Computers, Materials & Continua
DOI: 10.32604/cmc.2023.029163
Article

Tech Science Press

Data Mining with Comprehensive Oppositional Based Learning for Rainfall Prediction

Mohammad Alamgeer[1], Amal Al-Rasheed[2], Ahmad Alhindi[3], Manar Ahmed Hamza[4,*], Abdelwahed Motwakel[4] and Mohamed I. Eldesouki[5]

[1]Department of Information Systems, College of Science & Art at Mahayil, King Khalid University, Abha, 62529, Saudi Arabia
[2]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia
[3]Department of Computer Science, College of Computing and Information System, Mecca, 24382, Saudi Arabia
[4]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, 16278, Saudi Arabia
[5]Department of Information System, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, AlKharj, 16278, Saudi Arabia
*Corresponding Author: Manar Ahmed Hamza. Email: ma.hamza@psau.edu.sa
Received: 26 February 2022; Accepted: 30 March 2022

Abstract: Data mining process involves a number of steps from data collection to visualization to identify useful data from massive data set. the same time, the recent advances of machine learning (ML) and deep learning (DL) models can be utilized for effectual rainfall prediction. With this motivation, this article develops a novel comprehensive oppositional moth flame optimization with deep learning for rainfall prediction (COMFO-DLRP) Technique. The proposed CMFO-DLRP model mainly intends to predict the rainfall and thereby determine the environmental changes. Primarily, data pre-processing and correlation matrix (CM) based feature selection processes are carried out. In addition, deep belief network (DBN) model is applied for the effective prediction of rainfall data. Moreover, COMFO algorithm was derived by integrating the concepts of comprehensive oppositional based learning (COBL) with traditional MFO algorithm. Finally, the COMFO algorithm is employed for the optimal hyperparameter selection of the DBN model. For demonstrating the improved outcomes of the COMFO-DLRP approach, a sequence of simulations were carried out and the outcomes are assessed under distinct measures. The simulation outcome highlighted the enhanced outcomes of the COMFO-DLRP method on the other techniques.

Keywords: Data mining; rainfall prediction; deep learning; correlation matrix; hyperparameter tuning; metaheuristics

1 Introduction

Data mining (DM) is the process of examining massive dataset for the identification of patterns and relationships which helps to resolve business problem by data analysis [1]. DM tools and technologies can be used for predicting future trends and decision making [2]. It is commonly employed by business intelligence and data analytics teams, assisting them to extract knowledge for huge quantity of data. It aims to extract and discover patterns in large sized data with the inclusion of machine learning, statistics, and database systems [3]. It finds useful in several applications in different areas such as education, healthcare, environmental monitoring, finance, banking industry, etc [4,5]. On the other hand, rainfall prediction has existed for a long time with conventional models that utilize statistical approaches [6,7] for assessing the relationships among the rainfall, geographic coordinate includes latitude and longitude, and atmospheric factor includes temperature, pressure, humidity, and wind speed. But, the difficulty of rainfall namely non-linearity makes them hard to forecast.

Statistical and mathematical approaches could be time-consuming with minor consequences and employ complex computational power. Machine learning (ML) based approach employs self-learning capacity to attain hidden features of echo variations and displays association and good memory capability [8]. It is employed as numerical prediction and classification model in climate prediction shows the broad and potential predictions of employing neural network systems to radar echo extrapolation [9]. Especially, it has lately employed deep learning (DL) method for processing meteorological big data, shows stronger technical performance and advantages, that has gained considerable interest from the research [10].

This article develops a novel comprehensive oppositional moth flame optimization with deep learning enabled rainfall prediction (COMFO-DLRP) technique for environmental monitoring. The proposed CMFO-DLRP model undergoes data pre-processing and correlation matrix (CM) based feature selection processes. Besides, deep belief network (DBN) model is applied for the effective prediction of rainfall data. Furthermore, COMFO technique was derived by integrating the concepts of comprehensive oppositional based learning (COBL) with traditional MFO algorithm. At last, the COMFO algorithm is employed for the optimal hyperparameter selection of the DBN model. For inspecting the improved performance of the COMFO-DLRP approach, a comprehensive experimental analysis was performed and the outcomes are assessed under various measures.

The rest of the paper is organized as follows. Section 2 briefs the existing works, Section 3 discusses the proposed model, Section 4 offers experimental validation, and Section 5 draws conclusion.

2 Literature Review

Sun et al. [11] presented the convolution 3D-gated recurrent unit (GRU), named Conv3D-GRU technique for predicting the future rainfall intensities on a comparatively short interval of time in the ML viewpoint. Primarily, the spatial feature of radar echo map with distinct height was removed by 3D convolutional, next, the radar echo map on time series were coding and decoding with utilize GRU. At last, the training method was utilized for predicting the radar echo map from the next 1–2 h. In [12], different methods and techniques are executed for predicting the rainfall data. The comparison analysis is demonstrated concentrating on evolving and relating various ML techniques, estimating various conditions and time horizons, and predicting rainfall utilizing two kinds of techniques.

