Global NEST Journal, Vol 26, No 2, 05408

An intelligent weather prediction model using optimized 1D CNN with attention GRU

Hemamalini S.[1], Geetha Rani K.[2], Rajasekar B.[3] and Sadish Sendil M[4]

[1]Department of Artificial Intelligence and Data Science, Panimalar Engineering College, Anna University, Chennai 600123, Tamil Nadu, India

[2]Department of Computer Science and Engineering, Jain (Deemed-to-be University), Bangalore 560069, India

[3]Department of Electronics and Communication Engineering, Sathyabama Institute of Science and Technology, Chennai, 600119, India

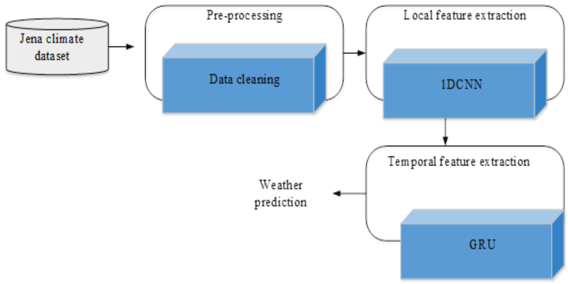
[4]Department of Emerging Technologies, Guru Nanak Institute of Technology, Telegana, 501506, India

Received: 27/09/2023, Accepted: 26/01/2024, Available online: 08/02/2024

*to whom all correspondence should be addressed: e-mail: hemamalini.annauniv2020@gmail.com

https://doi.org/10.30955/gnj.005408

Graphical abstract



Abstract

One of the main factors affecting human livelihoods is weather events. High weather disasters with forest fires, high air temperature, and global warming that cause drought. An efficient and accurate weather forecasting approach is required to take measures against climate disasters. Therefore, it is important to design an approach that makes better weather prediction. This work presents an optimized deep learning model, 1D convolutional neural network (CNN), with an attention gated recurrent unit (GRU) model for reliable weather forecasting. That is, to capture the local features of weather data, 1D CNN is used, and to capture the temporal features of the weather data, an optimized GRU is used. The attention mechanism is used for improving the performance, and the hyperparameter of GRU are optimized by the adaptive wild horse algorithm (AWHA). This work considered the Jena meteorological database which has 14 parameters, and the comparative analysis is carried out for different prediction measures. The proposed weather prediction model achieved better mean square error (MSE) and root mean square (RMSE) values.

Keywords: Weather forecasting, convolutional neural network, gated recurrent unit, adaptive wild horse algorithm, deep learning, mean square error

1. Introduction

Weather prediction plays an important role in early warning of weather effects on different features of people's livelihood. For example, weather prediction ensures proper decisions for self-driving vehicles to reduce traffic congestion and accidents, which is entirely based on prediction of environmental factors such as temperature, rainfall, and visibility of air. Ren X. *et al*., (2021) explained is an essential application that aims to forecast future weather events, particularly severe weather events. Correct and timely weather forecasting is always a major goal of meteorologists. But the traditional theory-based NWP (numerical weather prediction) model has challenges like imperfect concern of physical scheme and needs high computational sources. Thanarajan T *et al*., (2023) discussed some of the weather criteria such as temperature, wind, and humidity are very important to people's livelihood. Recently, temperature forecasting has become one of the hot research topics in weather monitoring. Traffic and congestion occur due to bad weather conditions. Chantry M *et al*. (2021) is to predict the warning of natural disasters for the early identification of the environment.

Some of the social advantages that can be obtained by correct weather prediction are the preservation of safety life, the encouragement of public safety, and the support of economic growth. To maximize crop productivity, farmers can benefit from statistical weather reports. Markovics D *et al*. (2022) informed weather is essential to prevent the spread of pests and diseases between crops. Non-linear approaches like weather prediction are overcome by the DL (deep learning) approach. Surendran R *et al*. (2023) developed the DL models have multiple-layered architectures and re-modify the data from the real data. These models overcome the hand-crafted features and automatically learn the features, thereby achieving better accuracy. Mehrkanoon S (2019) implements a deep analysis of the existing literature, several drawbacks have put forth several drawbacks. One of the most common issues is time and cost consumption during the examination process.

Balamurugan M. S *et al*. (2021) explained the weather prediction system is mainly dependent on time and spatial features. Conventional ML (machine learning) models are not adaptable to extract time and spatial features. Today, AI (artificial intelligence)-based ML techniques overcome the problems faced by existing methodologies by accurately learning the attributes. The ML converts the data into a useful way and helps determine the natural disaster. Watson-Parris D. (2021) created DL models such as CNN (convolutional neural network), GRU (gated recurrent unit) and LSTM (long-term short memory) is improved by the correlation between features and representation. In addition, Surendran R (2021) and Kirkwood C *et al*. (2021) developed DL models are well suitable for time series data. Therefore, scope of this study proposes an improved technique to help farmers identify the weather effectively. To the best of our knowledge, the proposed study overcomes all drawbacks faced in existing studies and shows outstanding performance with minimal time complexity. The main objectives of the proposed weather prediction model are as follows:

- To present an automated deep learning 1D CNN (convolutional neural network) with attention GRU (gated recurrent unit) weather prediction model for extracting the spatial and temporal features.
- To capture the local features of the weather data using 1D CNN and to capture the temporal features of weather data, an optimized GRU is used.
- Optimize the hyperparameters of GRU using the AWHA (adaptive wild horse algorithm).

The work sorted the remaining part of the work as follows: Section 2 shows related works based on weather prediction using different approaches, Section 3 shows the proposed weather prediction, Section 4 defines the analysis of the results, and Section 5 ends with the conclusion.

