Journal of Intelligent Systems and Internet of Things
Vol. 17, No. 01, PP. 342-359, 2025
DOI: https://doi.org/10.54216/JISIoT.170124

An Effective Mechanism for Early Pandemic Detecting COVID-19 Rediction based on Time Series Data and WPRO-Based Deep Learning RNN

Zaid Derea[1,*], Ammar Kazm[2], Manar Bashar Mortatha[2], Oday Ali Hassen[2,3] , Esraa Saleh Alomari[2]

[1]College of Computer Science and Information Technology, Wasit University, Wasit 52001, Iraq

[2]Department of Computer, College of Education for Pure Sciences, Wasit University. Iraq

[3]Ministry of Education, Wasit Education Directorate, Wasit, Iraq

Emails: zabdulameer@uowasit.edu.iq; aawaad@uowasit.edu.iq; manar@uowasit.edu.iq; odayali@uowasit.edu.iq; ealomari@uowasit.edu.iq

Abstract

Rapid spread of Corona virus 2019 (COVID-19) is predictable to create high contact on healthcare organization. Early detection of this disease is required to make precise treatment that further helps to increase the survival rate of humans. However, detecting the COVID-19 at beginning stage is one of a major challenge in the world because of rapid disease spread. Various existing methods are developed to detect the disease, but generating accurate result at the beginning stage still poses a complex task in the medical research community. Hence, an effective mechanism is modeled in this research to predict the pandemic at early with the time-series data using proposed Water Poor and Rich optimization-based Deep Recurrent Neural network (WPRO-based Deep RNN). Accordingly, proposed method is highly effective in generating the most appropriate results through deep learning classifier based on the high dimension features. However, the high dimensional data is generated through the data augmentation process by employing oversampling technique. The proposed method is more robust and increases the efficiency of the optimization algorithm by attaining global convergence results based on the fitness measure. Accordingly, the technical features of time series data to improve effectiveness of developed model. However, the proposed WPRO-based Deep RNN produced minimum Root Mean Square Error (RMSE) as well as MSE values of 0.4 and 0.1714 for confirmed cases, and obtained lesser MSE and RMSE values of 0.1887 and 0.433 for the cases of death. Moreover, proposed model achieved minimal RMSE and MSE of 0.447 and 0.1901 for the recovered cases.

Received: January 28, 2025 Revised: March 10, 2025 Accepted: April 02, 2025

Keywords: Deep learning; Water Cycle Algorithm (WCA); Epidemic prediction; Time-series; Water Poor and Rich optimization (WPRO)

1. Introduction

Data mining is the procedure to explore and analyze meaningful patterns, rules, models and knowledge either with the set of structured or the unstructured data. However, large analytics technology of data are employed to get useful details from unstructured data such that the data mining method is the big data analytics model used in the data processing system. The data mining is a procedure of comprehension and the pattern discovering with regards to OLAP technology, data warehousing, big data, and database such that data warehouse is the process of knowledge discovering in big data process (KDB) [17]. The healthcare industries are the hoarding and the engendering large volume of data that is specifically used to scrutinize and forecast healthcare ratio in whole world. With the data mining methods, it is more feasible to acquire useful and the essential information from database called as CBIS and KDD [16]. With the increasing growth of information technology (IT) in the medical and the health services field, the medical facilities gained more attraction and so the usage of healthcare platforms in the treatment and the diagnosis of medical area is achieved [25]. However, the growth of medical technology could result gradual increase in the lifespan of human and hence it causes rapid ageing in world wide. In general, the diseases are changed from acute to chronic type [20] [17]. The health care sciences and the data mining generate different reliable system to achieve early prediction process in related with the health care from medical diagnosis data. The data mining is depending on probability, statistics, machine learning and artificial intelligence [26] [18].

COVID-19 is the epidemic disease that broadens from the Wuhan city of china and transmitted over 213 countries crossways worldwide. According to report of World Health Organization (WHO) on February 2020, it is reported that 80% of patients with the viral disease have moderate fever and they are recovered, whereas the 2% of death cases are reported when compared with other diseases, such as SARS in 2003, and MERS between 2012 and 2019 such that they reported that death rate of 772 cases and confirmed cases are of 8089 that is around 10%, and nearly 858 cases are dead among 2494 cases with the value of 34%, respectively [2]. Covid is generally termed as SARS-COVID that is responsible to cause COVID-19 [1]. It is the respiratory disease with the pneumonia like qualities such that it is initially caused by the human contact by means of exotic fauna and eventually results person to person spread. However, this virus causes large international negative impact and is generally affected day to day to human lives to number of people through public gatherings, travel bans and the panic buying [13] [4]. Due to the emergence, the COVID disease affected people in worldwide with different problems. It has different negative impact in the people health and hence it interrupts the economy [14] [15]. Various countries implemented very effective interventions for controlling the viral spread of epidemic disease and also to minimize the effect of covid disease. The interventions may vary between the countries in such a way that the commonly used interventions are border closure, lockdown, banning of public events, travel banning, school closure, and social distancing [19] [3].

The covid pandemic is one of the serious problems in recent decades and it confronts the modern world due to the negative effects of human health. The impact of this disease is noticeable in the sensitive populations such that it includes people and elder with the chronic disease, like asthmatics. Hence, it becomes a major issue as it involves pharmaceutical industry, epidemiological experts, specialists in local authorities and diagnosis systems. To predict covid case numbers more appropriate becomes a major backbone for facilitating the usage of available resources in the medical center and to increase the management strategy for optimally managing the infected cases. In the recent decades, the deep learning and machine learning methods are gained as attractive field of research in different appliances, such as industry and academia [11] [9]. The deep learning method has gained more trends in the predictive modeling and data analysis in the recent decades and is considered as one of the advanced technologies. This is the leading and developing machine learning method in computer vision system. Accordingly, this method gained more enhancement in the classification performance for the large sized dataset [31]. Moreover, the deep learning gained attractive performance in various computer visions process in such a way that it learns the predictive features from the data specified in the large dataset [28]. In the recent decades, the deep learning methods are considered as the effective technique for decision making and prediction in various disciplines, like computer vision, natural language processing, gesture recognition, robotics, and health care system [32]. The deep learning is effectively useful in the prediction of virus spread and potentially helps to solve the impact of COVID-19 [29] [4].

The main contributions of this research are:

- Suggest the hybrid model (WPRO-Deep RNN) combine Poor and Rich Optimization and Water Cycle Algorithm to obtain the more accurate training with time-series data.
- Technical indicators extraction like CCI, SMI, RSI, EMA, CMO, Stochastic %K, and Aroon indicator are used to enhance the prediction.
- Use the technical features (oversampling) is increased the data augmentation and data dimension to achieve the obtain prediction.
- Achieved the better results comparing with competing models (GDCNN, DSPM, Deep NN and STNN).
- help to early prediction for covid-19 its help to support the medical decisions and refine the customize the medical resources.

1.1 Existing literature review

In this section, some of the conventional disease prediction schemes are explained along with their merits and demerits that motivate the researchers to design a new approach for accurate detection of COVID at early stage. Some of the established disease prediction schemes are investigated in this section. Babukarthik, R.G. *et al.* [1] introduced a Genetic deep learning Convolutional Neural Network (GDCNN) for covid-19 prediction. Here, features were taken from image to classify the covid cases and the normal images. The dataset used by this method was also used to detect the pneumonia disease. This method showed better nominal rate for the detection of disease in the unbalanced framework. It failed to use large sized database for obtaining higher classification accuracy. Shahid, F. *et al.* [2] developed a LSTM+GRU+Bi-LSTM model to find the covid cases. This method was employed to create three different categories, such as death cases, recovered and the confirmed cases. Rahman *et al.* [33] it achieved higher performance in finding the recovered and death cases. However, it failed to consider the features from the data. Abbasimehr, H. and Paki, R. [3] modeled a deep learning method+Bayesian optimization to detect covid-19. Deep learning incorporation schemes were done by multiple output forecast model that allowed to predict different time points. The optimization was employed to select optimal hyper parameters automatically in order to increase the performance of forecasting. It further required to acquire informative features and to integrate them with deep learning methods. Ayris, D. *et al.* [4] designed a Deep Sequential Prediction model (DSPM) for predicting the covid cases. This method was tested and trained with publicly present dataset. It accurately predicted disease spread and showed higher prediction performance. However, it failed to integrate NRM and DPSM features for refining prediction of developed scheme. It required number of training data for model to improve prediction.