Endalie et al. [13] implemented a rainfall prediction method to Jimma, an area placed from south-western Oromia, Ethiopia. It has presented the long short term memory (LSTM) based forecast technique able of predicting Jimma's daily rainfalls. An experiment is demonstrated for evaluating the presented technique utilizing different metrics. Venkatesh et al. [14] established a rainfall forecast method utilizing generative adversarial network (GAN) for analyzing rainfall data of India and forecasting the future rainfall. The presented technique utilized a GAN network in that LSTM network technique was utilized as generator and convolutional neural network (CNN) technique was utilized as discriminator. In [15], rainfall forecast was implemented for anticipating the damage to embankment. The rainfall forecast is executed utilizing LSTM dependent upon rainfall parameters: El-Nino and Indian Ocean Dipole (IOD). The experimentally are implemented with 2 methods: a primary technique utilized IOD and El-Nino parameters, but the secondary method utilized rainfall time series patterns.

Zhang et al. [16] presented the dual-input dual-encoder recurrent neural network (RNN) such as Rainfall Nowcasting Network (RN-Net), for solving this issue. It gets the past grid rainfall data included by automatic weather station and doppler radar mosaic data as input data after that predicts the grid rainfall data to the next 2h. It can be conducted experiments on the South-eastern China data set. Amine Ben Rhaiem et al. [17] utilized the everyday open rainfall data in the national observatory of Tunisian agriculture (ONAGRI) for developing an ETL (Extract, Transform, and Load) tools for automatically spatializing and loading the old information as to big data platforms with always increment a novel daily disseminated record. Besides, this work implements the Voronoi spatial analysis technique for estimating rainfall measures to the recently more spatial unit in OSM world mapping scheme. Next, according to these spatial estimations, the work studies the possibility of executing ARIMA to time series predicting.

3 The Proposed Model

In this study, a new COMFO-DLRP technique has been developed to predict rainfall and thereby determine the environmental changes. The presented COMFO-DLRP technique encompasses a series of processes namely data pre-processing, CM based feature selection, DBN based prediction, and COMFO based hyperparameter tuning. The COMFO algorithm is designed by incorporating the concepts of COBL with MFO algorithm. Fig. 1 illustrates the process flow of COMFO-DLRP technique.

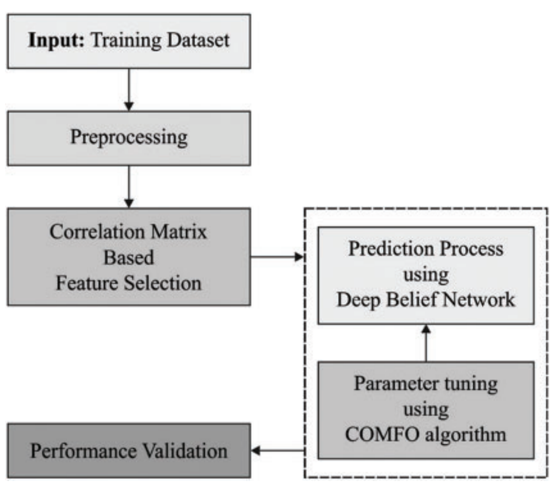

Figure 1: Process flow of COMFO-DLRP technique

3.1 Data Pre-processing and Feature Selection

Data pre-processing is an essential step in the rainfall prediction model, which transforms the raw data into useful format. Generally, pre-processing involves removal of categorical values, missing values, and structuring one hour of data into a vector row [18]. Besides, the weather condition codes understand weather id, weather description, weather main, and weather icon sub features, which comprises distinct records based on the weather type acquired in one hour. Besides, important features representing the variability of the input data are chosen by the use of CM. It eradicates the existence of repetitive features with no use of target features. Here, Pearson correlation coefficient is utilized for determining CM.

3.2 Design of DBN Based Predictive Model

During prediction process, the DBN model receives the preprocessed data as input to forecast the rainfall precisely. DBN is a basic DNN technique that contains distinct layers such as restricted Boltzmann machine (RBM) and multilayer perceptron (MLP). The RBM has visible and hidden units that are connected on the fundamental of weighted connection [19]. MLP is altered as feed-forward network which comprises output, input, and hidden layers. Consider there exist 2 RBMs like RBM1 and RBM2, and input to RBM1 is the feature vector attained in the big data. Input as well as hidden neurons from the input layer of RBM1 are formulated as:

$$\varepsilon^1 = \left\{\varepsilon_1^1, \varepsilon_2^1, \varepsilon_3^1, \ldots, \varepsilon_i^1, \ldots, \varepsilon_m^1\right\}; 1 \le i \le m \tag{1}$$

$$\beta^1 = \{\beta_1^1, \beta_2^1, \ldots, \beta_c^1, \ldots, \beta_f^1\}; 1 \le c \le f \tag{2}$$

where ε_i^1 implies the i^{th} input neuron which is existing from RBM1 and the amount of input neurons of RBMI was equivalent to dimensional of feature vectors. There is m neuron from the input layer of RBMI for performing classification. Assume that the entire amount of hidden neurons from the RBM1 be f and consider c^{th} hidden neuron from RBM2 be β_c^1. The bias of visible as well as hidden neurons of RBMI was formulated as:

