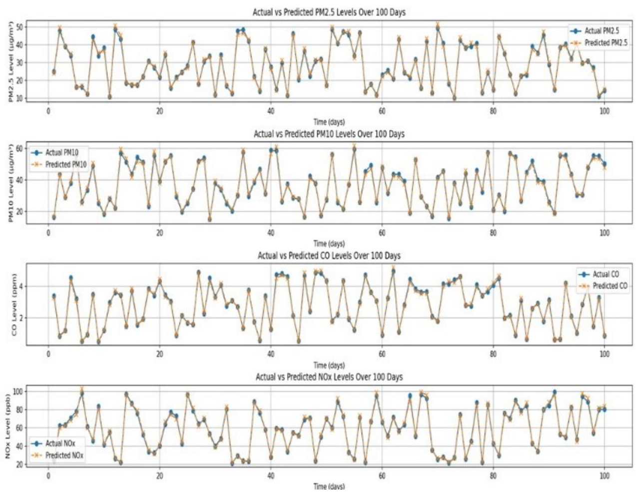
Global NEST Journal, Vol 26, No 6, 06192

Deep learning-based air pollution prediction model using modified gated recurrent unit

Periyanan A.[1] and Dr. S. Palanivel Rajan[2]
[1]Assistant Professor, Department of Electronics and Communication Engineering Sri Ranganathar Institute of Engineering and Technology Athipalayam, Coimbatore–641 110 Tamil Nadu, India
[2]Professor, Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology (Autonomous), Viraganoor, Madurai - 625009, India
Received: 22/05/2024, Accepted: 23/06/2024, Available online: 27/06/2024
*to whom all correspondence should be addressed: e-mail: periyanan12@gmail.com
https://doi.org/10.30955/gnj.006192

Graphical abstract



Abstract

In recent days air quality has been very poor in urban society due to a lot of industries, vehicles and so on. To overcome poor quality, an air quality prediction is much needed to manage an urban area and control pollution. In this article, deep learning (DL) namely a Modified Gated Recurrent Unit (GRU) neural network is used to forecast air pollution. Initially, a new activation function named Dual-Slope Leaky ReLU (DSL-ReLU) is introduced. The DSL-ReLU activation function includes two hyperparameters which control the slopes for positive and negative input values to achieve fine-tuned responses for varying data inputs. Then, the model parameters like learning rate, batch size, and dropout rates are tuned using the Spider Wasp Optimization (SWO) algorithm. The SWO algorithm is based on the female spider wasp's hunting and nesting behaviour that is applied to solve the optimization issues. This SWO model is used to GRU model parameters and provides robustness in the predictive modelling of air quality monitoring. For validation, the Kaggle data set is used whereas the proposed optimized GRU model considerably improves the accuracy of air pollution forecasts.

Keywords: GRU, air quality prediction, deep learning, SWO, optimization

1. Introduction

In recent decades, Air pollution has been a serious problem in developed and developing countries due to its poor air quality. This low quality of air threatens environmental health and also affects the public lifestyle (Manisalidis *et al*. 2020; Haleem *et al*. 2019). Due to enormous industrial activities, vehicular emissions, agricultural practices and residential heating in an urban area are the major reasons for pollution. In atmospheric air, there are pollutants like PM10, PM2.5, nitrogen oxides (NOx), carbon monoxide (CO), sulphur dioxide (SO2) and ozone (O3) etc are majorly presented to cause air pollution (Kumar and Pande 2023). It is clear that air pollution causes environmental degradation and affects ecosystems.

To avoid such poor air quality, air quality prediction is more important to control pollution for faster urbanization and industrialization (Binbusayyis *et al*. 2024). Accurate forecasting of air pollutant levels is to be predicted to take preventive measures and provide harmless health effects. Moreover, effective prediction models enable governments to apply guidelines and policies to control and manage pollution.

Previously lot of air quality forecast models were developed based on its chemical and physical behaviours. However, these models are not successful in capturing the complex and non-linear interactions among various pollutants. Meanwhile, it provides limited predictive accuracy, especially in dynamic and densely populated urban environments. To predict air quality based on historical data and meteorological conditions, most advanced DL models are applied for air pollution prediction.

The DL model performs data preprocessing and feature extraction stages separately for prediction. In DL models, time series models of recurrent neural networks (RNNs), Long Short-Term Memory (LSTM) and GRU models have proven excellent promise in addressing these challenges (Drewil and Al-Bahadili, 2022; Bekkar *et al*. 2021). Among these models, the GRU model gives a simplified architecture compared to LSTMs and RNNs with fewer parameters to optimize. The GRU model is computationally efficient with robust predictive performance.