Alazab, M. *et al.* [5] modeled a deep CNN for prediction of pandemic with real world dataset. It examined the chest images for finding the infected patients. This method was valuable for covid diagnosis with low cost. It showed higher accuracy performance. However, it did not analyze terrain effects, temperature, and humidity. Huang, C.J. *et al.* [6] developed deep network scheme for finding infected cases through the extraction of spatio temporal features. This method predicted the confirmed cases more accurately. It offered to increase the resource allocation in hospitals and also enable to make response and production plans during epidemic. It reduced the explosion of circulation from space correlation. De Oliveira, L.S. *et al.* [7] introduced an artificial neural network (ANN) to detect confirmed and the death cases of covid-19. The estimated values followed logical tendency such that it fit with respective curves. It obtained lower error rate and increased the performance of prediction. However, it further required to analyze the performance of network. Niu, Y.S. *et al.* [8] developed a spatio-temporal neural network (STNN) for covid prediction. It ensured flexibility and predictability performance. It offered more accurate performance in prediction and fitting in handling both the temporal and spatial data. It consumed less time for training and computing. However, it required large volume of training data. Zhang *et al.* [34].

It is more complex to determine when and where new infected cases are appeared and the governments are failed to analyze impact and scale of virus [4]. To forecast and accurately model the spread of recovered and infected cases of covid-19 cases and more important in such a way that to understand and make the decision-making strategy for arresting the viral disease spread is more complex [9]. Even though, RT-PCR test is assumed as clear diagnosis tool, limited supply and the strict need for the laboratory environment significantly delay the diagnosis of infected patients. Hence it is more complex to prevent the virus spread at the infected center of epidemic area [10]. Recently, the common way to find the infected cases is to undertake swab and to examine the material using chain reaction. A major issue is that swab is done only for the patients who have the symptoms and so the people with asymptomatic cases are not detected. Due to lack in the availability of data for the covid patients, the detailed study to report the solution for automatic detection with X-ray image is not available [12].

2. Time series data prediction

The time series [30] is a vector sequence represented as $z(r)$, here the term r denotes the elapsed time. For simplicity, let us assume only scalar sequences, but the considered techniques generalize readily to the series of vector. In theoretical, z may be the value that continuously varies with r, like temperature. In the practical way, for any physical system, z is sampled to generate discrete data series points that is spaced equally in time. In addition, the rate, where the samples produce the maximal resolution, but the high resolution is not at all has the best predictive power, thereby the appropriate result may be achieved by using only each p^{th} point in series.

The work in Neural Network (NN) has mainly concentrated on future development forecasting of the time series from the values of z till the current time. Thus, find the function $e:\hbar^{M} \rightarrow \hbar$ to determine the value of z at time $r+f$, from M time steps back from time r, hence,

$$z(r+f)=e(z(r),z(r-1),...,z(r-M+1)) \qquad (1)$$

where, $z(r+f)=e(u(r)$ such that $u(r)$ represents the N-ary vector of lagged z values. In general, f is the one, thereby e forecasts the next z value.

3. Deep RNN for Time series data prediction

Figure 1 portrays Deep RNN's architecture. Deep RNN [27] is the artificial neural network (ANN) with number of hidden layers and record forceful time series of the network by the directional linking of the nodes at hidden layer. It is entirely diverse from the feed forward network, as it records input sequences at various time states using the feedback and the feed forward connections amongst internal processing layers. The Deep RNN is more robust transformation on the technical features and thereby it increases the ability of prediction. This network is composed with different non-linear layers with respect to the time. However, the position of hidden layer is utility in position of hidden layer previously available and so this classifier is vertical expansion over time. Here, each data of sequence is effectively computed and present result is utilized to predict next result. It used information in random order and used historical data at some number of stages. Anyhow, it maps input sequence to order of hidden states and then this is used to map by output arrangements by means of feature learning. Input takes by the Deep RNN classifier is the data augmented result H that is passed to the input layer. Let us consider the input passed to the input layer at time c as H_c, and hidden layer state at time c as J_c. The input layer has $[1\times 11]$ dimension, whereas hidden state has $[1\times]$ dimension. The process of hidden state layer is given as,

$$J_c = y\left(H\,X^{JH} + J_{c-1}X^{JJ}\right) \qquad (2)$$

where, y denotes non-linear function, like tanh and the ReLu functions, J_{c-1} indicates previous hidden layer state, X denotes the weight factor, X^{JH} represents the weight among input and the hidden layer, X^{JJ} indicates the weight amongst hidden layers, and H represents the input. However, input processing unit are recorded in internal state by means of activation function. Accordingly, output computed at output layer is given as,

$$g_c = y\left(J_c X^{gJ}\right) \qquad (3)$$

where, g denotes the output, y represents the function, X^{gJ} signifies the weight aongst hidden and the output layer. Moreover, the dimension of output layer is given as $[1\times 8]$. The optimal weight value is computed by proposed WPRO algorithm through training process.

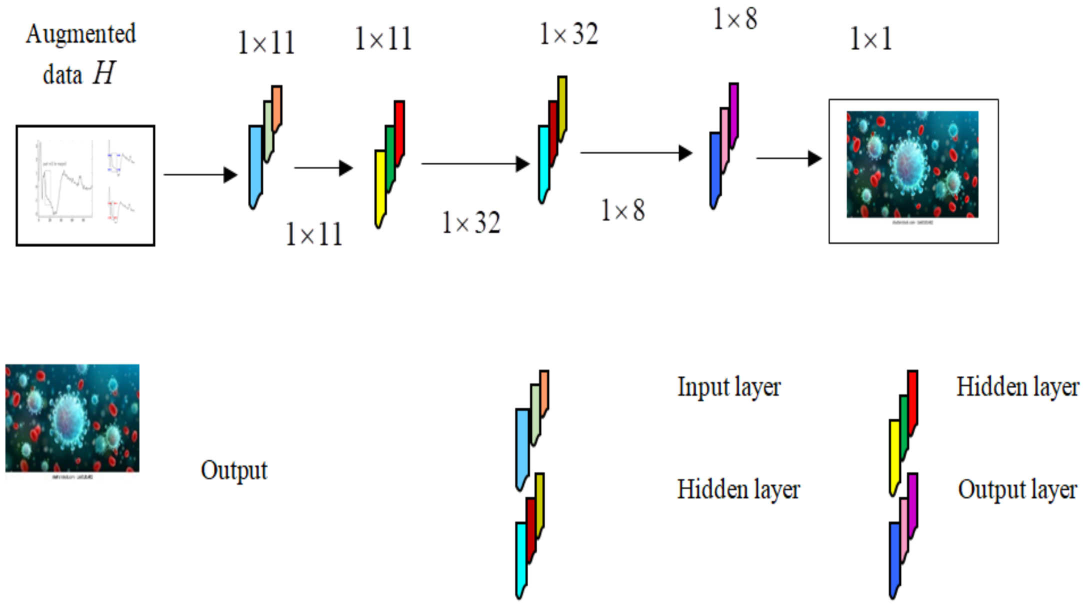


Figure 1. Structure of Deep RNN

4. Optimization driven Deep Recurrent Neural Network

To predict COVID epidemic is vital task for accurately monitoring and to reveal the death number, infected cases and recovered cases to acquire knowledge concerning trend of this disease. With the data augmented output, process of covid prediction is done based on Deep RNN classifier. Deep learning classifier gained significant improvement to deal with times-series data in various applications. Deep RNN utility modeling noun of data, each style is supposed to be based on prior data. Moreover, Deep RNN categorizer is a lot efficient for dealing by this type of data with raised parameter. Categorizer architecture and its training process hired to find of classifier's best weight is elucidated in the below section.