$$A^1 = \left\{A_1^1, A_2^1, A_3^1, \ldots, A_i^1, \ldots, A_m^1\right\} \tag{3}$$

$$O^1 = \{O_1^1, O_2^1, \ldots, O_c^1, \ldots, O_f^1\} \tag{4}$$

The bias of hidden as well as input layers of RBM1 are equivalent to that of the entire neurons from both the layer and weight of RBMI are represented as:

$$a^1 = \left\{a_{ic}^1\right\}; 1 \le i \le m; 1 \le c \le f \tag{5}$$

where a_{ic}^1 implies the weight of RBM1 and it can be weighted amongst i^{th} input and c^{th} hidden neurons of RBMI. Thus, the resultant of RBM1 is demonstrated as:

$$\beta_c^1 = fun\left[0_c^1 + \sum_i \tau_i^1 a_{ic}^1\right] \tag{6}$$

where *fun* denotes the activation function from RBMI and τ_i^1 implies the feature vector. The resultant of RBMI is formulated as:

$$\beta^1 = \left\{\beta_c^1\right\}; 1 \le c \le f \tag{7}$$

The output in RBMI is provided as input to RBM2 and the output of RBM2 was evaluated utilizing the above formulas. The output of RBM2 was referred as G_j^2 that is provided as input to MLP. The input neuron in MLP are represented as:

$$\lambda = \left\{\lambda_1, \lambda_2, \ldots, \lambda_e, \ldots, \lambda_f\right\}; 1 \le c \le f \tag{8}$$

where f refers the entire input neurons from the MLP. The hidden neuron of MLP is written as:

$$E = \left\{E_1, E_2, \ldots, E_x, \ldots, E_y\right\}; 1 \le x \le y \tag{9}$$

where y demonstrated the entire hidden neurons of MLP layer. The bias of hidden neurons is formulated as:

$$B = \{B_1, B_2, \ldots, B_w, \ldots, B_z\}; 1 \le w \le z \tag{10}$$

where z defines the output neuron from the MLP layer. The weight amongst the input as well as hidden layers are provided as:

$$a^{mlp} = \left\{a_{cx}^{mlp}\right\}; 1 \le c \le f; 1 \le x \le y \tag{11}$$

where u_{cx}^{mlp} signifies the weight vector amongst c^{th} input neurons, and x^{th} hidden neurons. The resultant of hidden layer from MLP was dependent upon the bias as well as weights and is formulated as:

$$O^{hid} = \left[\sum_{c=1}^{f} a_{cx}^{mlp} \times F_c\right] a_x \tag{12}$$

where a_x denotes the bias of resultant layer. The weight vector amongst the hidden as well as output layers are demonstrated as:

$$a\prime = \left\{a'_{xw}\right\}; 1 \le x \le y; 1 \le \iota v \le z \tag{13}$$

Therefore, the resultant of MLP was evaluated as:

$$B_w = \sum_{x=1}^{y} a'_{xw} \times O^{hid} \tag{14}$$

where a'_{xw} stands for the weight amongst the hidden as well as resultant neurons from MLP, and O^{hid} refers the output of hidden layer.

3.3 Design of COMFO Based Hyperparameter Tuning Process

For optimally adjusting the hyperparameter values of the DBN model, the COMFO algorithm has been employed. MFO technique is a population based metaheuristic technique that inspires moth's performance from the night nearby the flame [20]. The common stages of MFO technique are nearly similar to individuals of other metaheuristic techniques as described in the subsequent:

- Creating a group of arbitrary primary population of moths (for instance, matrix M);
- Making a group of haphazard primary flame (for sample, matrix F);
- Computing the moth efficiency utilizing the FF;
- Conduct spiral effort of moths nearby the flame;
- Upgrade the amount of flames;
- Estimating the end situation, and when it could not be fulfilled, returned to step3;
- Return the optimum place of moth as solutions.

Afterward, can be provided a detailed summary of MFO technique. Moth and flame were important modules of MFO technique. At night, the moth flies nearby the flame at set angle. Once the moth realizes the light sources, it is endured for flying from a straight-line nearby the light sources. When the moth technique to light sources, it moves nearby the light source from spiral direction. The moth is searching agent, and flames were the optimum place established previously. Thus, each population gets this place as solution.

The MFO technique is group of moths which is demonstrated as the subsequent matrix:

$$M = \left[M_{1,1} M_{1,d} \because\because M_{n,1} M_{n.d} \right] \tag{15}$$

where n implies the amount of moths and d signifies the amount of problem variables. The matrix OM is another matrix to save the fitness value of solutions which depicts the level of solution quality. The matrix OM is formulated as:

$$OM = [OM_1 OM_2 \ldots OM_n]^T \tag{16}$$

Furthermore, the flames were other modules of MFO technique. The matrix F is to represent flame as demonstrated as:

$$F = \left[F_{1,1}\ F_{1,d} \because\because F_{n,1}\ F_{n.d} \right] \tag{17}$$

n implies the amount of moths (or flame), and d denotes the dimensional of problems or the amount of variables of the problems. Noticeable the dimensional of matrices M and F are equivalent to everyone. The matrix OF comprises the fitness value to the flame, as provided as follows:

$$OF = [OF_1 OF_2 \ldots OF_n]^T \tag{18}$$

Certainly, both moth and flame refer the solutions. All the moths search the space nearby their flame and all iterations determine an optimum solution and the flame demonstrates optimum solutions established by all the moths [21].