In this work, a modified GRU model with a novel activation function for predicting air pollution. The DSLReLU function introduces two control parameters to control the slopes for positive and negative inputs. To optimize the GRU model parameters and the DSLReLU activation function, we use the SWO algorithm.

The paper is organized into the following chapters. Chapter 2 presents a literature review of air pollution prediction techniques. Chapter 3 describes the proposed GRU for air quality prediction. Chapter 4 presents the results; Chapter 5 presents the conclusions of the work.

2. Related work

Y. Han, *et al*. (2022) developed an air pollution prediction model using the Bayesian DL model. The proposed model includes two strategies for accurate prediction. It is based on domain specific which learns the statistical relationship between attributes strongly. In addition, the self-attention mechanism is used between layers.

The hybrid PM2.5 prediction model is proposed by C. -Y. Lo *et al*. (2022) Initially, the autoencoder is applied for data preprocessing. The LSTM model based on K-means is used for PM2.5 prediction. In their study (Yi *et al*. 2021), the authors proposed a DL model for air quality prediction. The output from different layers a processed and combined using a fusion layer for long-term air quality prediction. Compared to sole models, the fusion-based deep neural model shows higher accuracy and minimum error rates.

B. Liu *et al*. (2021) explored a weighted GRU model for PM2.5 forecasting. The time series of features of a data set is learned using data clustering. For each cluster, the GRU model is applied with different weights for forecasting. Graph Neural Networks (GNNs) have developed as powerful tools for learning from sequential data. Hierarchical Graph Neural Networks (HGNNs) extend this advantage by incorporating hierarchical structures into the learning process. This HGNN-based model is proposed by J. Han *et al* (2023) for air quality prediction. This model converts spatial features of air quality data into spatial temporal features for learning correlations among features accurately. In (Xu and Yoneda, 2021), the authors analyse the performance of multi-layer LSTM for air pollution prediction. The historical data is pre-processed by using autoencoders. By the construction of multi-layers in LSTM, the model effectively learns spatiotemporal patterns in data.

Echo State Networks (ESNs) model is based on the concept of reservoir. This reservoir transforms the input signals into a high-dimensional dynamic state space. This transformation is best suitable for time-series prediction. M. Xu, *et al* (2019) introduce an ESN-based prediction model for air quality. The performance of the model is analysed in terms of mean square error rates.

The hybrid model which integrates both Convolutional Neural Networks and Bi-directional LSTM (Bi-LSTM) model for air quality prediction is developed by S. Du *et al* (2021). Initially, NN involves the extraction of spatial features. Then, this model is applied to learning spatial-temporal features. Likewise, the authors (Jin *et al*. 2021) presented a DL model called multiple nested LSTM networks (MTMC-NLSTM) for PM2.5 prediction. Similarly, M. Singh *et al* (2023) presented an LSTM for the prediction of CO2 emissions in the environment.

Adversarial Meta-Learning is an advanced model that combines adversarial training and meta-learning. Z. Wu *et al* (2023) introduce a pollution prediction models using combined adversarial meta-learning. It involves Bayesian meta-learning to learn the hidden features.

To handle accurate short-term forecasting, a hybrid model is developed by Q. Tao *et al* (2019). This model combines convolutional neural networks and bidirectional GRU for handling both spatial and temporal features. Compared to other models, the bidirectional GRU effectively process the data in both directions and reduces prediction accuracy considerably.

B. Liu *et al*. (2019) analysed the performance of different air pollution prediction models. In addition, proposed a new recurrent model based on the attention mechanism. The AutoRegressive Integrated Moving Average also called the ARIMA model is a popular model for time-series forecasting. It processes the data based on three operations autoregression, differencing and moving average for accurate prediction. The modified ARIMA-based PM2.5 prediction model is proposed by U. A. Bhatti *et al*. (2021). In (Qin *et al*. 2019), a new optimized LSTM-based prediction model is proposed. Likewise, the modified LSTM model which integrates a Bayesian neural network for the prediction of PM2.5 is proposed (Nguyen *et al*. 2023).

H. Chen, *et al* (2021) proposed a fusion technique for meteorological forecasting. There are two types of LSTM models used for feature learning. In the fusion model, the output from LSTM models is combined and the final prediction is done by using XGBoosting models.

