Yugoslav Journal of Operations Research
34 (2024), Number 4, 669-686
DOI: https://doi.org/10.2298/YJOR2402016024A

CIRCLE CHAOTIC MAP TUNA SWARM OPTIMIZATION (CCMTSO) BASED FEATURE SELECTION AND DEEP LEARNING APPROACH FOR AIR QUALITY PREDICTION

Swamy ARADHYAMATADA

Department of Electronics & Communication, Proudha Devaraya Institute of Technology, Hosapete, Visvesveraya Technological University, Belagavi -590018. India
swamyam@gmail.com

Dr. Rohitha U.M

Department of Electronics & Communication, Proudha Devaraya Institute of Technology, Hosapete, Visvesveraya Technological University, Belagavi -590018. India
rohitha_ujjini@rediffmail.com

Received: February 2024 / Accepted: April 2024

Abstract: Air pollution has threatened human life in many countries worldwide due to human activity, industrialization, and urbanization over the past few decades. In air forecasting, particulate matter ($PM_{2.5}$) is a significant health concern. Thus, $PM_{2.5}$ concentrations must be accurately predicted to protect communities from air pollution. This work aims to increase air quality forecasting by predicting their quality. The significant achievement of this work was the design of a new FS (Feature selection) and prediction method for air quality. Circle Chaotic Map Tuna Swarm Optimization (CCMTSO) and FCNN-LSTM (Fully Convolutional Neural Network - Long short-term term Memory) based DL (Deep Learning) have been used to select features and estimate air quality prediction. The FCNN-LSTM algorithm is generated by CCMTSO using previous information from the target station and nearby stations with chosen attributes. The FCNN model uses geographical features to filter out pollution components, meteorological circumstances, and station interactions. Using the training set, the network is trained until convergence once the model's structure has been established. The suggested approach outperforms competing systems regarding the accuracy of $PM_{2.5}$ prediction and effectiveness in extracting spatiotemporal data. Three metrics are employed to assess the efficiency of the proposed framework: Root Mean Squared Error (RMSE), coefficient of determination (R^2), and Mean Absolute Error (MAE). The findings demonstrate that the results achieved by the proposed system are 7.214, 13.437, and 0.961 for MAE, RMSE, and R^2 at a batch size of 128. Utilizing LSTM and FCNN, this algorithm can extract the temporal and spatial components of the information with good precision and reliability.

Keywords: Air-pollution, Circle Chaotic Map Tuna Swarm Optimization (CCMTSO), Deep learning, Forecasting, FCNN-LSTM (Fully Convolutional Neural Network - Long Short Term Memory).

MSC: 68T27, 68W40, 92B20.

1. INTRODUCTION

Air pollution is currently an important issue for the entire world. Air pollution is a problem for most cities. Because of its significant impacts on the health of humans and the ecological balance, air pollution has currently received more attention. Industry in India is one of the primary causes of air pollution [1]. Information on how much industry harms the environment is available from the air pollutants index (Particulate Matter 2.5 $PM_{2.5}$), Nitrogen Dioxide (NO_2), Sulphur Dioxide (SO_2), Carbon Monoxide (CO), and trioxygen/Ozone (O_3) among others). According to the United Nations (UN) [2], around 56.15% of people will reside in urban areas by 2020. In addition, by 2050, 68% of people worldwide are predicted to reside in cities [3]. The problems caused by the increase in urbanization and industry include transportation, health care, and air quality, to name a few. The idea of a "smart city" was developed to address these problems and enhance the well-being of its residents by integrating fixed and mobile sensors with information and communication technology (ICT). These types are placed all across the city to observe genuine human activity. The concept eventually becomes a continuous flow of urban data.

Nearly 90% of people, as per the World Health Organization (WHO), respire air that is contaminated and has a higher quality than the WHO's recommended levels, which can cause respiratory issues [4]; additionally, it includes mortality and occurrences related to cardiovascular disease [5], even with a short exposure to $PM_{2.5}$ of a few hours to a few weeks. According to a study by the OECD (Organization for Economic Cooperation and Development), air pollution may rate the world's GDP (Gross Domestic Product) 1.00%. Li et al. [6] consider the possibility that air pollution may induce reproductive system disorders.