The MFO technique employs 3 functions for initializing the arbitrary places of moths (I), moving the moth from the searching space (P), and ending the searching state (T) based on Eq. (19):

$$MFO = (I, P, T) \tag{19}$$

where I demonstrates the function which initializing the primary population of moths as illustrated in Eq. (20).

$$I: \varnothing \rightarrow \{M, OM\} \tag{20}$$

In addition, P refers the function which moves the moths based on Eq. (20).

$$P: M \rightarrow M \tag{21}$$

The final function utilized is the function T. When the end state was fulfilled T return True, and when the end could not fulfilled, T return False as described in Eq. (22).

$$T: M- \rightarrow \{true, false\} \tag{22}$$

The moth and flame were the essential modules of MFO technique. The moth flies nearby the search space, but the flame illustrates the optimum place defined by the moth. The moth flies nearby the flame and upgrades its places by determining optimum places.

To enhance the efficacy of the traditional MFO algorithm, the COMFO algorithm has been derived with the use of COBL based population initiation. The OBL concept's aim is to discover the optimal solution by comparing the existing solution with opposite solution [22]. It is implemented by calculating the opposite solution X of X_j as follows:

$$\underline{X}_j = Ub_j + Lb_j - X_j j = 1, 2, \ldots, Dim \tag{23}$$

But there is an extension for the conventional OBL approach named complete OBL (COBL) [23] that enforces MFO algorithm approach for converging to the global solutions. The major concept of COBL is to change the solution $\underline{X} \in [Lb, Ub]$ to one of their opposite solution, such as Reflected extended opposition $\left(\underline{X}^{reo}\right)$ quasi-opposite $(\underline{X}^{qo})$, quasi-reflected $(\underline{X}^{qr})$ or Extended opposition $\left(\underline{X}^{eo}\right)$. Therefore, the COBL is expressed by.

$$\underline{X} = \{\underline{X}^{reo}\underline{X}^{qr}\underline{X}^{qo}\underline{X}^{eo}\gamma \le P_{reo}P_{reo} < \gamma \le P_{reo} + P_{qr}P_{reo} + p_{qr} < \gamma \le P_{reo} + P_{qr} + P_{qo}P_{reo} + P_{qr} + P_{qo} < \gamma \le 1 \tag{24}$$

whereas

$$\underline{X}_j^{reo} = \{X_j + (Ub_j - X_j) \times randX_i > (Lb_j + Ub_j)/2Lb_j + (X_j - Lb_j) \times randX_i < (Lb_j + Ub_j)/2,$$
$$j = 1, 2, \ldots, Dim \tag{25}$$

$$\underline{X}^{qr} = X_j + \left(Cb_j - X_j\right) \times rand, Cb_j = \frac{Lb_j + Ub_j}{2}, j = 1, 2, \ldots, Dim \tag{26}$$

$$\underline{X}_j^{qo} = Cb_j + \left(\underline{X}_j - Cb_j\right) \times rand, Cb_j = \frac{Lb_j + Ub_j}{2}, j = 1, 2, \ldots, Dim \tag{27}$$

$$\underline{X}_j^{eo} = \{\underline{X}_j + (Ub_j - \underline{X}_j) \times randX_i > (Lb_j + Ub_j)/2Lb_j + (\underline{X}_j - Lb_j \times randX_i < (Lb_j + Ub_j)/2,$$
$$j = 1, 2, \ldots, Dim \tag{28}$$

$\gamma \in [o, 1]$ an arbitrary value, P_{reo}, P_{qr}, and P_{qo} represents the probability of choosing $\underline{X}^{reo}$, $\underline{X}^{reo}$, and $\underline{X}$, correspondingly. Here, we follow a similar approach unutilized for upgrading the solution through COBL. At last, the optimal solution is carefully chosen from the existing solution X and opposite solution $\underline{x}$.

4 Performance Validation

This section assesses the rainfall forecasting outcome of the COMFO-DLRP technique using the rainfall data, collected for a duration of two years such as 2019-2021. Fig. 2 demonstrates the forecasting result analysis of the COMFO-DLRP technique under distinct runs. The figures reported that the COMFO-DLRP technique has obtained improved predictive outcomes under all rounds. Particularly, the difference between the actual and predicted rainfall falls are minimal.