In their study, Y. Huang *et al* (2020) proposed an optimized back propagation neural network for predicting air quality index. The weight of the neural network is tuned using an improved particle swarm optimization algorithm. In improved swarm optimization, the mutation strategy is applied to avoid local minima issues.

C. Liu, *et al* (2023) proposed an extreme learning model for air quality forecasting. The evolutionary algorithm of genetic optimization is integrated to optimize the activation functions and dropout rates in the extreme learning model. A transfer learning-based hybrid model is proposed by J. Yang *et al* (2023) for PM2.5 prediction. This model uses LSTM and residual neural units to capture long-term dependencies between pollution parameters.

M. H. Nguyen *et al* (2021) address the limits of existing air pollution prediction models. Further, proposed an LSTM combined encoder and decoder model for the accurate PM2.5 prediction. Results shows that the encoder model shows more accurate predictions than those of existing models.

K. Gu (2022) *et al* presented a temporal support vector regressor model for air quality prediction. The temporal Support Vector Regressor model is an extension of the traditional Support Vector Regression (SVR) that is specifically adapted to handle time-series data. It finds a function that deviates from the target values by a value no greater than a specified margin while being as flat as possible. Y. Cong *et al* (2022) introduce a self-attention-based feature-aware LSTM model to forecast air pollutants gas concentration. In the self-attention mechanism, the weight of the model is adjusted based on feature importance.

Natarajan S.K. *et al* (2024) Grey wolf optimization-based machine learning model used to predict the air quality index. The machine learning algorithm of the decision tree is optimized to reach a higher accuracy. In their study, the authors (Biswas *et al*. 2024) used an image processing technique to predict the air quality. The deep learning models of CNN are applied to extract features from the images.

3. Proposed model

In this work, a modified GRU model with a novel DSL-ReLU activation function is proposed for air quality forecast. This model is constructed to effectively capture the temporal dependencies in air pollution data and adapt to the varying magnitudes of pollutant concentrations. The steps involved in the proposed prediction are given in Figure 1. In the first step, the data set is collected and normalized. The SWO component processes to find the appropriate hyperparameters for the GRU model. The GRU model is trained with SWO. The error values are used as a fitness function for SWO.

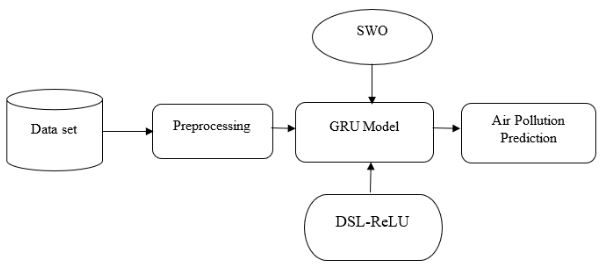


Figure 1. Overall workflow

3.1. GRU model

The GRU is a type of RNN architecture developed to address challenges associated with vanishing gradients in traditional RNNs (Zeng *et al*. 2022; Pan *et al*. 2020; Yao and Abisado, 2024; Li *et al*. 2020). The architecture of GRU is given in Figure 2. The key components of GRU models are gates to process the pollution data. There are two important gates: Update Gate (zt) and the Reset Gate (rt). The Update Gate determines how much past data should be carried over to the present. The Reset Gate decide how much past data should be forgotten. The Current Memory Content represents new information that will be added to the memory, and the Hidden State Update combines the old and new information to produce the updated hidden state. This architecture allows the GRU to effectively capture long-term dependencies in pollution data. The use of gating mechanisms enables the models to selectively process the air pollution data and to identify the correlations between pollutants effectively.

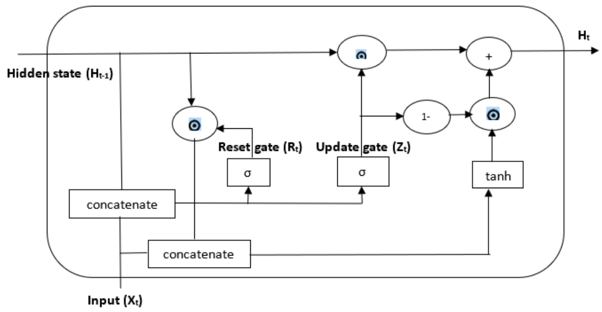


Figure 2. GRU Architecture

The GRU is characterized by its gating mechanisms which control the flow of information through the network that is given in the following equation (1-4).

$$zt = \sigma\left(Wz \cdot \left[ht-1, xt\right]\right) \tag{1}$$

$$rt = \sigma\left(Wr \cdot \left[ht-1, xt\right]\right) \tag{2}$$

$$h't = \tanh\left(W \cdot \left[rt \odot ht-1, xt\right]\right) \tag{3}$$

$$ht = \left(1-zt\right) \odot ht-1 + zt \odot h't \tag{4}$$

where xt is the input at time t, Wz, Wr, and W are the weight matrices for the corresponding gates, respectively. h’t and ht are current memory content and hidden state updates. σ denotes the sigmoid activation function, tanh denotes the hyperbolic tangent activation function and represents element-wise multiplication 5.