4.1 Procedure of the proposed WPRO algorithm

Deep RNN's training procedure is composed with proposed optimization algorithm to find ideal weight value through optimal position of optimization parameter. The proposed WPRO algorithm is derived by integration of PRO [21] and WCA [22], respectively. PRO algorithm is encouraged by dual groups of population, like poor as well as rich to attain wealth and increase their financial circumstances. Moreover, rich acquire the wealth and try to improve the class gap, whereas poor try for achieving wealth and decrease gap in class. It is operated based on the social phenomenon. Each poor member population attempts to increase status using learning from rich. The concept of WCA depends on water cycle observation and flow of stream and rivers towards sea. Initial population of water cycle is observed from raindrops. The behavioral parameters of both the optimization algorithm are integrated to achieve best global solution by eradicating local optimum. With hybrid optimization, the efficiency of the proposed method is increased through accurate training of classifier and so it achieved optimal performance.

Solution encoding: This is solution vector representation which shows computation of best solution for value factor at each iteration of the optimization with the fitness measure. The dimension of result is denoted by, $L = 1\times X$, the optimal value for X is calculated by optimization algorithm. Suggested method WPRO approach is illustrated following:

i) Initialization*:* Initial population is created arbitrarily with equal distribution based on lower and upper bound parameters. Sort initial population in ascending order by considering objective function. Population of optimization algorithm is collection of dual subpopulations, namely rich and poor in search space (L).

Accordingly, population is indicated as,

$$P_{op} = P_{poor} + P_{rich} \quad (4)$$

where, P_{op} indicates main population, P_{poor} denotes poor population, and P_{rich} signifies rich population.

ii) Fitness function's computation: Fitness finds solution for optimization problem using error function and is given as,

$$F = \frac{1}{\kappa}\sum_{\tau=1}^{\kappa}\left[O_\tau - g_\tau\right]^2 \quad (5)$$

where, F indicates fitness measure, O indicates output of target, κ denotes total number of training samples.

iii) Rich population's update position: For every iteration, member position in rich population is given as,

$$\overrightarrow{B^{new}_{rich,m}} = \overrightarrow{B^{old}_{rich,m}} + a\left[\overrightarrow{B^{old}_{rich,m}} - \overrightarrow{B^{old}_{poor,best}}\right] \quad (6)$$

$$\overrightarrow{B^{new}_{rich,m}} = \overrightarrow{B^{old}_{rich,m}}(1+a) - a\,\overrightarrow{B^{old}_{poor,best}} \quad (7)$$

Standard equation of WCA is denoted as,

$$B^{m+1}_{river} = B^{m}_{river} + rand\,A\left(B^{m}_{sea} - B^{m}_{river}\right) \quad (8)$$

$$B^{m+1}_{river} = B^{m}_{river}(1 - rand\,A) + rand\,A B^{m}_{sea} \quad (9)$$

$$B^{m}_{river} = \frac{B^{m+1}_{river} - rand\,A B^{m}_{sea}}{1 - rand\,A} \quad (10)$$

Let us assume $B^{m}_{river} = \overrightarrow{B^{new}_{rich,m}}$ and so the above is re-written in terms of PRO as,

$$\overrightarrow{B^{old}_{rich,m}} = \frac{\overrightarrow{B^{new}_{rich,m}} - rand\,A B^{m}_{sea}}{1 - rand\,A} \quad (11)$$

By substitution of Eq. (17) in Eq. (11),

$$\overrightarrow{B^{new}_{rich,m}} = \frac{\overrightarrow{B^{new}_{rich,m}} - rand\,A B^{m}_{sea}}{1 - rand\,A}(1+a) - a\,\overrightarrow{B^{old}_{poor,best}} \quad (12)$$

$$\overrightarrow{B^{new}_{rich,m}} = \frac{\overrightarrow{B^{new}_{rich,m}}}{1 - rand\,A}(1+a) - \frac{rand\,A B^{m}_{sea}}{1 - rand\,A}(1+a) - a\,\overrightarrow{B^{old}_{poor,best}} \quad (13)$$

$$\overrightarrow{B^{new}_{rich,m}} - \frac{\overrightarrow{B^{new}_{rich,m}}}{1 - rand\,A}(1+a) = -\frac{rand\,A B^{m}_{sea}}{1 - rand\,A}(1+a) - a\,\overrightarrow{B^{old}_{poor,best}} \quad (14)$$

$$\overrightarrow{B^{new}_{rich,m}}\left[1 - \frac{1+a}{1 - rand\,A}\right] = -\frac{rand\,A B^{m}_{sea}}{1 - rand\,A}(1+a) - a\,\overrightarrow{B^{old}_{poor,best}} \quad (15)$$

$$\overrightarrow{B_{rich,m}^{new}}\left[\frac{1-rand\,A-1-a}{1-rand\,A}\right]=-\frac{rand\,A B_{sea}^{m}}{1-rand\,A}(1+a)-a\,\overrightarrow{B_{poor,best}^{old}} \tag{16}$$

$$\overrightarrow{B_{rich,m}^{new}}\left[-\frac{rand\,A+a}{1-rand\,A}\right]=-\left[\frac{rand\,A B_{sea}^{m}(1+a)}{1-rand\,A}+a\,\overrightarrow{B_{poor,best}^{old}}\right] \tag{17}$$

$$\overrightarrow{B_{rich,m}^{new}}=\frac{1-rand\,A}{rand\,A+a}\left[\frac{rand\,A B_{sea}^{m}(1+a)}{1-rand\,A}+a\,\overrightarrow{B_{poor,best}^{old}}\right] \tag{18}$$

where, $\overrightarrow{B_{rich,m}^{new}}$ denotes new value of m^{th} place in rich population, a signifies parameter of class gap that has the value of $[0,1]$, $\overrightarrow{B_{rich,m}^{old}}$ indicates current value of m^{th} position, $\overrightarrow{B_{poor,best}^{old}}$ represents the current location of best member, $rand$ represents arbitrary number that ranges between interval of $[0,1]$, and A is the parameter that contains the value of $[1,2]$.

iv) Compute mutation: The mutation for poor and rich population is calculated according to mutation probability which is represented below,

$$\overrightarrow{B_{rich,m}^{new}}=\overrightarrow{B_{rich,m}^{new}}+s \tag{19}$$

$$\overrightarrow{B_{poor,m}^{new}}=\overrightarrow{B_{poor,m}^{new}}+s \tag{20}$$

where, $\overrightarrow{B_{rich,m}^{new}}$ denotes the mutation value of rich population, $\overrightarrow{B_{poor,m}^{new}}$ indicates mutation value of poor population, and s represents normal distribution of mean value as 0 and variance of 1, respectively.

v) Poor population's update equation: Here, poor population position is updated by below equation as,

$$\overrightarrow{B_{poor,m}^{new}}=\overrightarrow{B_{poor,m}^{old}}+\left[v\times K-\overrightarrow{B_{poor,m}^{old}}\right] \tag{21}$$

where, v indicates improvement factor of pattern that ranges in value of $[0,1]$, $\overrightarrow{B_{poor,m}^{new}}$ indicates new value of m^{th} position, and $\overrightarrow{B_{poor,m}^{old}}$ represents current value of m^{th} position.

$$K=\frac{\overrightarrow{B_{rich,best}^{old}}+\overrightarrow{B_{rich,mean}^{old}}+\overrightarrow{B_{rich,worst}^{old}}}{3} \tag{22}$$

Here, $\overrightarrow{B_{rich,best}^{old}}$ denotes best member of rich population, $\overrightarrow{B_{rich,mean}^{old}}$ represents average position, and $\overrightarrow{B_{rich,worst}^{old}}$ denotes location of worst member.

vi) Re-evaluating possibility: Fitness for every solution is calculated and ideal objective value is accepted as optimal value.

vii) End: Previous stages are iterated till global optimal result is found. Pseudo code of proposed WPRO_Deep RNN is shown in Algorithm 1.