Due to the detrimental impacts on people's health, other forms of life, and the environment, $PM_{2.5}$, the primary air pollutant, has drawn a lot of attention [7]. Numerous studies provide evidence that air pollution harms respiratory and cardiovascular health, resulting in mortality among both animal and plant populations. Additionally, air pollution contributes to acid rain, climate change, and global warming, thereby imposing economic losses and posing challenges to human survival within societies worldwide [8]. Using a comparative comparison of ML approaches, the impacts of $PM_{2.5}$ studied over the last 25 years are discussed; according to Ameer et al. [9], ozone exposure has been linked to an extra 250,000 deaths, and there have been an estimated 4.2 million deaths related to chronic extent to $PM_{2.5}$ in the atmosphere [7]. $PM_{2.5}$ ranked fifth worldwide in terms of risk factors for mortality and was responsible for 7.6% of all fatalities. According to extensive studies, three classic classifications may be used to essentially categorize air pollution forecasting approaches: Artificial intelligence approaches, statistical forecasting techniques, and numerical forecasting techniques.

In view of air pollution's established negative impacts on health, for the purpose of controlling air pollution and preventing health issues, daily $PM_{2.5}$ concentration forecasts is essential. For $PM_{2.5}$ predictions, several researchers have developed original methods. It is really challenging to have 5 generations. Considering it critically is thus necessary.

The air quality data also has a major temporal component, making it a part of time series and giving it the appearance of periodicity. Machine learning algorithms and regression models are the two primary empirical methodologies. Multiple linear regression, land use regression, and autoregressive moving averages are examples of traditional regression models that are straightforward [10]. In contrast, several machine learning models have been extensively used because of their excellent handling of nonlinear interactions, and they often provide sufficient results. Artificial Neural Network (ANN) is an instances of frameworks [11], Recursive Neural Network (RNN) [12], Support Vector Regression (SVR), and hybrid model.

Government decision-making and public health depend on an efficient system for monitoring and forecasting air pollution. RNN and its derivatives have received recent attention due to the growing popularity of artificial intelligence. Because it takes into consideration the time dependencies of characteristic phenomena found in the $PM_{2.5}$ concentration series, when predicting air quality, the most often used model is LSTM [13, 14]. Additionally, these models' properties significantly affect how well they predict outcomes. Therefore, several swarm intelligent optimization techniques have been used to choose features from the dataset, including GA (Genetic Algorithm), PSO (Particle Swarm Optimization), GWO (Gray Wolf Optimization), and CS (Cuckoo Optimization). By examining the feeding habits of tuna populations, Tuna Swarm Optimization (TSO) is presented. TSO can still be further accelerated and converges faster than other traditional meta-heuristics techniques. The global optimum value cannot be successfully found by TSO. The model accuracy is boosted and the robustness is also improved over previous techniques after employing these algorithms to choose the key characteristics. At the same time missing values presented in the dataset is handled by using linear spline imputation.

In this paper, Circle Chaotic Map Tuna Swarm Optimization (CCMTSO), and FCNN-LSTM has been employed to estimate air quality and choose features. Using CCMTSO, specific attributes are fed into the FCNN-LSTM model together with historical data from the target station and neighboring stations. The FCNN model may filter out spatial factors like pollution components and weather data between nearby stations. Results match other cutting-edge deep learning methods.

The overall organization of the study is structured as follows: Section 2 reviews about the details of existing air quality prediction methods, section 3 focuses on the proposed methodology with their steps and mathematical model. In section 4 presents the fallouts comparison of current methods, and suggested system with their evaluation metrics. Finally, the results achieved by proposed system, and findings of the present work is discussed at end of section 5.

2. LITERATURE REVIEW

Du et al. [15] discussed about the development of knowledge pertaining to the utilization of a hybrid deep learning architecture that involves the implementation of a new-edge deep learning algorithm intended to forecast air quality, with a concentration on $PM_{2.5}$ levels. This framework leverages the inherent spatial-temporal correlation characteristics and interrelationships present within multivariate time series data associated with air quality. One-dimensional CNN and Bi-LSTM may be used to learn multivariate air quality time series data. Use the framework to forecast $PM_{2.5}$ air pollution. Results are measured using the metrics like RMSE, and MAE. Han et al. [16]

proposed the particular domain in China and the UK, a Bayesian deep learning model is used to anticipate long-term air pollution. To improve the recommended model's performance, multi-step forecast techniques are coupled based on the appropriate uncertainty metrics. While integrating Bayesian approaches enables the merging of several forecast tactics to increase prediction accuracy, the Bayesian deep learning model lowers prediction errors. Results are measured using the metrics like accuracy, and error.

Hossain et al. [17] presented a prediction model for the air quality index for Chattogram and Dhaka, two of Bangladesh's most polluted cities. The two reliable variants of a RNN are the gated recurrent unit (GRU) and the LSTM. These two concepts are combined in this model. GRU and LSTM were utilized as the model's first and second hidden layers, respectively, before two dense layers. After the data was collected and processed, 80% of the total information was employed for the model training, and it was then tested against the remaining data. Results show that the suggested model can track the patterns in the real AQI for both locations. Results are measured using the metrics like MAE, MSE, and RMSE.