3.2. DSL-ReLu Model

The proposed model is named as DSL-ReLu which is an advanced variant of the traditional ReLu activation function. This function has additional two learnable parameters of α_p and α_n in it. These parameters control the functional slopes for both positive and negative input values. Based on this process, it carries a higher flexibility in shaping the activation function to improve data fitting and improve the neural network’s learning capacity. The DSL-ReLU function is defined as follows:

$$\begin{cases} \alpha_p . x \text{ if } x > 0 \\ \alpha_n . x \text{ if } x <= 0 \end{cases} \tag{5}$$

Where x is the input to the activation function. αp is the learnable factor that controls the slope for positive input values. α_n is the learnable factor that controls the slope for negative input values.

3.3. SWO Algorithm

The SWO model is based on female spider wasps' behaviour of hunting, nesting and mating activities (Abdel-Basset *et al*. 2023). This model has a unique strategy of prey searching, nest securing and reproduction to identify an optimal solution in a search space. The SWO model carries several phases as follows.

3.3.1. Initialization phase

Initialization is used to create a various set of potential solutions across the search space. Each spider wasp in the population signifies a possible solution vector in a D-dimensional space. It also initialises the positions of these wasps that are generated randomly within the defined bounds that are expressed as follows:

$$\mathrm{SW_i} = [\mathrm{x_1}, \mathrm{x_2} \ldots . \mathrm{x_D}] \tag{6}$$

The population can be randomly produced as follows:

$$SW_{i,t} = L + \vec{r} \times (H - L) \tag{7}$$

Where, $\vec{r}$ is a vector of random numbers. The random initialization helps in covering various regions of the search space and making the algorithm robust against local minima.

3.3.2. Searching phase (Exploration)

It mimics the random exploration spider wasp's behaviour of prey searching. It explored the search space thoroughly to recognize precise areas for further exploitation. Wasps explore randomly or around a dropped prey for random exploration and localized search that is given in the following equation.

$$SW_{i,t+1} = SW_{i,t} + \mu_1 \times (SW_{a,t} - SW_{b,t}) \tag{8}$$

$$\mu_1 = r_n \times r_1 \tag{9}$$

$$SW_{i,t+1} = SW_{c,t} + \mu_2 \times (L + r_2 \times (H - L)) \tag{10}$$

$$\mu_2 = B \times \cos(2\pi l) \tag{11}$$

$$B = \frac{1}{1 + e^l} \tag{12}$$

Where, a, b and c is a random index.

Following and Escaping Phase (Exploration and Exploitation)

It balances exploration and exploitation by pretending the chasing dynamic behaviors and adjusting the search direction. The wasps chase prey (exploitation) or manage their path when prey escapes (exploration) using adaptive equations that are given below.

$$SW_{i,t+1} = SW_{i,t} + c \times 2 \times r_5 \times (SW_{a,t} - SW_{i,t}) \tag{13}$$

$$C = 2 - \frac{2 \times t}{t_{max}} \times r_6 \tag{14}$$

Where *c* is a distance-controlling coefficient.

The escaping behaviour of the wasps can be expressed as follows:

$$SW_{i,t+1} = SW_{i,t} \times v_c, v_c \sim N(k,k) \tag{15}$$

$$k = 1 - \frac{t}{t_{max}} \times r_6 \tag{16}$$

3.3.3. Nesting behavior (Exploitation)

This stage exploits promising areas identified during the search by mimicking the nesting behavior of spider wasps which secure and optimize their nests. The wasps focus on pulling prey to a suitable nesting region and avoid overlap in nesting locations using specific equations. The strategy of pulling to a suitable nesting region can be expressed as follows:

$$SW_{i,t+1} = SW^* + \cos(2\pi l) \times ((SW^* - SW_{i,t})) \tag{17}$$

The behaviour of avoiding the same nest location can be stated as follows:

$$SW_{i,t+1} = SW_{a,t} + r_3 \times \gamma \times (SW_{a,t} - SW_{i,t}) + (1 - r_3) \times U \times (SW_{b,t} - SW_{c,t}) \tag{18}$$