Algorithm 1. Algorithm outline for WPRO- built on Deep RNN

stages	WPRO- built on Deep RNN
1	In **:** A , a , B_{sea}^{m}
2	Out **:** $\overrightarrow{B_{rich,m}^{new}}$
3	Begin
4	Population initialization
5	Sort the main population into rich and poor
6	for the rich population
7	Update the location using Equation (30)
8	Compute mutation
9	end for
10	for the poor population
11	Update the location using Equation (33)
12	Compute mutation
13	end for
14	Return best rich
15	Terminate
16	End

5. Proposed system architecture

COVID is the pandemic disease due to the faster transmissibility and seriousness. Hence, accurate detection at early stage is more vital in the medical healthcare system. The deep learning model shows an important part in evaluation and prediction of large outbreak data. This helps to diagnose the disease at early stage. Therefore, this research developed a Deep RNN classifier for detection of disease using time-series data. Accordingly, proposed method involves different phase, like technical indicators extraction, augmentation of data, and prediction of disease. Input data developed from dataset is passed to extraction phase of technical metrics, as technical indicators like CCI, SMI, RSI, EMA, CMO, Stochastic %K, and Aroon indicator are effectively captured from this type of data. Thereafter, dimension related to technical attributes is increased using data enhancement method through upsampling technique. This last stage of suggested disease estimation, where Deep RNN is employed to find the number of death cases, recovered cases, infected (confirmed) cases, percentage of infection, percentage of death, and percentage of recovered cases. Schematic representation of proposed method is presented in figure 2.

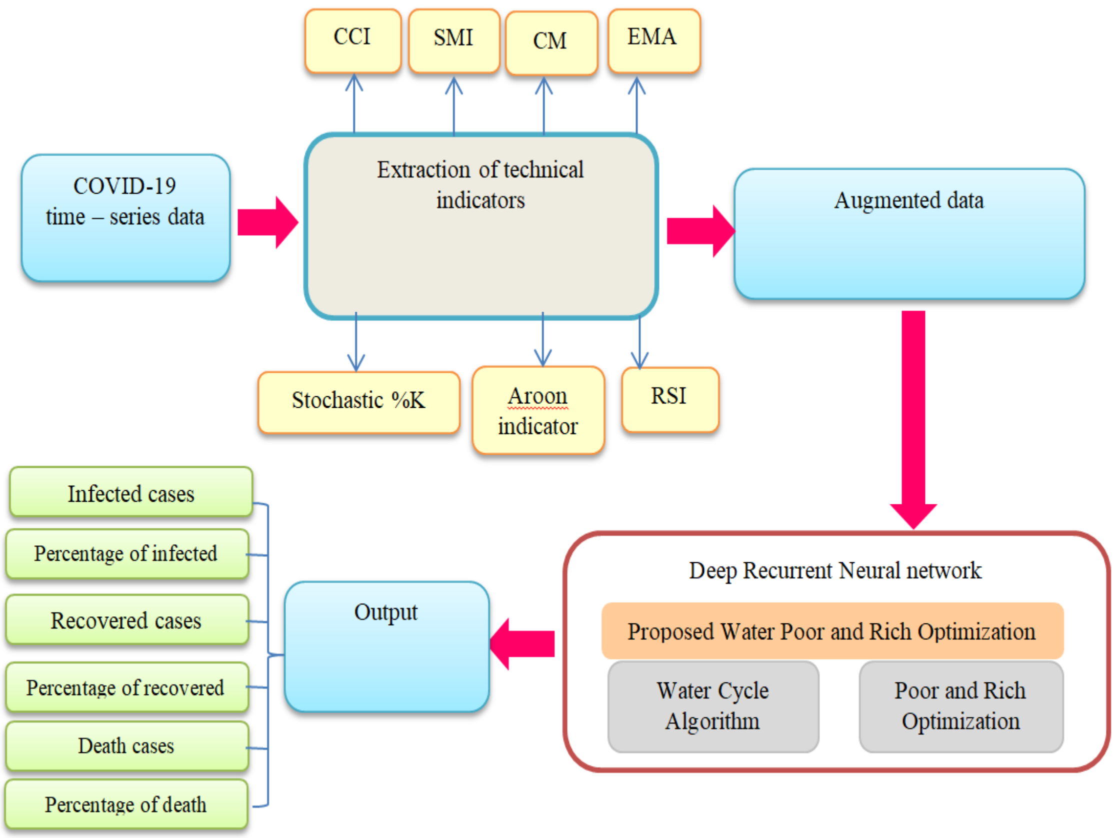


Figure 2. Schematic presentation of proposed WPRO-based Deep RNN for prediction of COVID

5.1 Input time-series data acquisition

Data collected at consistent time breaks is termed as the time series data such that each data point is spaced similarly in terms of time. Data of time-series is utilized to forecast upcoming patterns or trends with historical dataset using technical features. To predict COVID cases, it is more effective as the input data contains temporal components, so time series data is utilized to achieve forecast of COVID. Consider dataset D having m value of sequential data C that is expressed following:

$$D = \{C_1, C_2, ..., C_i, ..., C_w\} \quad (23)$$

where, C_i denotes the input sequential data situated as i^{th} pointer of data, w implies total data number, and D denotes dataset.

5.2 Technical indicators extraction

Time-series data C_i selected from dataset is exposed to practical parameters derivation module to acquire practical attributes from the data. The technical indicators use technical rules that rely on the historical data over specific time period to predict the future cases using mathematical functions. However, the extraction of technical indicators can be made more effective only with the time-series data, where the historical data is clearly defined over certain time period [23]. Some of the technical features captured from C_i are explained as follows:

CCI: Oscillator to find if data is high or low. Anyhow, it represents relationship amongst data which is computed below,

$$T_1 = \frac{\left(b^n - H_w\left(b^n\right)\right)}{0.05\sum_{j=1}^{w}\left|b_{n-j+1} - H_w\left(b^n\right)\right|/w} \tag{24}$$

where, b^n indicates total of high data, low data, and the close data value on the day n that is $b^n = U_n + V_n + S_n$, $H_w\left(b^n\right)$ represents the simple moving average (SMA) of b^n values measured over w days, and w indicates total number of days.

SMI: Calculated based on high and low data values and mathematically formulated as,

$$T_2 = 100 \times \left(\frac{2U_n}{V_n - S_n}\right) \tag{25}$$

where, U_n denotes number of closest data on day n, V_n indicates number of high data on day n, and S_n implies the low data on day n.

RSI: It compares the recent low with the recent high data to show the strength or weakness of data value and is expressed as,

$$T_3 = 100 - \frac{100}{\left(1 + H_w\left(k^+\right)/H_w\left(k^-\right)\right)} \tag{26}$$

where, H denotes SMA, k^+ represents positive directional movement, and k^- implies negative directional movement.

$$k^+ = \max\left(U_n - U_{n-1}, 0\right) \tag{27}$$

$$k^- = \min\left(U_n - U_{n-1}, 0\right) \tag{28}$$

EMA: It is moving average class that considers weight value for computation of future prediction and is calculated as,

$$T_4 = \sum_{j=0}^{w-1} W_j U_{n-j} \tag{29}$$

where, W denotes the weight.

CMO: Calculated according to difference positive and negative directional movement which is given below,

$$T_5 = 100 \times \frac{H_w\left(k^+\right) - H_w\left(k^-\right)}{H_w\left(k^+\right) + H_w\left(k^-\right)} \tag{30}$$

Stochastic %K: This technical feature is employed to verify the number of data over period of w days. However, it is represented as,

$$T_6 = 100 \times \frac{U_n - V_n}{V_n - S_n} \tag{31}$$

Aroon indicator: It is used to find the trend changes in the data with respect to the time and is computed as,

$$T_7 = AR_{up} - AR_{down} \tag{32}$$

$$AR_{up} = 100 \times \frac{25 - H_{25}}{25} \quad (33)$$

$$AR_{down} = 100 \times \frac{25 - L_{25}}{25} \quad (34)$$

where, H_{25} denotes high data for 25 days period, and L_{25} indicates low data for 25 days period. Accordingly, time-series data C_i extract technical indicators and this is represented below,

$$T = \{T_b\}; b \in \{1,...,7\} \quad (35)$$

where, T indicates extracted technical features with the dimension of $[277 \times 12]$.

5.3 Augmentation of data by oversampling method

Technical features developed from input data are given to augmentation of data phase, in which dimensionality of feature is enlarged to improve the efficiency of prediction. Augmentation of data is process employed for enhancing data dimensions using up sampling technique. Upsampling is the mechanism to create many numbers of training samples by partitioning data built on class categories. As, its time-series data, there exists no class value. Hence data augmentation is carried out by taking minimum and maximum value of training samples as threshold value for generating sample numbers. Hence, augmentation of data results showed increase in dimension of data through the generation of data samples. Size of data is increased by data augmentation that additionally supports the classifier for generating reliable prediction result. Anyhow, H is the augmented result with the dimension $[100277 \times 12]$.

