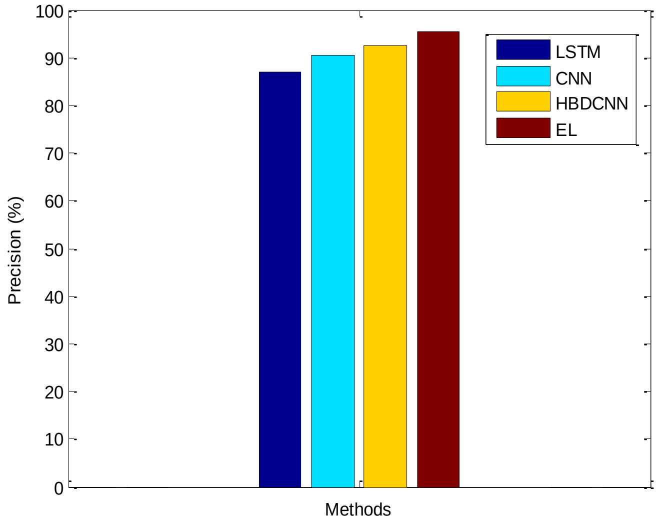

Figure 11. Precision comparison

Precision of the proposed ensemble learning is compared with the existing LSTM, CNN and HBDCNN methods are depicted in figure 11.
Techniques are plotted in x-axis and precision in y-axis. For introduced technique, the classification is done using Ensemble Learning (EL). Here outputs of the three Hyperparameter based Convolutional Neural Network (HCNN) classifiers are taken to improve the true positive rate. Outcome of result proved that the introduced algorithm attains higher precision of 95.32% whereas other methods such as LSTM, CNN and HBDCNN attains 87.70%, 90.55% and 92.41% respectively.

3. Recall

In covid-19 prediction, recall is proportion of actual true cases that are successfully found.

$$Recall = \frac{TP}{TP+FN} \quad (20)$$

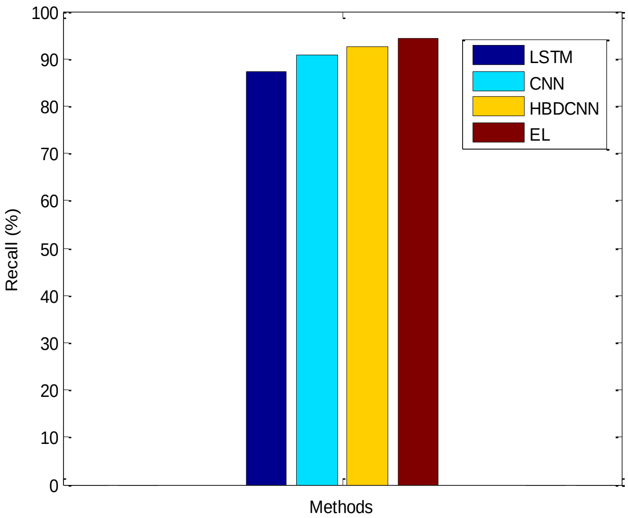

Figure 12. Recall comparison

Figure 12 shows the performance of proposed Ensemble Learning (EL) and existing LSTM, CNN and HBDCNN methods in terms of recall.
Techniques are plotted in x-axis and recall in y-axis. Outcome of result concluded that the introduced system attains 94.18% of recall whereas other method such as LSTM, CNN and HBDCNN attains 87.88%, 90.75% and 92.56% respectively.

4. F1 Score

F1 score is the weighted mean of precision and recall and is employed for numerical metric to rate the classifier's execution. Hence, it considers false positives and false negatives. It is formulated as

$$F - measure = 2\frac{Precision * Recall}{Precision + Recall} \quad (21)$$

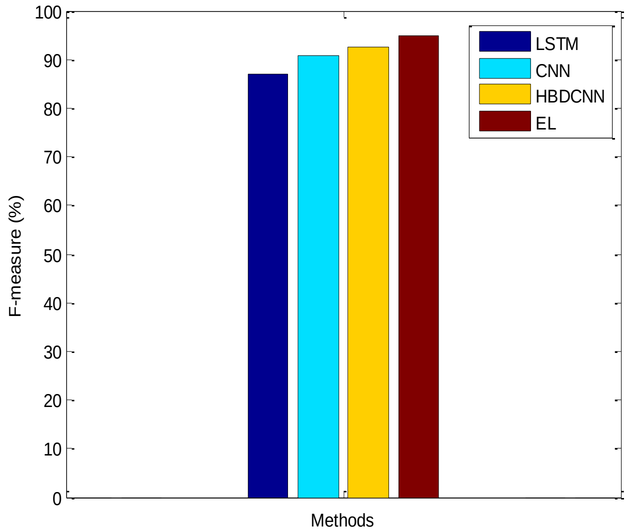

Figure 13. F-measure comparison

F-measure of the proposed EL approach is contrasted against present LSTM, CNN and HBDCNN methods as depicted in figure 13. Methods are considered in x-axis and F-measure in y-axis. Outcome of the results indicated that the introducedEL algorithm gives higher f-measure value of 94.75% when other methods such as LSTM, CNN and HBDCNN 87.79%, 90.65% and 92.49% respectively.