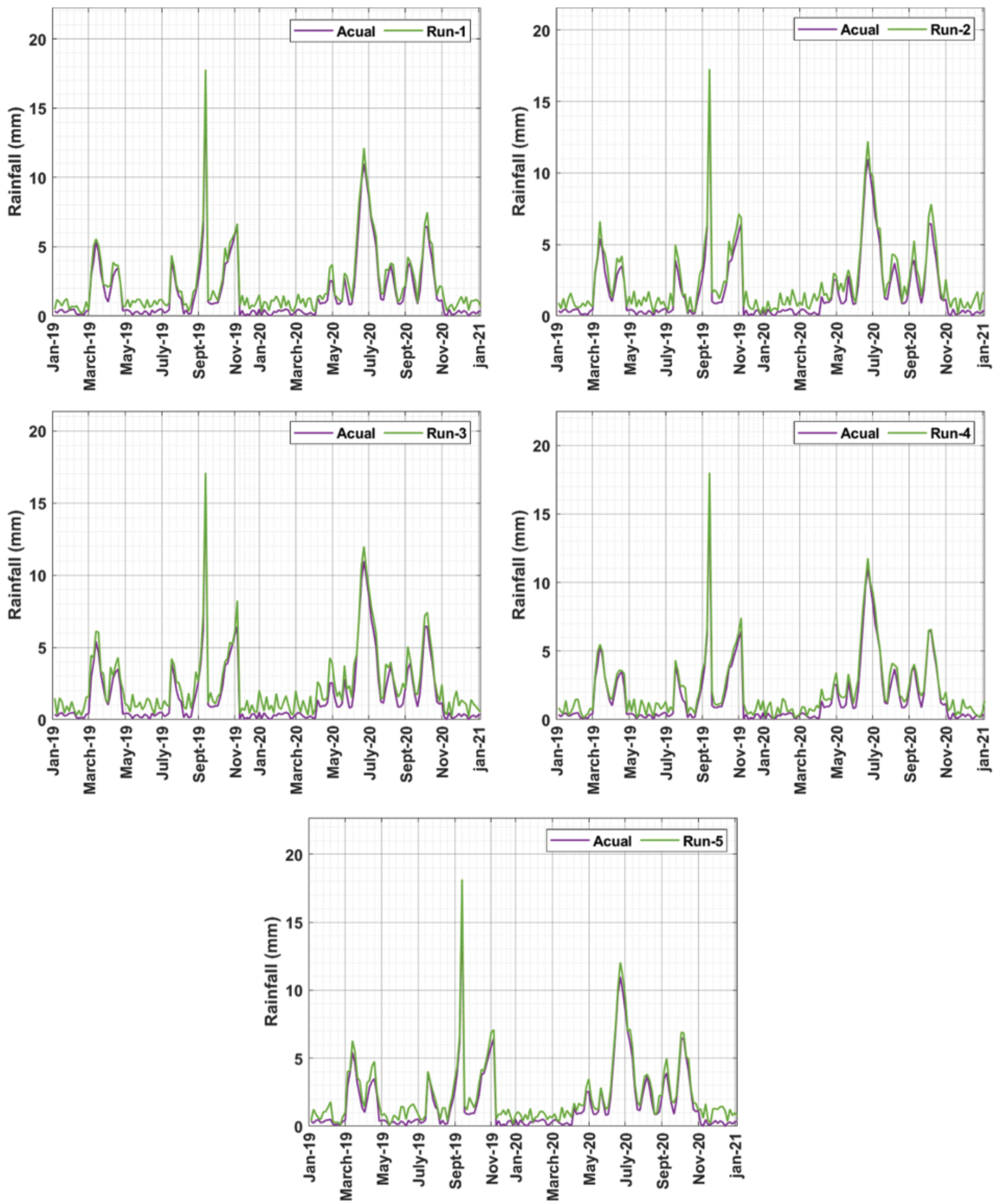

Figure 2: Rainfall prediction analysis of COMFO-DLRP technique

Fig. 3 illustrates the overall rainfall prediction result analysis of the COMFO-DLRP technique. The results show that the COMFO-DLRP technique has effectively forecasted the rainfall under all distinct runs.

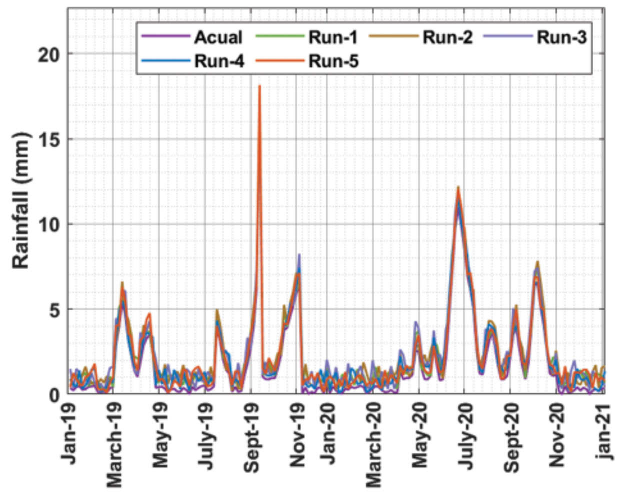

Figure 3: Overall rainfall prediction analysis of COMFO-DLRP technique

Tab. 1 and Fig. 4 demonstrate the overall mean square error (MSE) results of the COMFO-DLRP technique under distinct runs and durations.