2. Related works

The literature works based on weather prediction using different approaches and the datasets used are listed here.

Bi K *et al*. (2023) presented an accurate medium stage weather prediction using the 3D neural network. The prediction performance was carried out with the unperturbed initial states of ERA5 and the RMSE performance was recorded. Liu C *et al*. (2023) developed series-wise model TLA (temporal lag attention) for extracting NWP data. There are four segments like TLA, BSAM (block sparse attention model), lag recognition, and feature fusion were carried out. The BSAM was used for screening the wide distributions of the TLA and the lag recognition was used for computing the similarity between the NWP time series and original wind speed. Hewage P *et al*. (2021) presented light weight DL based fine-grained weather prediction approach. Here, the temporal features were extracted using the LSTM and the patterns identified. TCN (temporal convolutional network) was used for examining long-range patterns. Comparative analysis was performed for the prediction of LSTM for 50 random samples.

Singh M *et al*. (2023) introduced cubed sphere for weather prediction and time stepping was carried out by U-Net. Here, the U-Net was used to transform the univariate to multivariate. Here, the weather bench and global forecast data sets were used for the experimental analysis and reduced computational time for the real-time prediction. Venkatachalam K *et al*. (2023) presented transductive LSTM for weather forecasting considering the rainfall parameter. Here, the weights were automatically learned and the performance was demonstrated by minimum and maximum humidity and temperature. The experimental analysis was carried out on datasets such as Jena and HHWD datasets and the accuracy attained was 98.2%.

Tekin S.F *et al*. (2021) presented attention to recurrent networks to extract spatial-temporal extraction. Here, the encoder with decoder model was used for long-term prediction and convolutional LSTM. The experiment was carried out on the ERA5 hourly data on the pressure level dataset. Finally, the MSE comparison was tested on the validation and test scores and provided better spatial and temporal solutions. Karevan Z *et al*. (2020) developed Transductive Lstm to obtain a time series model for weather prediction. In this work, the quadratic cost was set using the regression model and the object function localization was performed by the weighted quadratic cost. Finally, the cosine similarity was computed for the training samples and the test set. Performance was carried out by varying activation functions such as tanh and Sigmoid.

Surendran R *et al*. (2023) presented CNN with RNN-based DL model for weather prediction. The experiment was carried out on the Jena data set and the comparison was made with the conventional ML and DL models. The experimental analysis proved that CNN with RNN outperformed other models and achieved better R2 and MAE values of 0.98 and 0.12. Moosavi A. *et al*. (2021) demonstrate utilization of ML approach to analyze the numerical uncertainty of weather prediction. Initially, systematic errors were estimated, which improved the forecasting efficiency. Then, the physical processes were considered to forecast the uncertainty. The difference in model outcomes and observations of past series was used to learn the relation between physical processes and weather prediction.

Rajagopal S. *et al*. (2023) developed an adaptive multiplicative LSTM (AM-LSTM) for efficient weather forecasting. For extracting the features algorithms like DWT (discrete wavelet transform) and OVMD (optimal variational mode decomposition). Finally, denoising low- and high frequencies of all parameters are provided and forecasted using the AM-LSTM. The experimental results were analyzed from three stations and the performance such as temperature and wind speed was computed. Santhanaraj R. K. *et al*. (2023) developed a parallel convolutional neural network (CNN) and a gated recurrent unit (GRU) with an enhanced ResNet are used in this paper to forecast short-term load. The proposed model exhibits a 40% and 30% lower mean average percentage error than the GRU and serial CNN-GRU-iResNet models. This model outperforms the parallel CNN-LSTM-iResNet model by 12% in MAPE.

Tamilvizhi T *et al*. (2022) proposed that the model mentioned above functions as the primary element of a framework that is specifically engineered to preprocess training and testing data, streamline the training process of the model, and then implement the model onto a microcontroller that has restricted resources. Through the application of genuine meteorological datasets obtained from weather stations located in Spain, Greece, and China, we provide evidence that the predicted accuracy of the proposed model surpasses that of previous studies. Abdelwahab M. H. *et al*. (2023) Accurate wind power projections help plan the power grid and maximize renewable energy. A temporal pattern attention (TPA) multiobjective optimized recurrent neural network is presented in this paper. It handles wind farm randomization and unpredictability while forecasting regional wind power. Choosing wind farm meteorological elements using the Taguchi approach reduces waste and increases efficiency. A gated recurrent unit (GRU) and a denoising autoencoder (DAE) produce the stacked model next. This improves the temporal correlation and stability of the hidden state.In existing work, different DL approaches are combined with the model to improve prediction performance. However, these approaches were not considered to extract local and temporal characteristics. Furthermore, the DL models utilized in the existing models have overfitting issues. Therefore, the proposed work extracts local and temporal features using the 1D CNN with attention-based GRU for weather forecasting.

3. Methodology

The proposed work aims to present a DL model and utilize it in time series weather prediction. Here, the Jena climate dataset is used for weather forecasting. Initially, 1D CNN is used to capture the local features, and attention with GRU is used to capture the temporal features of weather data. Finally, the AWHA is used for optimizing the hyper-parameters and enhancing the prediction performance. The proposed model enhances prediction and forecast efficiency. The area of weather forecasting has made extensive use of the 1D Convolutional Neural Network (CNN) models, which combines the Attention Gated Recurrent Unit (GRU) model with the convolutional layers of the neural network. This has been done to make the most of the possibilities offered by deep learning techniques. An attention-based gated recurrent unit (GRU) and a one-dimensional convolutional neural network (CNN) are the components that make up the model that has been proposed. A Convolutional Neural Network (CNN) is used to extract features, and an Attention-Based Gated Recurrent Unit (GRU) is utilized to capture the temporal correlations among these qualities in meteorological data. Both of these processes are performed by a neural network. It is possible to demonstrate how accurate weather forecasts truly are by utilizing a model that combines a 1D convolutional neural network (CNN) with an attention gated recurrent unit (GRU). In a specific study, the model demonstrated an impressive level of accuracy when it came to forecasting the historical weather data contained within the dataset. When employing the attention gated recurrent unit (GRU) model to train a one-dimensional convolutional neural network (CNN), a significant amount of processing power is required. However, the model has an impressive capacity to reach a high level of accuracy, which makes it an extremely promising instrument for the purpose of weather forecasting. Figure 1 presents the framework of the proposed time series weather prediction model.