5. Error Rate

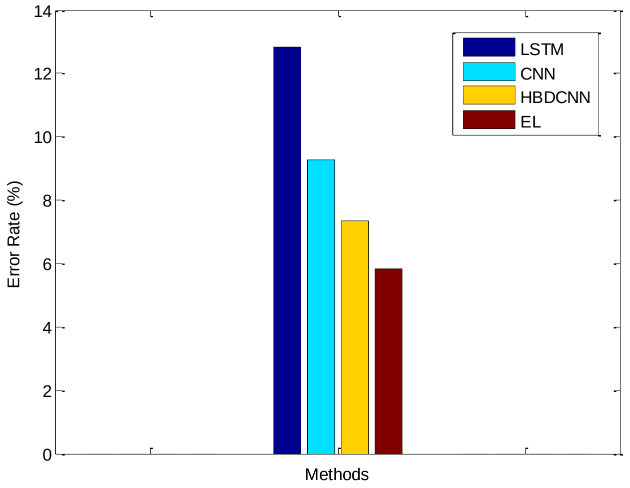

Figure 14. Error rate comparison

The error rate of the proposed system is contrasted against LSTM, CNN and HBDCNN methods as depicted in figure 14. Methods are considered for x-axis and error rate for y-axis. Outcome of results shows that introduced EL approach attains 5.82% of error rate where as other methods such as LSTM, CNN and HBDCNN achieves 12.07%, 9.25% and 7.42% respectively.

Conclusion

The proposed system designed a new Improved Linear Factor based Grasshopper Optimization Algorithm with Ensemble Learning (ILFGOA with EL) for covid-19 forecasting. In order to normalize covid-19 dataset, min-max normalization is utilized. To achieve higher classification accuracy, Improved Linear Factor based Grasshopper Optimization Algorithm (ILFGOA) is utilized for optimal features selection. According to the selected features, Ensemble Learning (EL) which includes Hyperparameter based Convolutional Neural Network (HCNN) is utilized to predict affected and demise cases across India for over a period. The outcome of analysis shows that the proposed system attains 94.87% of accuracy, 95.32% of precision, 94.18% of recall, 94.75% of f-measure and 5.82% of error rate which is better than previous methods such as LSTM, CNN and HBDCNN. In future various dimension reduction methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA) and etc are utilized to reduce the computation time.

References

Prasad CTB, Rajamani S, Dharshan HV, Patil S, Roy P, Srikantiah C, Raghavendra GA. Coronavirus (COVID-19) forecasting in India: Application of ARIMA and periodic regression models. *International Journal of Advanced Scientific Research* 2020; 5(1): 24-28.

Zhang X, Ma R, Wang L. Predicting turning point, duration and attack rate of COVID-19 outbreaks in major Western countries. *Chaos, Solitons & Fractals* 2020; 135: 109829.

Muhammad LJ, Islam MM, Usman SS, Ayon SI. Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. *SN Computer Science* 2020; 1(4): 1-7.

Rehman MU, Shafique A, Khalid S, Driss M, Rubaiee S. Future Forecasting of COVID-19: A Supervised Learning Approach. *Sensors* 2021; 21(10): 1-17.

Pandey G, Chaudhary P, Gupta R, Pal S. SEIR and Regression Model based COVID-19 outbreak predictions in India. *arXiv preprint arXiv:2004.00958*, 2020.

Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, Atkinson PM. Covid-19 outbreak prediction with machine learning. *Algorithms* 2020; 13(10): 249.

Kumari R, Kumar S, Poonia RC, Singh V, Raja L, Bhatnagar V, Agarwal P. Analysis and predictions of spread, recovery, and death caused by COVID-19 in India. *Big Data Mining and Analytics* 2021; 4(2): 65-75.

Yadav M, Perumal M, Srinivas M. Analysis on novel coronavirus (COVID-19) using machine learning methods. *Chaos, Solitons & Fractals* 2020; 139: 110050.

Chimmula VKR, Zhang L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals* 2020; 135: 109864.

Hu Z, Ge Q, Li S, Jin L, Xiong M. *Artificial intelligence forecasting of covid-19 in china.* arXiv preprint arXiv:2002.07112, 2020: 1-20.

Prasanth S, Singh U, Kumar A, Tikkiwal VA, Chong PH. Forecasting spread of COVID-19 using Google Trends: A hybrid GWO-Deep learning approach. *Chaos, Solitons & Fractals* 2021; 142: 110336.

Rustam F, Reshi AA, Mehmood A, Ullah S, On BW, Aslam W, Choi GS. COVID-19 future forecasting using supervised machine learning models. *IEEE access* 2020; 8: 101489-101499.

Tamang SK, Singh PD, Datta B. Forecasting of covid-19 cases based on prediction using artificial neural network curve fitting technique. *Global Journal of Environmental Science and Management* 2020; 6(19): 53-64.

Shetty RP. Realtime forecasting of COVID 19 cases in Karnataka state using Artificial neural network (ANN). *Research square* 2020: 1-9.

Ahamad MM, Aktar S, Rashed-Al-Mahfuz M, Uddin S, Liò P, Xu H, Moni MA. A machine learning model to identify early stage symptoms of SARS-Cov-2 infected patients. *Expert systems with applications* 2020: 1-12.

Jia L, Li K, Jiang Y, Guo X. *Prediction and analysis of Coronavirus Disease 2019.* arXiv preprint arXiv:2003.05447, 2020: 1-19.

Claassen J, Doyle K, Matory A, Couch C, Burger KM, Velazquez A, Rohaut B. Detection of brain activation in unresponsive patients with acute brain injury. *New England Journal of Medicine* 2019; 380(26): 2497-2505.

Kaseda ET, Levine AJ. Post-traumatic stress disorder: A differential diagnostic consideration for COVID-19 survivors. *The Clinical Neuropsychologist* 2020; 34(7-8):