5.4 COVID prediction by proposed WPRO-based Deep RNN

Deep RNN classifier predict the disease that is trained using proposed WPRO algorithm. Input of Deep RNN classifier is data augmented output. Accordingly, proposed WPRO algorithm is the integration of Water Cycle Algorithm (WCA) and Poor and Rich Optimization (PRO), respectively. The proposed WPRO-based Deep RNN predicts the infected, recovered, deaths and their percentages.

6. Time series data description

The experimentation of developed method is done using Novel Coronavirus (COVID-19) Cases Data [24]. This dataset is the part of COVID-19 pandemic, gathered for system science and Engineering by Johns Hopkins University center from different sources. This dataset contains various fields, such as country/region, province/state, last update, suspected, deaths, recovered and confirmed fields. Total number of rows and attributes number present in death case, confirmed case, death percentage, and confirmed percentage is 277 and 513. The recovered cases and the recovered percentage have the rows and columns as 262 and 513, respectively.

6.1 Evaluation metrics

The performance of developed method is analyzed by considering MSE and RMSE.

MSE: It represents the average of squared difference between actual and the predicted value and is specified in equation (5).

RMSE: It is the square root of signify squared loss faction that measures the residuals dispersion measure. It is represented following:

$$RMSE = \sqrt{\frac{1}{\kappa} \sum_{\tau=1}^{\kappa} [O_\tau - g_\tau]^2} \quad (36)$$

7. Predictive graph for the different methods

Figure 3 shows the estimation evaluation implemented with the confirmed cases. By given Azerbaijan country, the original confirmed cases is 334715, whereas the prediction result generated by the existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN is 167231, 58905, 54919, 66545, and 274508 such that error between original according to predicted confirmed cases by GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN is 167484, 275810, 279796, 268170, and 60207.

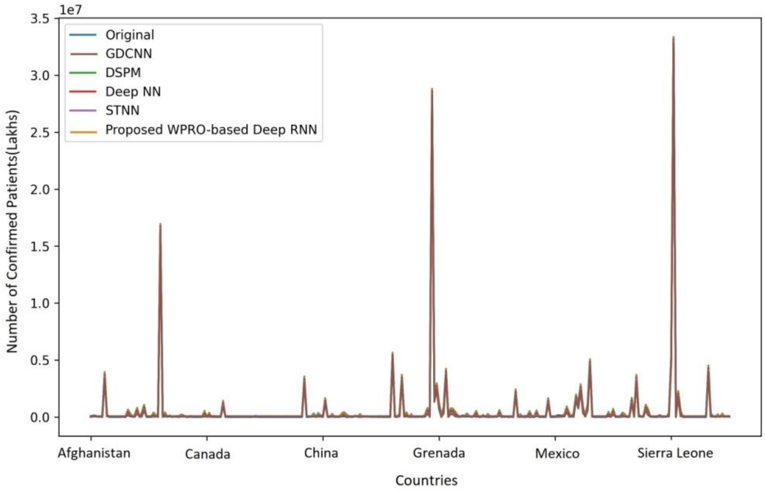


Figure 3. Analysis with confirmed cases

Figure 4 depicts analysis of death case prediction. By considering Sierra Leone country, the original death cases are 82, whereas prediction result generated by the existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN is 605, 3822, 2116, 1282, 4040 as such the error between original according to predicted recovered cases by GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN is 523, 3740, 2034, 1200, and 3958.

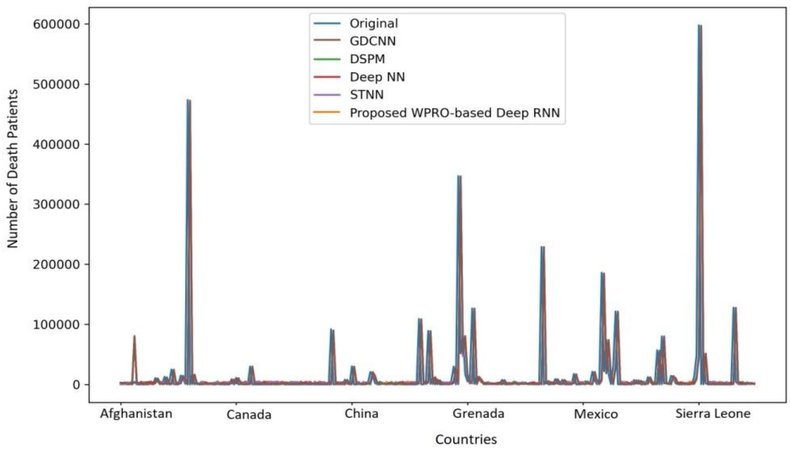


Figure 4. Assessment by death cases

Figure 5 illustrates forecast analysis of recovered cases. By considering Mexico country, the original cases recovered is 1939596, but prediction result generated by the existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN is 1724715, 1603375, 1464678, 1329813, and 1839596 as such the error between original according to predicted recovered cases by GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN is 214881, 336221, 474918, 609783, and 100000.

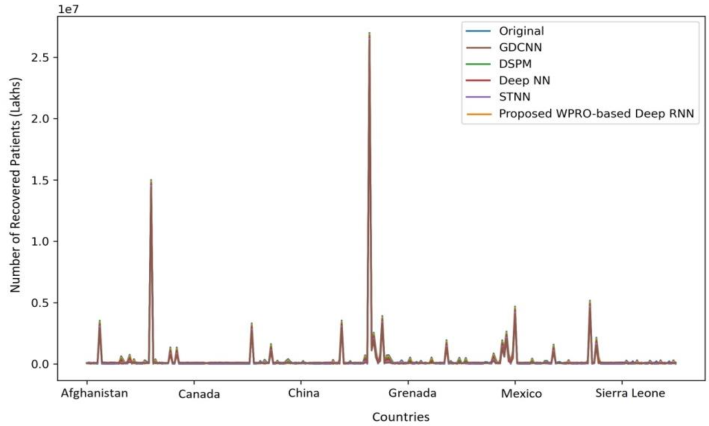


Figure 5. Analysis based on recovered cases

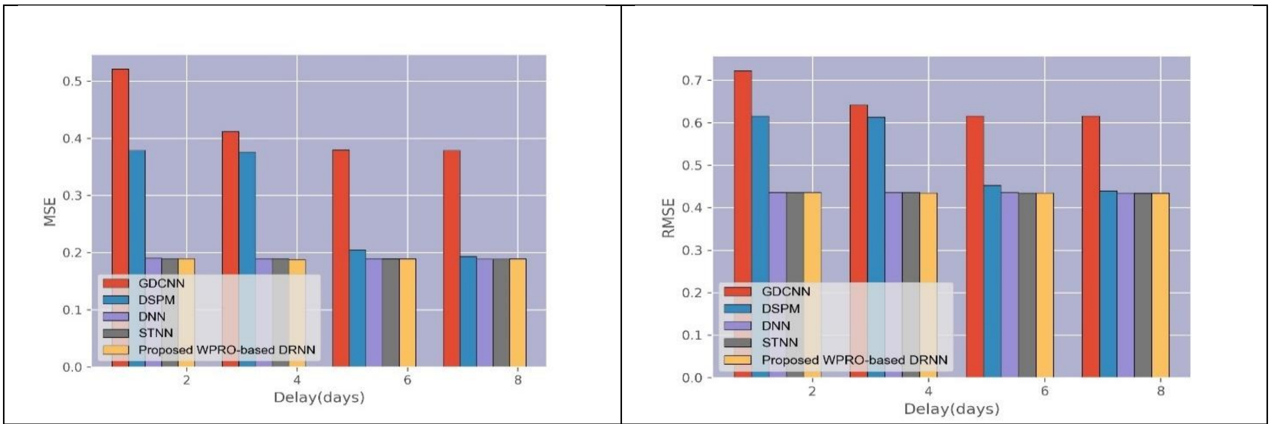
8. Comparative validation

The efficiency enhancement for developing proposed approach analyzing by present methods, such as Genetic Deep Learning Convolutional Neural Network (GDCNN) [1], Deep Sequential Prediction Model (DSPM) [4], Deep Neural Network [6], and Spatio-Temporal Neural Network (STNN) [8], respectively. Here, the delay is considered as days.