Where γ is the levy flight distribution and U is a binary vector:

$$U = \begin{cases} 1 \, if r_4 > r_5 \\ 0 \; otherwise \end{cases} \tag{19}$$

3.3.4. Mating Behavior

This stage simulates the reproductive process, generating new solutions (offspring) from existing ones to introduce diversity and improve solution quality. New solutions are generated through crossover and mutation operations, reflecting the mating and gender determination behaviours of spider wasps.

$$SW_{i,t+1} = crossover(SW_{i,t}, SW_{m,t}, CR) \tag{20}$$

$$SW_{i,t+1} = SW_{i,t} + e^t \times \beta \times \overrightarrow{v_1} + (1 - e^t) \times \beta_1 \times W_{m,t}, \overrightarrow{v_2} \tag{21}$$

Where. β and β_1 are normally distributed, and:

$$\overrightarrow{v_1} = \begin{cases} \overrightarrow{x_a} - \overrightarrow{x_i} \; if \overrightarrow{x_a} < f(\overrightarrow{x_i}) \\ \overrightarrow{x_i} - \overrightarrow{x_a} \; otherwise \end{cases} \tag{21}$$

$$\overrightarrow{v_2} = \begin{cases} \overrightarrow{x_b} - \overrightarrow{x_c} \; if \overrightarrow{x_b} < f(\overrightarrow{x_c}) \\ \overrightarrow{x_c} - \overrightarrow{x_b} \; otherwise \end{cases} \tag{22}$$

3.4. GRU parameter tuning using SWO

Initially, population size and mutation rates are initialized key parameters. The population of potential solutions is represented as a vector in a multidimensional space. The exploration phase begins with random exploration. Wasps explore both randomly and around potential targets to examine the search space to identify the exact area. It is expressed through equations that guide the wasps' movements and decision-making processes.

Also, the SWO algorithm delicate balance between exploration and exploitation. By dragging prey to suitable nesting areas and evading nesting overlaps, the algorithm maximizes its effectiveness in locating optimal solutions within the search space.

Finally, mating behaviour simulates the reproductive process. Through crossover and mutation operations, new solutions are generated from existing ones. The pseudocode of SWO-based parameter tuning is given in Algorithm 1.

Algorithm 1: SWO Algorithm

```
Initialize parameters such as population size, mutation rates, and other relevant variables
Initialize population P with randomly generated solutions
Evaluate fitness for each solution in P
Sort population P based on fitness
Repeat until convergence criteria are met:
    FOR each solution in P:
        Conduct random exploration
            FOR each solution in P:
        Perform local search
    FOR each solution in P:
        Simulate chasing and escaping behavior
          FOR each solution in P:
        Mimic nesting behavior
            FOR each solution in P:
        Simulate mating behavior
    Replace solutions in P with the new solutions.
    Sort population P based on fitness.
Perform tuning for the GRU model.
END Spider Wasp Optimization (SWO) Algorithm
```

4. Results and discussion

The proposed GRU model is trained using a historical dataset obtained from the Kaggle website (Open-source). This dataset consists of hourly time-series data from various monitoring stations in India, covering the period from 2017 to 2020. It includes hourly measurements for multiple pollutants such as PM2.5, PM10, CO, NOx, O3 SO2 etc. For the training and testing of the model, the four pollutants are selected: PM2.5, PM10, CO, and NOx. The dataset was split into two parts: 75% of the data is allocated for training the model and the remaining 25% is used for testing purposes. The proposed GRU model is evaluated using RMSE, MAE, and MSE metrics, as defined in equations 23-25. These metrics evaluate the variance between the predicted and actual values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\bar{y} - y)^2 \quad (23)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\bar{y} - y)^2} \quad (24)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\bar{y} - y| \quad (25)$$

Initially, a GRU model is configured with the following hyperparameters: 50 units per layer, 2 layers, a learning rate of 0.001, a batch size of 32 and a dropout rate of 0.2 with 100 epochs. After optimization, through iterative adjustments, the hyperparameters are refined, resulting in a model with 120 units per layer, 3 layers, a learning rate of 0.0005, a batch size of 64, and a dropout rate of 0.3. Figure 3 displays the learning curve, depicting the gradual decrease of the loss function over epochs. Also, the graph observed that the loss drops very suddenly due to the proportions of the training data. This indicates that the GRU model is effective in forecasting air pollution. Figure 4 is the visualization of predicted and actual values of pollutants.