Table 1: Result analysis of COMFO-DLRP technique with different runs

MSE Values					
Day/Year	Run-1	Run-2	Run-3	Run-4	Run-5
Jan-19	0.369713	0.496275	0.512775	0.311138	0.385938
Feb-19	0.136063	0.296713	0.526088	0.321688	0.383338
Mar-19	0.504750	0.622938	0.723425	0.161425	0.562438
Apr-19	0.257575	0.434113	0.934613	0.310125	0.813313
May-19	0.568188	0.904788	0.695563	0.520888	0.258688
Jun-19	0.469900	0.290300	0.739800	0.346638	0.781475
Jul-19	0.217500	0.608450	0.488400	0.414850	0.106513
Aug-19	0.438625	0.547938	0.836313	0.232250	0.464363
Sep-19	0.373663	0.427425	0.374338	0.362550	0.569075
Oct-19	0.342063	0.715625	0.609488	0.495638	0.383075
Nov-19	0.712913	0.412275	0.633013	0.355250	0.455913
Dec-19	0.662000	0.333850	1.157400	0.322713	0.389125
Jan-20	0.317750	0.679450	0.681100	0.296788	0.218913
Feb-20	0.682188	0.711550	0.834400	0.271700	0.450438
Mar-20	0.267138	0.565075	0.949863	0.395925	0.279850
Apr-20	0.570638	0.774675	0.961300	0.415550	0.240263
May-20	0.574238	0.565025	0.726788	0.698088	0.561875
Jun-20	0.355013	0.615513	0.768313	0.430263	0.427763

(Continued)

Table 1: Continued

MSE Values					
Jul-20	0.242688	0.787375	0.511400	0.414500	0.369000
Aug-20	0.390750	0.519250	0.727313	0.191100	0.389013
Sep-20	0.601813	0.953588	0.568300	0.147163	0.225775
Oct-20	0.465150	0.550788	0.886175	0.471475	0.743463
Nov-20	0.597063	0.938325	0.495213	0.273563	0.699375
Dec-20	0.570638	0.565025	0.726788	0.414500	0.225775
Jan-21	0.465150	0.768313	0.147163	0.240263	0.374338

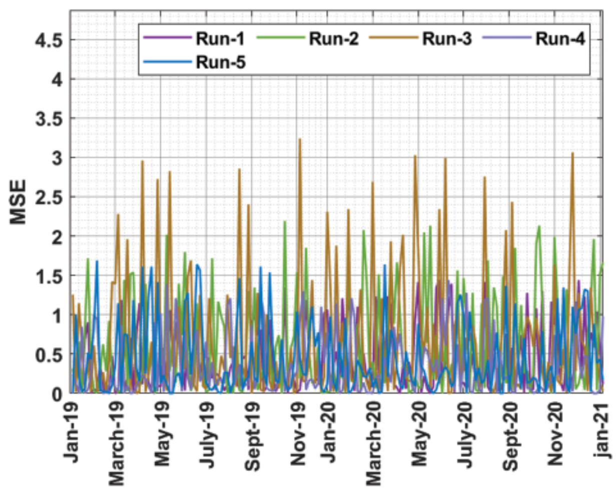

Figure 4: Result analysis of COMFO-DLRP technique with distinct runs

The results show that the COMFO-DLRP technique has resulted in enhanced performance with minimal values of MSE under all runs and seasons. For instance, on Jan-19, the COMFO-DLRP technique has obtained least MSE of 0.369713, 0.496275, 0.512775, 0.311138, and 0.385938 under runs 1-5 respectively. In Addition, on Apr-19, the COMFO-DLRP system has obtained lower MSE of 0.257575, 0.434113, 0.934613, 0.310125, and 0.813313 under runs 1–5 correspondingly. Moreover, on Aug-19, the COMFO-DLRP algorithm has reached least MSE of 0.438625, 0.547938, 0.836313, 0.232250, and 0.464363 under runs 1–5 respectively. Furthermore, on Jan-20, the COMFO-DLRP technique has gained minimum MSE of 0.317750, 0.679450, 0.681100, 0.296788, and 0.218913 under runs 1–5 correspondingly. Besides, on Apr-20, the COMFO-DLRP technique has obtained least MSE of 0.570638, 0.774675, 0.961300, 0.415550, and 0.240263 under runs 1–5 correspondingly. Along with that, in Jul-20, the COMFO-DLRP methodology has attained minimum MSE of 0.242688, 0.787375, 0.511400, 0.414500, and 0.369000 under runs 1–5 respectively. Meanwhile, on Oct-20, the COMFO-DLRP technique has gained least MSE of 0.465150, 0.550788, 0.886175, 0.471475, and 0.743463 under runs 1–5 correspondingly. Finally, on Jan-21, the COMFO-DLRP methodology has obtained least MSE of 0.465150, 0.768313, 0.147163, 0.240263, and 0.374338 under runs 1–5 correspondingly.

Tab. 2 provides a brief comparative rainfall forecasting outcomes of the COMFO-DLRP technique with other methods interms of MSE and root mean square error (RMSE) [24]. The table values highlighted that the COMFO-DLRP technique has resulted in effectual outcomes with minimal values of MSE and RMSE.