8.1 Analysis based on death case

Figure 6 shows assessment of proposed technique by considering the death cases. MSE analysis is depicted in figure 6 a). When delay (days) is taken as 4, the MSE calculated by present methods, such as GDCNN, DSPM, Deep NN, and STNN are 0.4113, 0.3758, 0.1894, and 0.1890, while proposed WPRO-based Deep RNN achieved MSE as 0.1879, accordingly. When delay=6, the MSE obtained by existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.3790, 0.2044, 0.1893, 0.1888, and 0.1887.

Figure 6 b) illustrate RMSE analysis. For delay is taken as 4, RMSE calculated by present approaches, like GDCNN, DSPM, Deep NN, and STNN are 0.641, 0.613, 0.435, and 0.435, while suggested WPRO_Deep RNN obtained RMSE its 0.434, accordingly. When delay is 6, RMSE found by considering existing GDCNN, DSPM, Deep NN, STNN, and proposed method are 0.616, 0.452, 0.435, 0.434, and 0.434.

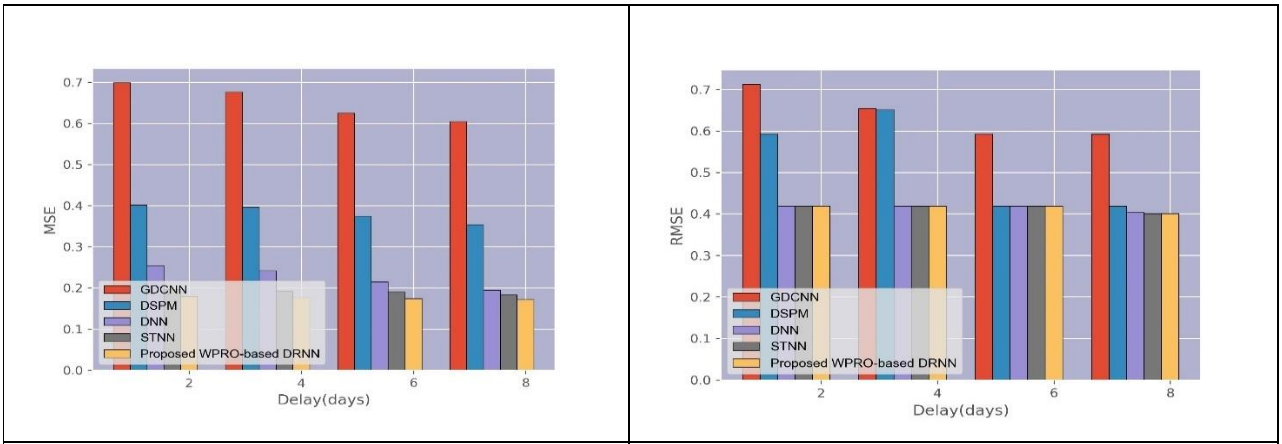


a)	b)

Figure 6. Analysis with death cases, a) MSE, b) RMSE

Figure 7 depicts the assessment of the developed model by considering confirmed cases. The figure 7 a explained the MSE analysis. In case delay is 4, MSE found by existing GDCNN, DSPM, Deep NN, and STNN are 0.6760, 0.3954, 0.2413, and 0.1924, while the developed WPRO- built on Deep RNN obtained the MSE its 0.1765. For delay 6, MSE obtained using existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.6041, 0.3541, 0.1941, 0.1824, and 0.1714.

Figure 7 b) illustrates RMSE assessment. For delay=4, RMSE computed using GDCNN, DSPM, Deep NN, and STNN are 0.654, 0.651, 0.419, and 0.419, as proposed approach found the RMSE as 0.419, respectively. For delay 6, the RMSE obtained by existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.592, 0.419, 0.404, 0.401, and 0.4.



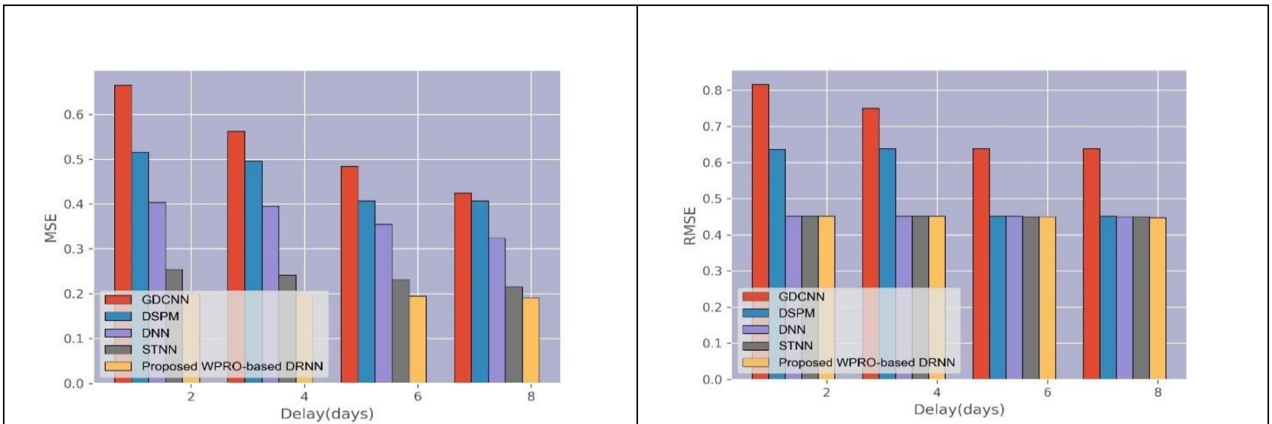
a)	b)

Figure 7. Assessment based on confirmed cases, a) MSE, b) RMSE

8.3 Assessment based on recovered cases

Figure 8 shows assessment of developed approach for recovered cases. MSE analysis is given in figure 8 a). For delay 4, MSE achieved by existing GDCNN, DSPM, Deep NN, and STNN are 0.5618, 0.4954, 0.3954, and 0.2413, while proposed WPRO-based Deep RNN achieved MSE as 0.1965. For delay as 8, the MSE obtained by existing GDCNN, DSPM, Deep NN, STNN, while suggested WPRO- built on Deep RNN are 0.4241, 0.4063, 0.3241, 0.2143, and 0.1901.

Figure 8 b) illustrate RMSE assessment. For delay=4, RMSE computed by GDCNN, DSPM, Deep NN, and STNN are 0.750, 0.637, 0.451, and 0.451, as proposed model attained the RMSE as 0.451, respectively. While considering delay as 6, RMSE obtained by existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.637, 0.451, 0.450, 0.450, and 0.447. Proposed technique had better performance in acquiring lower error value.



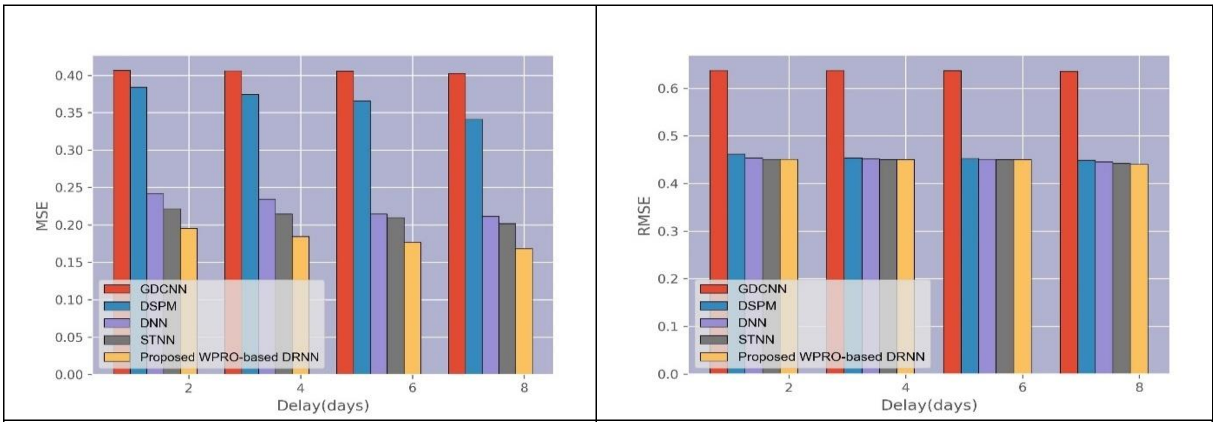
a)	b)

Figure 8. Analysis with cases of recovery, a) MSE, b) RMSE

8.4 Assessment with percentage of death cases

Figure 9 depicts analysis made by proposed technique by percentage of death cases. the figure 9 a explained the MSE analysis. in case delay is considered as 6, MSE value achieved by the traditional GDCNN, DSPM, Deep NN, and STNN is 0.4057, 0.3654, 0.2143, and 0.2095, as proposed model attained MSE as 0.1765, respectively. If delay is 8, MSE found using present GDCNN, DSPM, Deep NN, STNN, and WPRO with Deep RNN is 0.4241, 0.4063, 0.3241, 0.2143, and 0.1901.