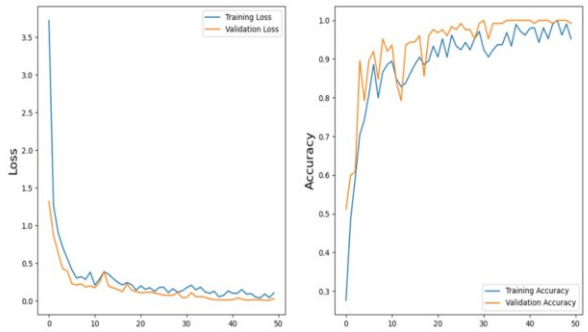


Figure 3. Model learning curve

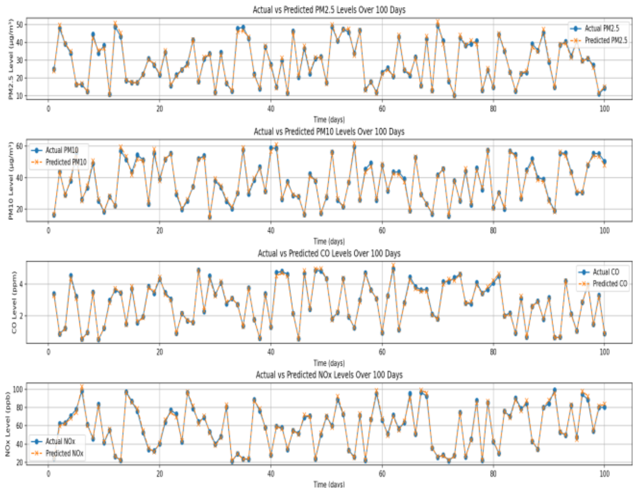


Figure 4. The real versus predicted values of Pollutants

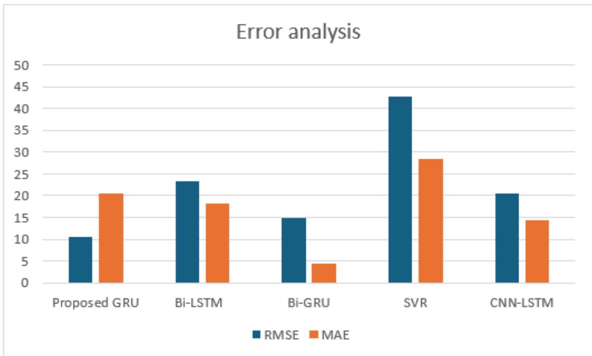


Figure 5. Error rate analysis

The average obtained values of RMSE, MAE, and MSE are given in Table 1. This proposed GRU model achieved the lowest MSE of 113.13 which denotes low prediction errors compared to other models. Its RMSE and MAE values are also relatively low, showing good predictive performance. The RMSE and MAE of the Bi-GRU model are quite good when compared to other models. Next, the CNN-LSTM model achieved 420.8, 20.52 and 14.3 MSE, RMSE and

MAE rates respectively. The SVR model attains the highest MSE among all models which indicates relatively higher prediction errors compared to other models. The results are graphically shown in Figure 5.

Table 1. Performance analysis

Method	MSE	RMSE	MAE
Proposed GRU	113.13	10.63	20.4
Bi-LSTM	549.68	23.43	18.3
Bi-GRU	221.41	14.87	4.35
SVR	1837.61	42.86	28.5
CNN-LSTM	420.8	20.52	14.3

5. Conclusion

A new GRU model for the accurate prediction of air pollution is proposed. The major problem of the GRU model is the optimum selection of hyperparameters. The successful solution for identifying this hyperparameter is achieved through the use of the Spider Wasp Optimization (SWO) algorithm. By integrating the GRU model with the Dual-Slope Leaky ReLU (DSLReLU) activation function, the model exploits the adaptive capabilities of the activation function to better handle the different ranges of pollutant data. The inclusion of the SWO algorithm for fine-tuning the GRU model parameters and the slopes of the DSLReLU function ensures an optimized and robust performance. In future, both data mining and remote sensing imaging techniques applied to improve the prediction accuracy further.

References

Abdel-Basset M., Mohamed R., Jameel M. *et al*. (2023). Spider wasp optimizer: a novel meta-heuristic optimization algorithm. *Artificial Intelligence Review* **56**, 11675–11738.