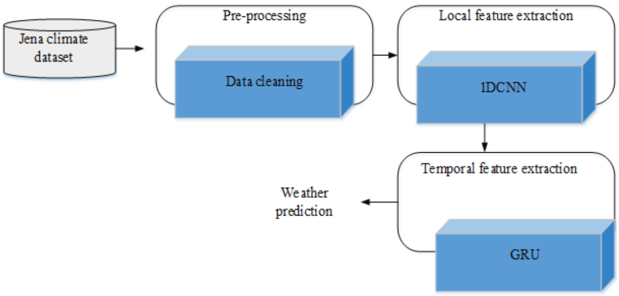


Figure 1. Framework of the proposed time series weather prediction model

3.1. Pre-processing

After the collection of the dataset, it should be preprocessed before the prediction process. Initially, the data cleaning process is performed; the data set has unnecessary attributes and are not useful for weather prediction. Therefore, in this work, unnecessary attributes and missing values are removed to obtain better prediction performance.

3.2. 1D-CNN for local feature extraction

In this work, 1D-CNN is used for modeling weather data, and this network extracts the local 1D patch from sequences for capturing the local sequences. The combination of a 1D Convolutional Neural Network (CNN) and an Attention Gated Recurrent Unit (GRU) model has been discussed as a promising approach to weather prediction. Significant improvements in the accuracy of weather forecasting have emerged from its usage, surpassing the efficacy of conventional methods, thanks to its ability to interpret and represent both spatial and temporal patterns in meteorological data. For each patch, the same input is performed; hence the 1D-CNN is translational invariant. The convolutional layer is used to perform the convolution operation. For weather prediction of m− layer to produce future predictionsT = (t, t+1, t+2, , t+n), this work considers the historical weather data $\{g_i^T\}$ of the m− layer as input. Then, the weather prediction matrix is given as equ (1):

$$G = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_m \end{bmatrix} \tag{1}$$

The first layer kernel carries out a convolutional operation on consecutive input sequential vectors P_c of G for recognizing the relevant features and update the features in the training stage in equ (2).

$$P_c = [g_1^{(t-n+c)}, g_2^{(t-n+c)}, \ldots\ldots, g_m^{(t-n+c)}]^T \quad (2)$$

The convolutional layer obtains the low-level features from the prior layer for modelling the high-level features. Output feature maps are generated using the filtering operation. The convolution operation provides the q^{th}feature map and it is calculated by equ (3):

$$M_c^q = \mathrm{Re}LU(W_c^q + P_c^q) + B_c^p \quad (3)$$

Where W is the weight and it is updated using the BP (back propagation) algorithm and B is the bias. The activation function used is ReLU and it is given as in equ (4).

$$g_{ReLU}(L) = \max(0, L) \quad (4)$$

Where L is the maxpooling function and one of the major problems in the network is overfitting. This problem is overcome by the dropout function. In this Figure 2, the dimension of input layer is C × S × 26 in which C and S are the number of connections in the training stage and the number of prior order steps. In this work, weather forecasts are carried out on the basis of temperature, humidity, and pressure. The structure model is the 1D convolutional neural network (CNN) and the gated recurrent unit (GRU) with attention mechanisms make up the bulk of the model known as the 1D CNN with Attention GRU. To learn the temporal correlations between these properties, an attention-based GRU is paired with CNN, which is responsible for collecting relevant characteristics from the meteorological data.

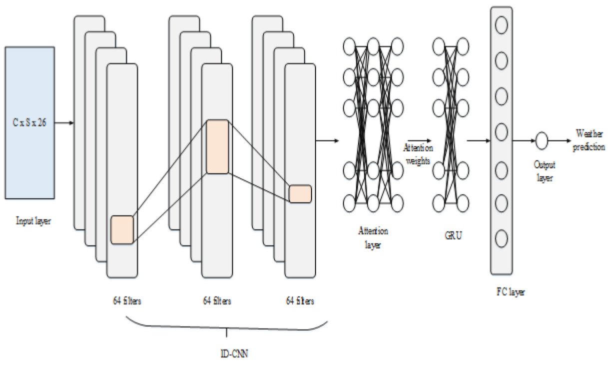


Figure 2. Structure of 1D CNN with attention GRU

3.3. Attention module

The aim of the attention module is to learn the past weather effects from the future weather effects. This correlation is influenced by different criteria, such as temperature, rainfall, and humidity. It has a high complexity in weather forecasting, and the prediction is based on the context of the input features. The GRU model may lose the long-sequence context and the prediction information may lose. Therefore, the attention module is used to convert the input data sequences to the matrix for identifying accurate predictions. The attention module is represented in equ (5):

$$A_t = \delta_a(\alpha(M) = Sig(W\alpha(l) + \beta)) \quad (5)$$

where δ_a is the attention learning model, α is the term that interconnect the input and hidden neurons. Wis the weight of the neurons, l is the vector of the convolution matrix M and β is the bias. Finally, the weighted weather matrix is given as $MA_t = M \cdot A_t$ and it is the input to the GRU model. For learning the A_t, the FC (fully connected) network with the hidden layer and Sigactivation function are utilized.