Figure 9 b) depicts RMSE assessment. For delay=4, RMSE calculated using existing GDCNN, DSPM, Deep NN, and STNN is 0.637, 0.453, 0.451, and 0.450, as the proposed WPRO-based Deep RNN attained lower value of RMSE as 0.450, respectively. At 8 delay, the RMSE achieved by existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.635, 0.448, 0.445, 0.442, and 0.440.



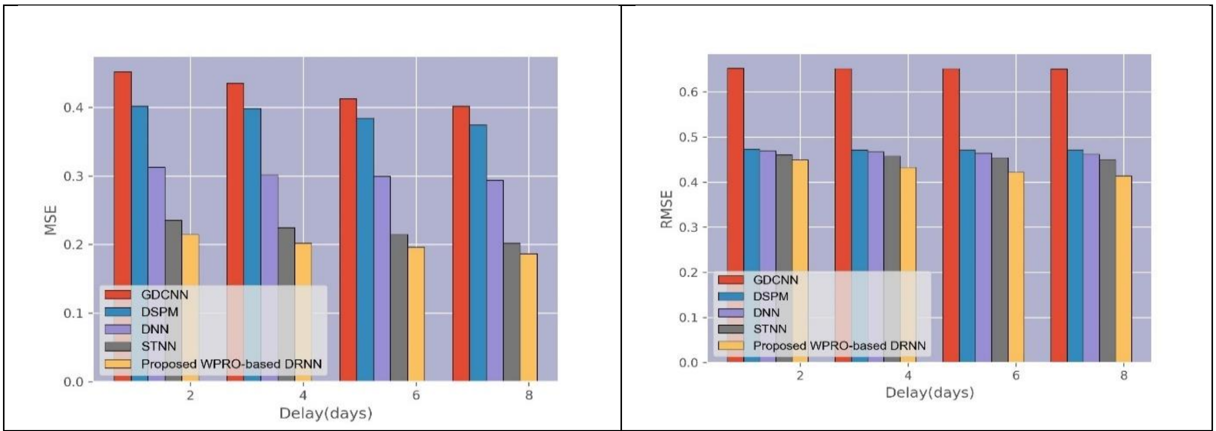
a)	b)

Figure 9. Assessment based on death cases percentage, a) MSE, b) RMSE

8.5 Assessment with percentage of confirmed cases

Figure 10 shows assessment made by proposed approach by percentage of confirmed cases. the figure 10 a explained the MSE analysis. For delay 6, MSE obtained using existing GDCNN, DSPM, Deep NN, STNN, and suggested WPRO-built on Deep RNN are 0.4126, 0.3841, 0.2997, 0.2142, and 0.1954. If delay is 8, MSE value achieved by the traditional GDCNN, DSPM, Deep NN, and STNN are 0.4013, 0.3741, 0.2934, and 0.2014, as proposed model attained the MSE value as 0.1865, respectively.

Figure 10 b) illustrate RMSE assessment. For delay of 4, RMSE calculated by existing GDCNN, DSPM, Deep NN, and STNN are 0.6514, 0.4712, 0.4665, and 0.4574, as proposed WPRO-based Deep RNN achieved lower value of RMSE as 0.4322, respectively. At 8 delay, the RMSE achieved by existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.6504, 0.4701, 0.4612, 0.4501, and 0.4126.



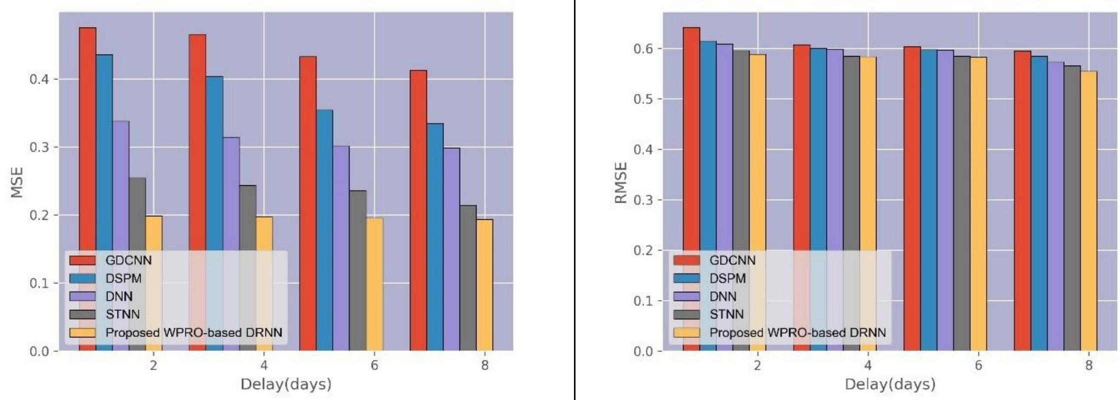
a)	b)

Figure 10. Assessment based on confirmed cases percentage, a) MSE, b) RMSE

8.6 Assessment with percentage of recovered cases

Figure 11 shows the assessment of proposed approach by percentage of recovered cases. the figure 11 a explained the MSE analysis. If delay is 6, MSE found using present GDCNN, DSPM, Deep NN, STNN, and suggested WPRO-built on Deep RNN are 0.4325, 0.3541, 0.3013, 0.2354, and 0.1954. If delay is 8, MSE value achieved by the traditional GDCNN, DSPM, Deep NN, and STNN are 0.4126, 0.3345, 0.2984, and 0.2143, as the proposed model achieved the MSE value as 0.1933, respectively.

Figure 11 b) portrays RMSE assessment. For delay of 4, RMSE calculated by existing GDCNN, DSPM, Deep NN, and STNN is 0.607, 0.6, 0.598, and 0.584, as suggested WPRO-built on Deep RNN attained reduced value of RMSE as 0.583, respectively. At 8th delay, the RMSE achieved by existing GDCNN, DSPM, Deep NN, STNN, and proposed WPRO-based Deep RNN are 0.594, 0.584, 0.573, 0.565, and 0.554.



a) b)

Figure 11. Assessment based on recovered cases percentage, a) MSE, b) RMSE

8.7. Discussive results

Comparative analysis with discussion for developed scheme is given in Table 1. Below table represent achievement obtained by suggested technique by given delay value as 8. For the cases of death, MSE found by proposed method is 0.1887, as existing methods achieved higher MSE value. By considering the confirmed cases, RMSE and MSE value found by proposed approach is 0.400 and 0.1714, respectively. RMSE and the MSE value acquired by proposed model for the recovered cases is 0.447 and 0.1901. RMSE and MSE values calculated by considering percentage of death cases is 0.440 and 0.1685.