Bekkar A., Hssina B., Douzi S. *et al*. (2021). Air-pollution prediction in smart city, deep learning approach. Journal Big Data **8**, 161. https://doi.org/10.1186/s40537-021-00548-1

Bhatti U.A. *et al*. (2021). Time Series Analysis and Forecasting of Air Pollution Particulate Matter (PM2.5): An SARIMA and Factor Analysis Approach, in *IEEE Access*, **9**, 41019–41031, doi: 10.1109/ACCESS.2021.3060744.

Binbusayyis A., Khan M.A., Ahmed A M.M. *et al*. (2024). A deep learning approach for prediction of air quality index in smart city. *Discover Sustainability* **5**, 89.

Biswas P., Kar N. and Deb S. (2024). ML based assessment and prediction of air pollution from satellite images during COVID-19 pandemic. *Multimed Tools Application* (2024). https://doi.org/10.1007/s11042-023-18102-x

Chen H., Guan M. and Li H. (2021). Air Quality Prediction Based on Integrated Dual LSTM Model, in *IEEE Access*, **9**, 93285–93297, doi: 10.1109/ACCESS.2021.3093430.

Cong Y., Zhao X., Tang K., Wang G., Hu Y. and Jiao Y. FA-LSTM: A Novel Toxic Gas Concentration Prediction Model in Pollutant Environment, *in IEEE Access*, **10**, 1591–1602, doi: 10.1109/ACCESS.2021.3133497.

Du S., Li T., Yang Y. and Horng S.-J. (2021). Deep Air Quality Forecasting Using Hybrid Deep Learning Framework, in *IEEE Transactions on Knowledge and Data Engineering*, **33**(6), 2412–2424, doi: 10.1109/TKDE.2019.2954510.

Ghufran Isam Drewil, Riyadh Jabbar Al-Bahadili. (2022). Air pollution prediction using LSTM deep learning and metaheuristics algorithms, Measurement: Sensors, **24**, 100546.

Gu K., Liu H., Liu J., Yu X., Shi T. and Qiao J. (2022). Air Pollution Prediction in Mass Rallies with a New Temporally-Weighted Sample-Based Multitask Learner, in *IEEE Transactions on Instrumentation and Measurement*, **71**, 1–15, Art no. 2505915, doi: 10.1109/TIM.2022.3157365.

Haleem A.M. *et al*. (2019). Air quality assessment of some selected Hospitals within Baghdad City, *Engineering and Technology Journal*., **37** (1), 59–63,

Han J., Liu H., Xiong H. and Yang J. (2023). Semi-Supervised Air Quality Forecasting via Self-Supervised Hierarchical Graph Neural Network, in *IEEE Transactions on Knowledge and Data Engineering*, **35**(5), 5230–5243, doi: 10.1109/TKDE.2022.3149815.

Han Y., Lam J.C.K., V. O. K. Li and Zhang Q. (2022). A Domain-Specific Bayesian Deep-Learning Approach for Air Pollution Forecast, in *IEEE Transactions on Big Data*, **8**(4). 1034–1046, 1 Aug. 2022, doi: 10.1109/TBDATA.2020.3005368.

Huang Y., Xiang Y., Zhao R. and Cheng Z. (2020). Air Quality Prediction Using Improved PSO-BP Neural Network, *in IEEE Access*, **8**, 99346–99353, doi: 10.1109/ACCESS.2020.2998145.

Jin N., Zeng Y., Yan K. and Ji Z. (2021). Multivariate Air Quality Forecasting with Nested Long Short Term Memory Neural Network, in *IEEE Transactions on Industrial Informatics*, **17**(12), 8514–8522, doi: 10.1109/TII.2021.3065425.

Kumar K. and Pande B.P. (2023). Air pollution prediction with machine learning: a case study of Indian cities. *International Journal of Environmental Science and Technology*. **20**, 5333–5348.

Li X., Xu X., Wang J., Li J., Qin S. and Yuan J. (2020). Study on Prediction Model of HIV Incidence Based on GRU Neural Network Optimized by MHPSO, *in IEEE Access*, **8**, 49574–49583, doi: 10.1109/ACCESS.2020.2979859.

Liu B. *et al*. (2019). A Sequence-to-Sequence Air Quality Predictor Based on the n-Step Recurrent Prediction, in IEEE Access, 7, 43331–43345, doi: 10.1109/ACCESS.2019.2908081.

Liu B., Yan S., Li J., Li Y., Lang J. and Qu G. (2021). A Spatiotemporal Recurrent Neural Network for Prediction of Atmospheric PM2.5: A Case Study of Beijing, in *IEEE Transactions on Computational Social Systems*, **8**(3), 578–588, June 2021, doi: 10.1109/TCSS.2021.3056410.