3.4. Temporal feature extraction

For modelling temporal features, both long-term and short-term dependencies play major roles in weather prediction. The GRU network has the benefit of utilizing gated neurons to capture long-term and short-term dependencies. GRU has two gates, the two gates like update gate X_f and reset gate R_f. These gates maintain long-term features and the mathematical expression of GRU are explained in the following section.

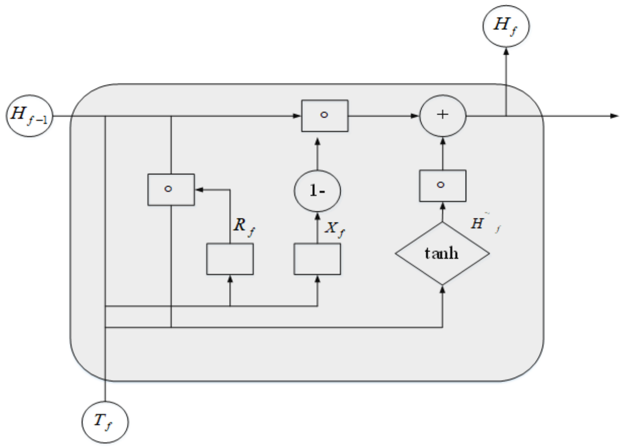


Figure 3. Structure of GRU

To obtain the short-term temporal features, the GRU inputs for short-term features are represented as $T_f = [g_1^{(t-n+f)}, g_2^{(t-n+f)}, \quad g_m^{(t-n+f)}]$. For every historical features, output of l^{th}layer is given as $H_l = (H_0^l, H_1^l, \ldots\ldots\ldots H_{n-1}^l)$. The below expressions are used for generating the temporal features and it is given in equ (6,7,8):

$$X_f = \varphi_f(W_x T_f + U_x H_{f-1} + b_x) \quad (6)$$

$$R_f = \varphi_f(W_r T_f + U_x H_{f-1} + b_q) \quad (7)$$

$$H_f = (1 - Z_f) \circ \tilde{H}_f + Z_f \circ H_{f-1}) \quad (8)$$

where∘is the element-wise multiplication, ϕ_f is the activation function, $H_{f-1}$is the hidden state at f−1unit and $\tilde{H}_f$ is the present memory unit. Figure 3 shows the structure of the GRU.

3.5. Fully Connected (FC) layer

Finally, the FC layer is used for forecasting the weather; 1D-CNN is used for extracting high-stage features, and the attention mechanism with GRU is used for extracting temporal features. To make the prediction process, the FC focuses on linear and non-linear features and predicts the output. The AWHA optimizer is used to minimize the loss function MSE (mean square error) and is calculated as in equ (9):

$$MSE = \frac{\sum_{j=1}^{M}\left[s_j^{(m)} - \hat{s}_j^{(m)} \right]}{M} \tag{9}$$

where $s_j^{(m)}$ and $\hat{s}_j^{(m)}$ are the estimated and predicted weather data.

3.6. Adaptive Wild Horse Algorithm (AWHA)

AWHA mimics the social life characteristics of wild horses and has different characteristics such as mating, grazing, pursuing, domination, and command. AWHA is based on standard WHA and Cuckoo search (CS) optimization. The initial population is split into M(population) and S (set of groups). Generally, horses live in a group and have different foals and mares. There are five stages that are involved in the AWHA and it is explained in the following section:

Grazing characteristics: For simulating the grazing characteristics, the following expression is utilized and is given in equ (10):

$$X_{j,L}^{k} = 2Y\cos(2\pi RY)\times(Leader^{k} - X_{j,L}^{k}) + Leader^{k} \tag{10}$$

where $X_{j,L}^{k}$ is the present foal location, $Leader^{k}$ $Leader^{k}$ is the leader position, R is the random number and Yis the adaptive parameter computed using the following expression in equ (11):

$$\begin{aligned} &W = \vec{Rand}1 < T_{dr};\, I_{dx} = (W == 0); \\ &Y = Rand2 \times I_{dx} + \vec{Rand}3(\sim I_{dx}) \end{aligned} \tag{11}$$

where W is the vector value, Rand2 is the random number, $\vec{Rand}1$ and $\vec{Rand}3$ are the random vectors. I_{dx} is the index value of $\vec{Rand}1$ and T_{dr} is the adaptive parameter. T_{dr} is computed by equ (12):

$$T_{dr} = 1 - iteration \times \left(\frac{1}{\max\_iteration} \right) \tag{12}$$

where iteration is the present iteration and max_iteration is the maximum iteration

Mating characteristics: For implementing mating characteristics, the foal moves from the original set j to temporary set k. For simulating this mating characteristic, the cross-over operator is utilized and given in equ (13):

$$X_{L,M}^{p} = Crossover\left(X_{L,j}^{q}, X_{L,k}^{Y} \right) \tag{13}$$

where $X_{L,M}^{p}$ is the horse position p from the set M, $X_{L,j}^{q}$ is the foal position q and $X_{L,k}^{Y}$ is the horse mating position Z.

Group leadership: In the AWHA model, the leaders make the sets for the water hole. Leaders fight for the water hole., hence the winning set can use the water hole at first, and after that the rest group uses the water hole. This stage is given as follows in equ (14):

$$Leader_{L_j} = \begin{cases} 2Y\cos(2\pi RY)\times(W_h - Leader_{L_j}) + W_h & when\, rand > 0.5 \\ 2Y\cos(2\pi RY)\times(W_h - Leader_{L_j}) - W_h & when\, rand \le 0.5 \end{cases} \tag{14}$$

where $Leader_{L_j}$ is the next leader position and W_h is the water hole position.