Table 2: Comparative analysis of COMFO-DLRP technique with recent approaches interms of MSE and RMSE

Methods	MSE	RMSE
MLP	7.73285	2.78080
1-DCNN	5.24135	2.28940
LSTM	5.33379	2.30950
DWRPM	4.71585	2.17160
COMFO-DLRP	0.35260	0.59380

Fig. 5 provides a comparative MSE examination of the COMFO-DLRP technique with recent methods. The results indicated that the MLP model has resulted in ineffectual outcomes with higher MSE of 7.73285. Followed by, the 1-DCNN and LSTM models have obtained slightly reduced MSE of 5.24135 and 5.33379 respectively. Moreover, the deep and wide rainfall prediction model (DWRPM) technique has resulted in to even decreased MSE of 4.71585. However, the COMFO-DLRP technique has accomplished superior results with the least MSE of 0.35260.

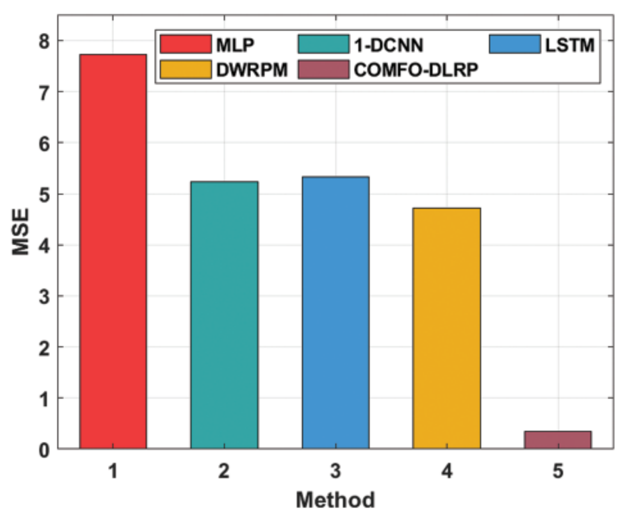

Figure 5: MSE analysis of COMFO-DLRP technique with recent approaches

Fig. 6 offers a comparative RMSE examination of the COMFO-DLRP approach with recent methods. The results referred that the MLP technique has resulted in ineffectual outcomes with superior RMSE of 2.78080. Likewise, the 1-DCNN and LSTM methodologies have gained slightly reduced RMSE of 2.28940 and 2.30950 correspondingly. Eventually, the DWRPM technique has resulted in even decreased RMSE of 2.17160. At last, the COMFO-DLRP methodology has accomplished superior outcomes with the minimum RMSE of 0.59380.

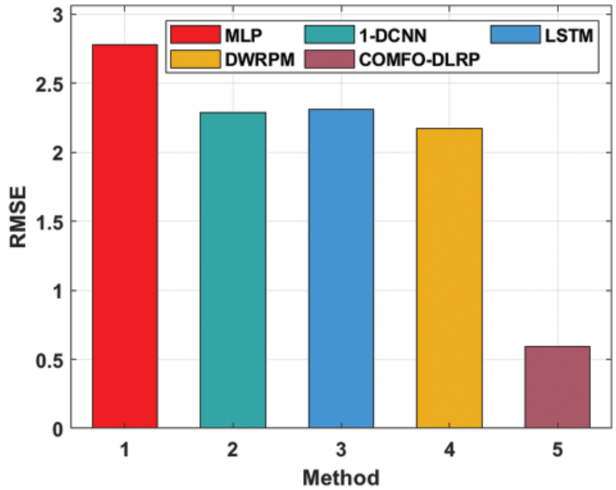

Figure 6: RMSE analysis of COMFO-DLRP technique with recent approaches

From the result analysis, it is ensured that the COMFO-DLRP technique has effectively forecasted the rainfall over the other techniques.

5 Conclusion

In this study, a novel COMFO-DLRP algorithm was developed to predict rainfall and thereby determine environmental changes. The presented COMFO-DLRP technique encompasses a series of processes namely data pre-processing, CM based feature selection, DBN based prediction, and COMFO based hyperparameter tuning. The COMFO algorithm is designed by incorporating the concepts of COBL with MFO algorithm. For inspecting the improved performance of the COMFO-DLRP approach, a comprehensive experimental analysis was performed and the outcomes are assessed under various measures. The simulation outcome highlighted the improved outcomes of the COMFO-DLRP method on the other techniques. Therefore, the COMFO-DLRP technique has the ability to attain improved predictive outcomes. In future, metaheuristics based feature selection algorithms can be derived to boost the classification results.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work under grant number (RGP 2/180/43). Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R235), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4270206DSR01).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems with Applications*, vol. 166, no. 2, pp. 114060, 2021.

[2] P. E. Cruz, R. Godina and E. M. G. Rodrigues, "A review of data mining applications in semiconductor manufacturing," *Processes*, vol. 9, no. 2, pp. 305, 2021.