Liu C., Pan G., Song D. and Wei H. (2023). Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine, *in IEEE Access*, **11**, 67086–67097, doi: 10.1109/ACCESS.2023.3291146.

Lo J.C.K. *et al*. (2022). Recurrent Learning on PM2.5 Prediction Based on Clustered Airbox Dataset, in *IEEE Transactions on Knowledge and Data Engineering*, **34**(10), 4994–5008, 1 Oct, doi: 10.1109/TKDE.2020.3047634.

Manisalidis E., Stavropoulou A., Stavropoulos E., Bezirtzoglou (2020). Environmental and health impacts of air pollution: a review, Front. PublicHealth, **8**, 10.3389/fpubh.2020.00014

Natarajan S.K., Shanmurthy P., Arockiam D. *et al*. (2024). Optimized machine learning model for air quality index prediction in major cities in India. *Scientific Reports* **14**, 6795. https://doi.org/10.1038/s41598-024-54807-1

Nguyen H.A.D. *et al*. (2023). Long Short-Term Memory Bayesian Neural Network for Air Pollution Forecast, in *IEEE Access*, **11**, 35710–35725, 2023, doi: 10.1109/ACCESS.2023.3265725.

Nguyen M.H., Le Nguyen P., Nguyen K., Le V.A., Nguyen T.-H. and Ji Y. (2021). PM2.5 Prediction Using Genetic Algorithm-Based Feature Selection and Encoder-Decoder Model, *in IEEE Access*, **9**, 57338–57350, doi: 10.1109/ACCESS.2021.3072280.

Pan M. *et al*. (2020). Water Level Prediction Model Based on GRU and CNN, *in IEEE Access*, **8**, 60100, doi: 10.1109/ACCESS.2020.2982433.

Qin D., Yu J., Zou G., Yong R., Zhao Q. and Zhang B. (2019). A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM2.5 Concentration, *in IEEE Access*, **7**, 20050–20059, 2019, doi: 10.1109/ACCESS.2019.2897028.

Singh M.and Dubey R.K. (2023). Deep Learning Model Based CO2 Emissions Prediction Using Vehicle Telematics Sensors Data," in *IEEE Transactions on Intelligent Vehicles*, **8**(1), 768–777, Jan. 2023, doi: 10.1109/TIV.2021.3102400.

Tao Q., Liu F., Li Y. and Sidorov D. (2019). Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU, in *IEEE Access*, **7**, 76690–76698, doi: 10.1109/ACCESS.2019.2921578.

Wu Z., Liu N., Li G., Liu X., Wang Y. and Zhang L. (2023). Learning Adaptive Probabilistic Models for Uncertainty-Aware Air Pollution Prediction, in *IEEE Access*, **11**, 24971–24985, 2023, doi: 10.1109/ACCESS.2023.3247956.

Xu M., Yang Y., Han M., Qiu T. and Lin H. (2019). Spatio-Temporal Interpolated Echo State Network for Meteorological Series Prediction, in *IEEE Transactions on Neural Networks and Learning Systems*, **30**(6), 1621–1634, doi: 10.1109/TNNLS.2018.2869131.

Xu X. and Yoneda M. (2021). Multitask Air-Quality Prediction Based on LSTM-Autoencoder Model, in *IEEE Transactions on Cybernetics*, **51**(5), 2577–2586, doi: 10.1109/TCYB.2019.2945999.

Yang J., Ismail A.W., Li Y., Zhang L. and Fadzli F.E. (2023). Transfer Learning-Driven Hourly PM2.5 Prediction Based on a Modified Hybrid Deep Learning, in *IEEE Access*, **11**, 99614–99627, 2023, doi: 10.1109/ACCESS.2023.3314490.

Yao L. and Abisado M. (2024). Prediction Method of O2O Coupon Based on Multi-Grained Attention Mechanism of CNN and Bi-GRU, *in IEEE Access*, **12**, 16902–16914, doi: 10.1109/ACCESS.2024.3359052.

Yi X., Duan Z., Li R., Zhang J., Li T. and Zheng Y. (2022). Predicting Fine-Grained Air Quality Based on Deep Neural Networks, in *IEEE Transactions on Big Data*, **8**(5), 1326–1339, doi: 10.1109/TBDATA.2020.3047078.

Zeng C., Ma C., Wang K. and Cui Z. (2022). Parking Occupancy Prediction Method Based on Multi Factors and Stacked GRU-LSTM, *in IEEE Access*, **10**, 47361–47370, doi: 10.1109/ACCESS.2022.3171330.