Exchange Leaders and Selection: In this phase, the leaders are selected on the basis of fitness. The position of the leader and the group members are changed on the basis of the following expression in equ (15).

$$Leader_{L_j} = \begin{cases} X_{L,j} & when\, cost\left(X_{L,j}\right) < cost(Leader_{L_j}) \\ Leader_{L_j} & when\, cost\left(X_{L,j}\right) < cost(Leader_{L_j}) \end{cases} \tag{15}$$

The proposed AWHA is based on the CS optimizer and for every iteration of AWHA, the new solution is obtained by the Levy flight and is given in in equ (16):

$$X_{j,L} = X_{j,L} - \alpha\left(X_{j,L} - X_l\right) \oplus Levy(\gamma) = X_{j,L} + \frac{0.01\,a}{|b|^{1/\gamma}}\left(X_{j,L} - X_l\right) \tag{16}$$

where α is the size of step scale, X_l is the best global solution and γ is the exponent of Levy's flight. Finally, the a and bare denoted in equ (17)

$$a \sim \left(0, \sigma_a^2\right), b \sim \left(0, \sigma_b^2\right) \tag{17}$$

σ_a and σ^b are the standard deviations and a is defined in equ (18):

$$\sigma_a = \left[\frac{\sin\left(\frac{\gamma\pi}{2}\right)\Gamma(1+\gamma)}{2^{(\lambda-1)}\Gamma\left(\frac{1+\gamma}{2}\right)} \right], \quad \sigma_b = 1 \tag{18}$$

where Γ is the Gamma term. The main advantage of the proposed AWHA is that it has the ability to balance local exploitation and global exploration. Figure 4 shows the flow chart of the proposed AWHA.

4. Analysis of results

This section presents the results analysis of the proposed and existing models. Evaluation of the work is carried out on the Python platform with 8GB of memory. Table 1 presents the hyperparameters utilized for the experimental analysis.

4.1. Dataset details

The Jena climate data set (2016) is utilized for the weather prediction process and has fourteen different parameters, such as pressure (mbar), temperature (degC), temperature at the dew point (degC), relative humidity (rh) (%), saturation vapor pressure (VPmax), vapor pressure (VPact), vapor pressure deficit (VPdef), specific humidity (sh), water vapor concentration (H2OC), airtight (rho), wind speed (wv), maximum wind speed (max. wv) and wind direction (wd). The readings are recorded every ten minutes and it has data from January 2009 to December 31, 2017.

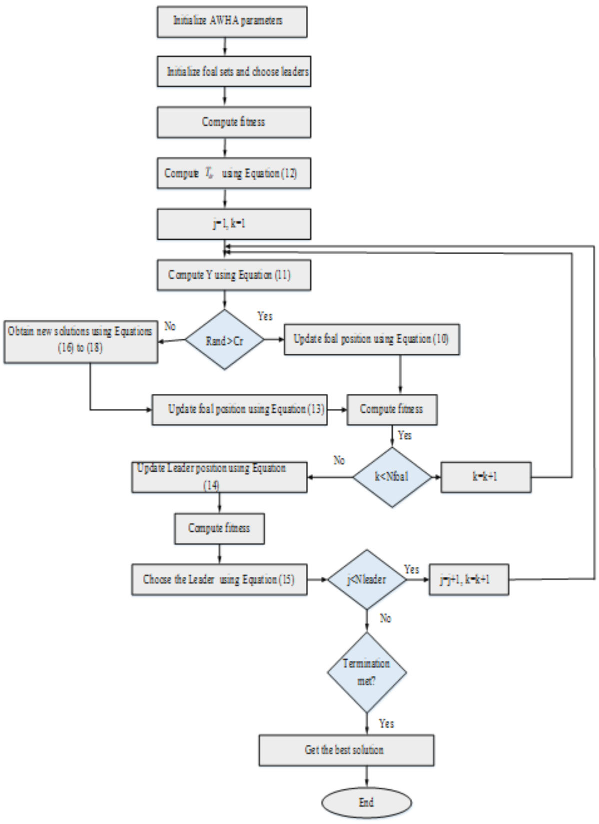


Figure 4. Flow chart of the proposed AWHA

Table 1. Hyperparameters

Hyper-parameters	Values
Dimension of input layer	(30100), (4), (26)
Filters in convolutional layer	64
Number of convolutional layers	3
Stride of convolutional layer	(1, 1)
Numbers of attention layers	3
Numbers of GRU layers	2
Dropout	0.1
Neurons in GRU-dense layer	64
Neurons in last dense layer	1
Size of population	50
Maximum iteration	100
Crossover percentage Cr	0.13
Number of separate run	20

4.2. Performance measures

In weather prediction to determine the performance of the 1D CNN with GRU-AWHA attention, the following measures are measured, such as RMSE (root mean square), MAE (mean absolute error), and R2.

RMSE: It is the square root of the mean squared variation between the actual and observed weather values. It is defined as:

$$RMSE = \sqrt{\frac{1}{M}\sum (Z-\hat{Z})^2} \tag{19}$$

MAE: It is the absolute variation between the actual and observed weather values. It is defined as:

$$MAE = \frac{1}{M}\sum (Z-\hat{Z}) \tag{20}$$

R2: This measure shows how the predictive approach sets the dependent parameter and it is given as:

$$MAE = 1 - \frac{\sum (Z-\hat{Z})^2}{\sum (Z-\bar{Z})^2} \tag{21}$$

where M is the number of samples, Z is the actual weather value $\hat{Z}$ is the observed weather value and $\bar{Z}$ is the average of all samples.