Table 1: Comparative analysis with discussive results

Metrics/ Methods	**Cases of Death**		**Confirmed Cases**		**Recovered Cases**		**Death cases in %**		**confirmed cases in %**		**Recovered cases in %**	
	MSE	*RMSE*	*MSE*	*RMSE*	*MSE*	*RMSE*	*MSE*	*RMSE*	*MSE*	*RMSE*	*MSE*	*RMSE*
GDCNN	0.3785	0.615	0.6041	0.592	0.4241	0.637	0.4024	0.635	0.4013	0.6504	0.4126	0.594
DSPM	0.1929	0.439	0.3541	0.419	0.4063	0.451	0.3413	0.448	0.3741	0.4701	0.3345	0.584
Deep NN	0.1885	0.434	0.1941	0.404	0.3241	0.45	0.2113	0.445	0.2934	0.4612	0.2984	0.573
STNN	0.1882	0.434	0.1824	0.401	0.2143	0.45	0.2014	0.442	0.2014	0.4501	0.2143	0.565
Proposed WPR-based Deep RNN	0.1887	0.433	0.1714	0.4	0.1901	0.447	0.1685	0.44	0.1865	0.4126	0.1933	0.554

9. Conclusion

In this research, an efficient method is developed to predict COVID disease at starting stage by proposed WPRO with Deep RNN. Deep learning predictor is efficient in processing the high dimensional data such that the data resulted by data augmentation process is high dimensional data. Input data is exposed to the extraction of technical indicators stage for extracting technical features. Accordingly, extracted features are passed to augmentation of data phase, as oversampling method is employed to produce much number of training samples as it increases data dimensionality. Disease prediction is made by Deep RNN classifier that is tuned by proposed WPRO. Moreover, developed approach shows higher efficiency in the prediction of disease using deep learning classifier with time-series data. However, the proposed WPRO-based Deep RNN attained minimum RMSE and MSE as 0.4 and 0.1714 for confirmed cases, lower RMSE and MSE of 0.433 and 0.1887 for cases of death. Moreover, developed method attained minimal RMSE and MSE of 0.447 and 0.1901 for the recovered cases. Future research would be considering other deep learning classifier to further improve efficiency of prediction.

Reference

[1] R. G. Babukarthik, V. A. K. Adiga, G. Sambasivam, D. Chandramohan, and J. Amudhavel, "Prediction of COVID-19 using genetic deep learning convolutional neural network (GDCNN)," *IEEE Access*, vol. 8, pp. 177647–177666, 2020.

[2] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons Fractals*, vol. 140, p. 110212, 2020.

[3] H. Abbasimehr and R. Paki, "Prediction of COVID-19 confirmed cases combining deep learning methods and Bayesian optimization," *Chaos, Solitons Fractals*, vol. 142, p. 110511, 2021.

[4] D. Ayris *et al.*, "Deep learning models for early detection and prediction of the spread of Novel Coronavirus (COVID-19)," *arXiv preprint*, arXiv: 2008.01170, 2020.

[5] M. Alazab, A. Awajan, A. Mesleh, A. Abraham, V. Jatana, and S. Alhyari, "COVID-19 prediction and detection using deep learning," *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 12, pp. 168–181, 2020.

[6] C.-J. Huang, Y. Shen, P.-H. Kuo, and Y.-H. Chen, "Novel spatiotemporal feature extraction parallel deep neural network for forecasting confirmed cases of coronavirus disease 2019," *Socioecon. Plann. Sci.*, p. 100976, 2020.

[7] L. S. De Oliveira, S. B. Gruetzmacher, and J. P. Teixeira, "COVID-19 time series prediction," *Proc. Comput. Sci.*, vol. 181, pp. 973–980, 2021.

[8] Y.-S. Niu, W. Ding, J. Hu, W. Xu, and S. Canu, "Spatio-temporal neural network for fitting and forecasting COVID-19," *arXiv preprint*, arXiv:2103.11860, 2021.

[9] Zeroual, F. Harrou, A. Dairi, and Y. Sun, "Deep learning methods for forecasting COVID-19 time-Series data: A Comparative study," *Chaos, Solitons Fractals*, vol. 140, p. 110121, 2020.

[10] Zheng *et al.*, "Deep learning-based detection for COVID-19 from chest CT using weak label," *medRxiv*, 2020.

[11] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," *Comput. Methods Programs Biomed*, vol. 196, p. 105608, 2020.

[12] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-COVID: Predicting COVID-19 from chest X-ray images using deep transfer learning," *Med. Image Anal.*, vol. 65, p. 101794, 2020.

[13] S. Kumar, S. Mishra, and S. K. Singh, "Deep transfer learning-based COVID-19 prediction using chest X-rays," *medRxiv*, 2020.

[14] M. A. Achterberg *et al.*, "Comparing the accuracy of several network-based COVID-19 prediction algorithms," *Int. J. Forecast.*, 2020.

[15] Rajesh, H. Pai, V. Roy, S. Samanta, and S. Ghosh, "CoVID-19 prediction for India from the existing data and SIR (D) model study," *medRxiv*, 2020.

[16] M. N. Sohail, R. Jiadong, M. M. Uba, and M. Irshad, "A comprehensive looks at data mining techniques contributing to medical data growth: A survey of researcher reviews," in *Recent Developments in Intelligent Computing, Communication and Devices*. Singapore: Springer, 2019, pp. 21–26.

[17] C. W. Song, H. Jung, and K. Chung, "Development of a medical big-data mining process using topic modeling," *Cluster Comput.*, vol. 22, no. 1, pp. 1949–1958, 2019.

[18] M. Nikooghadam, A. Ghazikhani, and M. Saeedi, "COVID-19 prediction classifier model using hybrid algorithms in data mining," *Int. J. Pediatr.*, vol. 9, no. 1, pp. 12723–12737, 2021.

[19] S. H. Yoo *et al.*, "Deep learning-based decision-tree classifier for COVID-19 diagnosis from chest X-ray imaging," *Front. Med.*, vol. 7, p. 427, 2020.

[20] M. A. Rahman, M. S. Hossain, and A. Ghoneim, "Federated deep learning approach for COVID-19 prediction using IoT and wearable data," Future Generation Computer Systems, vol. 135, pp. 130–141, 2022.

[21] S. H. S. Moosavi and V. K. Bardsiri, "Poor and rich optimization algorithm: A new human-based and multi populations algorithm," *Eng. Appl. Artif. Intell.*, vol. 86, pp. 165–181, 2019.

[22] H. Eskandar, A. Sadollah, A. Bahreininejad, and M. Hamdi, "Water cycle algorithm–A novel metaheuristic optimization method for solving constrained engineering optimization problems," *Comput. Struct*, vol. 110, pp. 151–166, 2012.

[23] S. Vaiz and M. Ramaswami, "A study on technical indicators in stock price movement prediction using decision tree algorithms," *Amer. J. Eng. Res.*, vol. 5, no. 12, pp. 207–212, 2016.

[24] "Novel Coronavirus (COVID-19) Cases Dataset," Humanitarian Data Exchange, 2021. [Online]. Available: https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

[25] S. Minaee, R. Kafieh, M. Sonka, S. Yazdani, and G. J. Soufi, "Deep-covid: Predicting covid-19 from chest x-ray images using deep transfer learning," Medical Image Analysis, vol. 65, pp. 101794, 2020.

[26] S. Bhattacharya *et al.*, "Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey," *Sustain. Cities Soc.*, vol. 65, p. 102589, 2021.

[27] D. Liu, Z. Wu, and S. Sun, "Study on subway passenger flow prediction based on deep recurrent neural network," *Multimedia Tools Appl.*, pp. 1–14, 2020.

[28] R. Thomas and M. J. S. Rangachar, "Hybrid optimization based DBN for face recognition using low-resolution images," *Multimedia Res.*, vol. 1, no. 1, pp. 33–43, 2018.

[29] K. M. Gangappa, C. Kiran Mai, and P. Sammulal, "Enhanced crow search optimization algorithm and hybrid NN-CNN classifiers for classification of land cover images," *Multimedia Res.*, vol. 2, no. 3, pp. 12–22, 2019.

[30] A. Mohammed, A. B. et al., "Developing plastic recycling classifier by deep learning and directed acyclic graph residual network," Eastern-European Journal of Enterprise Technologies, vol. 2, no. 10, pp. 116, 2022.

[31] B. Mohammed *et al.*, "Developing plastic recycling classifier by deep learning and directed acyclic graph residual network," *East.-Eur. J. Enterp. Technol.*, vol. 2, no. 10, p. 116, 2022.

[32] M. A. Rahman, M. S. Hossain, and A. Ghoneim, "Federated deep learning approach for COVID-19 prediction using IoT and wearable data," *Future Gener. Comput. Syst.*, vol. 135, pp. 130–141, 2022.

[33] H. O. Ahmad and S. U. Umar, "Sentiment analysis of financial textual data using machine learning and deep learning models," *Informatica*, vol. 47, no. 5, 2023.

[34] H. Adhab and N. A. H. Ali, "Influence of wireless novel routing protocol by using MDPC algorithm," in *Proc. 4th Int. Conf. Technol. Adv. Comput. Sci. (ICTACS)*, 2024.