[3] D. Shin and J. Shim, "A systematic review on data mining for mathematics and science education," *International Journal of Science and Mathematics Education*, vol. 19, no. 4, pp. 639–659, 2021.

[4] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Applied Sciences*, vol. 11, no. 1, pp. 237, 2020.

[5] A. Khan and S. K. Ghosh, "Student performance analysis and prediction in classroom learning: A review of educational data mining studies," *Education and Information Technologies*, vol. 26, no. 1, pp. 205–240, 2021.

[6] Y. Xiang, L. Gou, L. He, S. Xia and W. Wang, "A SVR-ANN combined model based on ensemble EMD for rainfall prediction," *Applied Soft Computing*, vol. 73, no. 2002, pp. 874–883, 2018.

[7] J. D. Sierra and M. del Jesus, "Long-term rainfall prediction using atmospheric synoptic patterns in semi-arid climates with statistical and machine learning methods," *Journal of Hydrology*, vol. 586, pp. 124789, 2020.

[8] Y. Dash, S. K. Mishra and B. K. Panigrahi, "Rainfall prediction for the Kerala state of India using artificial intelligence approaches," *Computers & Electrical Engineering*, vol. 70, pp. 66–73, 2018.

[9] R. Kumar, M. P. Singh, B. Roy and A. H. Shahid, "A comparative assessment of metaheuristic optimized extreme learning machine and deep neural network in multi-step-ahead long-term rainfall prediction for all-indian regions," *Water Resources Management*, vol. 35, no. 6, pp. 1927–1960, 2021.

[10] S. Sharadqah, A. M. Mansour, M. A. Obeidat, R. Marbello and S. M. Perez, "Nonlinear rainfall yearly prediction based on autoregressive artificial neural networks model in central jordan using data records: 1938–2018," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 2, pp. 240–247, 2021.

[11] D. Sun, J. Wu, H. Huang, R. Wang, F. Liang *et al.,* "Prediction of short-time rainfall based on deep learning," *Mathematical Problems in Engineering*, vol. 2021, no. 5, pp. 1–8, 2021.

[12] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed *et al.,* "Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1651–1663, 2021.

[13] D. Endalie, G. Haile and W. Taye, "Deep learning model for daily rainfall prediction: Case study of Jimma Ethiopia," *Water Supply*, pp. ws2021391, 2021.

[14] R. Venkatesh, C. Balasubramanian and M. Kaliappan, "Rainfall prediction using generative adversarial networks with convolution neural network," *Soft Computing*, vol. 25, no. 6, pp. 4725–4738, 2021.

[15] D. Z. Haq, D. C. R. Novitasari, A. Hamid, N. Ulinnuha, Y. Farida *et al.,* "Long short-term memory algorithm for rainfall prediction based on El-Nino and IOD data," *Procedia Computer Science*, vol. 179, no. 1, pp. 829–837, 2021.

[16] F. Zhang, X. Wang, J. Guan, M. Wu and L. Guo, "RN-Net: A deep learning approach to 0-2 h rainfall nowcasting based on radar and automatic weather station data," *Sensors*, vol. 21, no. 6, pp. 1981, 2021.

[17] M. A. B. Rhaiem, A. B. Hassine, A. Ouertani and I. R. Farah, "A hybrid rainfall prediction model for tunisian agriculture regions based on osm data, voronoi spatial analysis and long short term memory deep learning," in *EGU General Assembly Conf. Abstracts*, pp. EGU21–12993, 2021.

[18] E. O. Omuya, G. O. Okeyo and M. W. Kimwele, "Feature selection for classification using principal component analysis and information gain," *Expert Systems with Applications*, vol. 174, pp. 114765, 2021.

[19] S. M. Mujeeb, R. P. Sam and K. Madhavi, "Trust and energy aware routing algorithm for Internet of Things networks," *International Journal of Numerical Modelling: Electronic Networks, Devices and Fields*, vol. 34, no. 4, pp. e2858, 2021.

[20] R. A. Khurma, I. Aljarah and A. Sharieh, "A simultaneous moth flame optimizer feature selection approach based on levy flight and selection operators for medical diagnosis," *Arabian Journal for Science and Engineering*, vol. 46, no. 9, pp. 8415–8440, 2021.

[21] M. G. Arani, A. Souri, F. Safara and M. Norouzi, "An efficient task scheduling approach using moth-flame optimization algorithm for cyber-physical system applications in fog computing," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 2, pp. e3770, 2020.
[22] H. R. Tizhoosh, "Opposition-based learning: a new scheme for machine intelligence," in *Int. Conf. on Computational Intelligence for Modelling, Control and Automation and Int. Conf. on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, Vienna, Austria, 1, pp. 695–701, 2005.
[23] M. A. Elaziz and I. Attiya, "An improved Henry gas solubility optimization algorithm for task scheduling in cloud computing," *Artificial Intelligence Review*, vol. 54, no. 5, pp. 3599–3637, 2021.
[24] V. Bajpai, A. Bansal, K. Verma and S. Agarwal, "Prediction of rainfall in rajasthan, india using deep and wide neural network," 2010. [Online]. Available: https://arxiv.org/abs/2010.11787.