4.3. Comparative analysis

This section shows the comparative analysis of the proposed and existing weather prediction models. Initially, the weather predicted by the proposed model is given. Then, a comparison of performances like RMSE, MAE, and R2 is presented. Then, the accuracy-loss curves, the confusion matrix, and the efficiency of the optimization approach are tested. Multiple models were used for weather prediction, including the 1D Convolutional Neural Network (CNN) with Attention Gated Recurrent Unit (GRU) model, a statistical model, and a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM). Both models exhibit lower accuracy and generalization performance compared to the 1D CNN with Attention GRU model. Figure 5 presents the weather prediction performance of the proposed 1D CNN with GRU-AWHA attention. Here, the actual vs. predicted values of temperature, pressure, humidity, and wind speed are plotted. The predicted values are forecast outcomes of varying time steps from 1 to 100000. Once the weather performance is predicted, the predicted outcomes are compared with the actual values. It is observed that the predicted value of the proposed 1D CNN with attention to GRU-AWHA is exactly matched with the actual values. Hence, it is proved that parameters like temperature, pressure, humidity, and wind speed are essential for weather prediction.

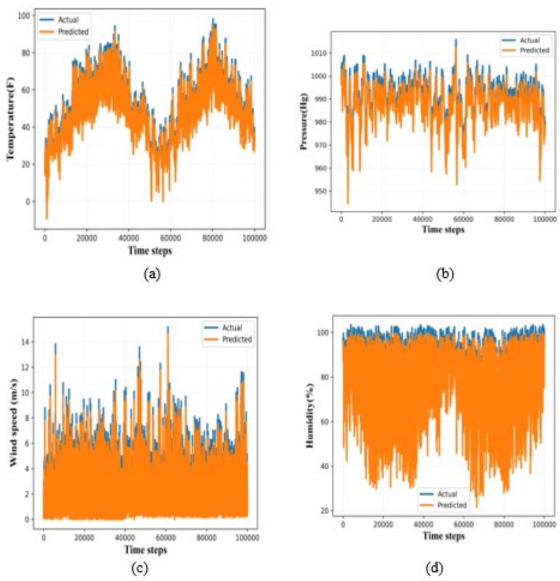


Figure 5. Actual vs. predicted values of (a) temperature, (b) pressure, (c) humidity, and (d) wind speed

Figure 6 presents the performance of the proposed 1D CNN with GRU-AWHA attention with respect to the accuracy and loss curves. In this observation, the performance is evaluated by varying the values for 140 epochs. The graphs are plotted between the train and test values. It is noted that the model is not underfit or overfit, and it has better generalization. Therefore, it is proved that the proposed 1D CNN with GRU-AWHA attention can be used in the weather prediction process.

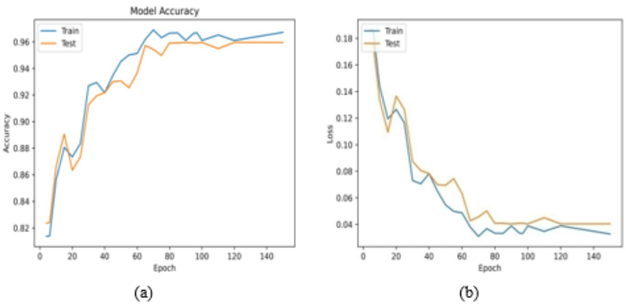


Figure 6. Performance of the proposed 1D CNN with attention GRU-AWHA (a) accuracy and (b) loss curf

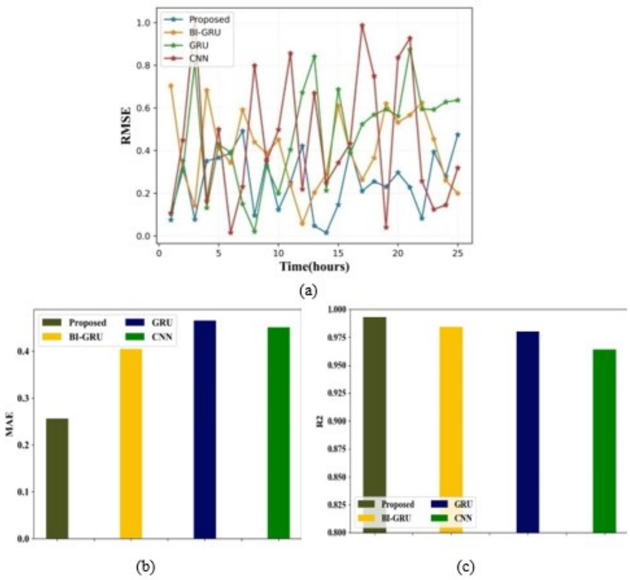


Figure 7. Performance of various techniques (a) RMSE, (b) MAE, and (c) R2

Figure 7 shows the performance of various techniques such as CNN, GRU, Bi-GRU, and the proposed 1D CNN with the attention of GRU-AWHA. Here, the measures like RMSE, MAE and R2 are computed. Figure 7 (a) shows the RMSE performance of the different DL approaches. This comparison is considered by varying the time from 1 to 25 hours. It is observed from the graph that the RMSE value of the proposed model is much less than that of the other approaches. Similarly, for better weather prediction, the model should have a lower MAE value and in Figure 7 (b), the proposed 1D CNN with attention GRU-AWHA has a lower MAE value. For better weather prediction, the model should have a high R2 value and in Figure 7 (c), the proposed 1D CNNS with attention GRU-AWHA has a high R2 value. In all comparisons, the proposed model attained better results due to better weight selection by AWHA. Furthermore, the proposed model uses only present weather characteristics as the main feature for predicting the weather, but the conventional approaches utilize historical weather characteristics. This may limit the robustness; hence, these conventional approaches attain poor performance.

Figure 8 presents the confusion matrix of the proposed 1D CNN with GRU-AWHA attention. The percentage matrix between correctly and incorrectly predicted samples is given on the left side and the absolute values are given on the right side. Here, the fourteen parameters like it has fourteen different parameters such as pressure (mbar), temperature (degC), temperature at the dew point (degC), relative humidity (rh), saturation vapor pressure (VPmax), vapor pressure (VPact), vapor pressure deficit (VPdef), specific humidity (sh), water vapor concentration (H2OC), airtight (rho), wind speed (wv), maximum wind speed (max. The proposed model accurately predicted the wind direction (wv) and the wind direction (wd).

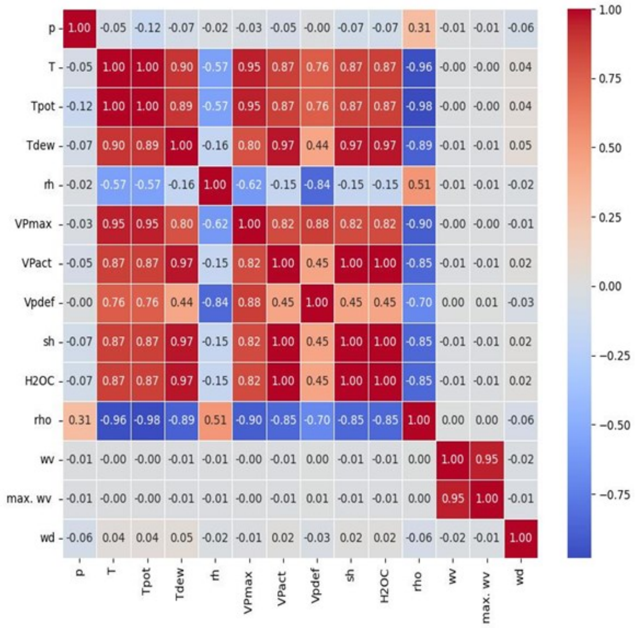


Figure 8. Confusion matrix of the proposed 1D CNN with attention GRU-AWHA

Figure 9 shows the convergence analysis of the proposed AWHA and WHA. Here, the performance is carried out for 100 iterations. For the 20th iteration, the fitness value achieved by the proposed AWHA is 4 and the WHA is 26. Similarly, for all iterations, the proposed AWHA achieved better fitness values. This achievement is due to the new solution obtained by the Levy flight in the AWHA. Hence, this algorithm overcomes the slow convergence and is trapped by local optima. Therefore, it is proved that this proposed AWHA provides a better performance enhancement in weather prediction.

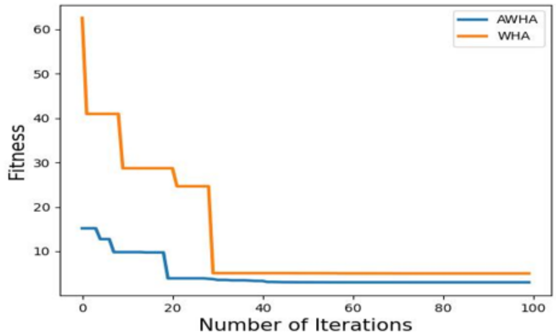


Figure 9. Convergence Analysis

The empirical results demonstrate that the use of the 1D Convolutional Neural Network (CNN) with the Attention Gated Recurrent Unit (GRU) model is an effective tool for weather prediction. The improved precision of the model in predicting weather conditions can be primarily attributed to its ability to comprehend spatial and temporal patterns in meteorological data. Compared to alternative models, the model consistently exhibits reduced mean absolute error (MAE) and root mean square error (RMSE) values, indicating its extraordinary efficacy in predicting temperatures and precipitation. Taking these factors into account, it can be concluded that the use of the 1D Convolutional Neural Network (CNN) with Attention Gated Recurrent Unit (GRU) model exhibits significant potential in enhancing the precision and dependability of weather prediction.

5. Conclusions

This work presented an enhanced DL model for weather forecasting. Here, 1D-CNN was used for local feature extraction, the attention module was used for extracting the salient features of the weather, and GRU was used for extracting temporal features. The AWHA metaheuristic optimizer was used to optimize the hyperparameters of the network. The performance of the proposed model and existing models were analyzed in the Jena data set. In all comparisons, the proposed weather prediction model outperformed conventional approaches. The MAE and R2 values achieved by the proposed weather prediction model were 0.5 and 0.982, respectively. It was observed that the proposed 1D CNN with attention GRU-AWHA can be able to forecast the weather better. In the future, different deep learning models with different numbers of attributes will be considered for weather forecasting.

Ethics declarations

Funding statement

No funding was received for this study.

Acknowledgment

The authors would like to thank the Panimalar Engineering College, Jain (Deemed-to-be University), Sathyabama Institute of Science and Technology, and Guru Nanak Institute of Technology for providing the resources.

Data availability statement

Available Based on Request. The datasets generated and or analyzed during the current study are not publicly available due to the extension of the submitted research work. They are available from the corresponding author upon reasonable request.

Conflict of interest

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

References

Abdelwahab M.H., Mostafa H. and Khattab A.A. (2023). low footprint olive grove weather forecasting using a single-layered seasonal attention encoder-decoder model. *Ecological Informatics*, **75**, 102113.

Balamurugan M.S. and Manojkumar R. (2021). Study of short term rain forecasting using machine learning based approach. *Wireless networks,* **27,** 5429–5434. https://doi.org/10.1007/s11276-019-02168-3

Bi K., Lingxi X., Hengheng Z., Xin C., Xiaotao G. and Qi T. (2023). Accurate medium-range global weather forecasting with 3D neural networks. *Nature,* 1–6. https://doi.org/10.1038/s41586-023-06185-3

Chantry M., Hatfield S., Dueben P., Polichtchouk I. and Palmer T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. Journal of Advances in Modeling *Earth Systems,* **13,** 7, e2021MS002477. https://doi.org/10.1029/2021MS002477

Dataset J.C. (2016). Weather time series dataset recorded at the Weather Station of the Max Planck Institute for Biogeochemistry in Jena, Germany. https://www.kaggle.com/mnassrib/jena-climate.

Hewage P., Trovati M., Pereira E. and Behera A. (2021). Deep learning-based effective fine-grained weather forecasting model. *Pattern Analysis and Applications,* **24,** 1, 343–366. https://doi.org/10.1007/s10044-020-00898-1

Karevan Z. and Suykens J.A. (2020). Transductive LSTM for time-series prediction: An application to weather forecasting. *Neural Networks*, **125,** 1–9. https://doi.org/10.1016/j.neunet.2019.12.030

Kirkwood C., Economou T., Odbert H. and Pugeault N. (2021). A framework for probabilistic weather forecast post-processing across models and lead times using machine learning. *Philosophical Transactions of the Royal Society A*, **379,** 2194, 20200099. https://doi.org/10.1098/rsta.2020.0099

Liu C., Xuemin Z., Shengwei M., Qingyu Z. and Hang F. (2023). Series-wise attention network for wind power forecasting considering temporal lag of numerical weather prediction. *Applied Energy,* **336,** 120815. https://doi.org/10.1016/j.apenergy.2023.120815

Markovics D. and Mayer M.J. (2022). Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renewable and Sustainable Energy Reviews*, **161**, 112364. https://doi.org/10.1016/j.rser.2022.112364

Mehrkanoon S (2019). Deep shared representation learning for weather elements forecasting. *Knowledge-Based Systems*, **179**, 120–128. https://doi.org/10.1016/j.knosys.2019.05.009

Moosavi A., Rao V. and Sandu A. (2021). Machine learning based algorithms for uncertainty quantification in numerical weather prediction models. *Journal of Computational Science,* **50,** 101295. https://doi.org/10.1016/j.jocs.2020.101295

Rajagopal S., Thanarajan T., Alotaibi Y. and Alghamdi S., Brain Tumor. (2023). Hybrid Feature Extraction Based on UNet and 3DCNN. *Computer Systems Science & Engineering*, **45**, no. 2. http://dx.doi.org/10.32604/csse.2023.032488

Ren X., Li X., Ren K., Song J., Xu Z., Deng K. and Wang X. (2021). Deep learning-based weather prediction: a survey. *Big Data Research,* **23,** 100178. https://doi.org/10.1016/j.bdr.2020.100178

Santhanaraj R.K., Rajendran S., Romero C.A.T., and Murugaraj S.S. (2023). Internet of Things Enabled Energy Aware Metaheuristic Clustering for Real Time Disaster

Management. *Computer Systems Science and Engineering,* **45,** 1561–1576.

Singh M., Acharya N., Patel P., Jamshidi S., Yang Z.L., Kumar B., Rao S., Gill S.S., Chattopadhyay R., Nanjundiah R.S. and Niyogi D. (2023). A modified deep learning weather prediction using cubed sphere for global precipitation. *Frontiers in Climate*, **4,** 1022624. https://doi.org/ 10.3389/fclim.2022.1022624

Surendran R., Alotaibi Y. and Subahi A.F. (2023). Lens-Oppositional Wild Geese Optimization Based Clustering Scheme for Wireless Sensor Networks Assists Real Time Disaster Management. *Computer Systems Science and Engineering,* **46,** 835–851. http://dx.doi.org/10.32604 /csse.2023.036757

Surendran R., Alotaibi Y. and Subahi A.F. (2023). Wind Speed Prediction Using Chicken Swarm Optimization with Deep Learning Model. *Computer Systems Science & Engineering,* **46,** 3. http://dx.doi.org/10.32604/csse.2023.034465

Surendran R., Tamilvizhi T., and Lakshmi S. (2021). Integrating the Meteorological Data into a Smart City Service Using Cloud of Things (CoT). In Emerging Technologies in Computing: *4th EAI/IAER International Conference,* iCETiC 2021, Virtual Event, August 18–19, 2021, Springer International Publishing, **4,** 94–111. https://doi.org/10.1007/978-3-030-90016-8_7.

Tamilvizhi T., Surendran R., Romero C.A.T. and Sadish M. (2022), Privacy preserving reliable data transmission in cluster based vehicular adhoc networks, *Intelligent Automation & Soft Computing,* **34**, 1265–1279.

Tekin S.F., Karaahmetoglu O., Ilhan F., Balaban I. and Kozat S.S. (2021). Spatio-temporal weather forecasting and attention mechanism on convolutional lstms. *arXiv preprint,* arXiv:2102.00696, **4**.

Thanarajan T., Alotaibi Y., Rajendran S. and Nagappan K. (2023). Improved wolf swarm optimization with deep-learning-based movement analysis and self-regulated human activity recognition. *AIMS Mathematics,* **8,** 5, 12520–12539. DOI: 10.3934/math.2023629

Venkatachalam K., Trojovský P., Pamucar D., Bacanin N. and Simic V. (2023). DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM). *Expert Systems with Applications*, **213,** 119270. https://doi.org/10.1016/ j.eswa.2022.119270

Watson-Parris D. (2021). Machine learning for weather and climate are worlds apart. *Philosophical Transactions of the Royal Society A*, **379.** 2194 20200098. https://doi.org/10.1098/rsta.2020.0098