Heydari et al. [18] suggested the LSTM-MVO hybrid model to forecast and quantify combined cycle power plant air pollution. Optimizing LSTM parameters with MVO reduces prediction error. Utilizing information from Kerman, Iran's combined cycle power plant, the model has been assessed. RMSE, MAE, and MAPE (Mean Absolute Percentage Error) have been employed for results comparison.

Bekkar et al. [19] implemented a CNN-LSTM-DL method that incorporates historical pollution data, weather information, and the hourly forecast of $PM_{2.5}$ values in Beijing, China, with $PM_{2.5}$ readings from nearby stations. LSTM, Bi-LSTM, GRU (Gated Recurrent Unit), Bidirectional Gated Recurrent Unit (Bi-GRU), CNN, and a hybrid CNN-LSTM framework have all been utilized in experiments with deep learning algorithms. According to experimental findings, in terms of predictive performance, the "hybrid CNN-LSTM multivariate" method performs better and is able to provide predictions that are more accurate than any of the stated conventional models. Three indicators like MAE, RMSE, and the R^2 has been used for results comparison.

Dairi et al. [20] proposed a versatile and an efficient approach for forecasting ambient pollution concentrations is deep learning-driven. The construction of a variational autoencoder (VAE) and an attention appliance has been accomplished through the utilization of the predictive modelling approach is founded on the advanced Integrated Multiple Directed Attention (IMDA) DL framework. Using information on air pollution from four US states, the suggested forecasting model is then put through experimental validation to see how well it performs. The suggested IMDA-VAE model may successfully enhance the effectiveness of air pollution forecasting, according to the results Validation metrics such as R^2, RMSE, MAE, explained variance (EV), MAPE, Mean Bias Error (MBE), and Relative Mean Bias error (rMBE) are used to evaluate the predicted accuracy as well as contrast the frameworks.

Zhang et al. [21] proposed to forecast changes in $PM_{2.5}$ levels in Chinese cities, researchers have developed the hybrid DL algorithm VMD-BiLSTM, which combines Variational Mode Decomposition (VMD) and BiLSTM. Using VMD based on the frequency domain, the complicated time series of $PM_{2.5}$ data is first divided into several sub-signal components. The forecasting accuracy was then greatly increased through the utilization of BiLSTM to anticipate each sub-signal component individually. The most reliable strategy for predicting is the proposed VMD-BiLSTM model. Chang et al. [22]

suggested an ALSTM (Aggregated LSTM) algorithm depends on the LSTM DL technique. Create a new model that incorporates regional, global, and industrial monitoring stations for air quality. Three LSTM models combined with data from neighboring industrial air quality sensors to build a prediction model and exterior pollution sources will have higher forecast accuracy. The outcomes demonstrate that the suggested aggregated model may significantly raise prediction accuracy. MAE, RMSE, and MAPE are employed for assessing the efficiency.

Kothandaraman et al. [23] proposed several ML models, containing Linear Regression, Random Forest (RF), K-Nearest Neighbor (KNN), ridge, lasso, XGBoost, and AdaBoost, have been utilized for the prediction of $PM_{2.5}$. There are five primary pollutants commonly found in urban areas with high levels of pollution. The new models were shown to perform better and have lower error rates than the current models in predicting the $PM_{2.5}$ pollutants. Drewil and Al-Bahadili[24] suggested a framework built using both the LSTM DL procedure and the Genetic Algorithm (GA) method. Using $PM_{2.5}$, PM_{10}, CO, and NOx as the four different kinds of pollutants, the framework provides the necessary hyperparameters for LSTM and the pollution level for the succeeding day. Compared to LSTM models and machine learning models, the suggested model that has been changed by optimization algorithms displays results that are more accurate with less experience and more rapid.

Some of the works related to deep learning methods with other applications are described as follows: Khodaverdian et al. [25] suggested using GRU and CNN to select the best migration candidate Virtual Machine (VM), resulting in an assessment of whether or not the VM is latency sensitive. The hybrid approach based on CNN and GRU that Khodaverdian et al. [26] presented is employed to classify VMs in the Microsoft Azure cloud service. Abbasimehr and Khodizadeh Nahari [27] proposed LSTM for time series forecasting. RahmatAbadi and Mohammadzadeh [28] suggested DL-based Collaborative Filtering (CF) recommender systems. It follows the procedure of Autoencoders and Restricted Boltzmann Machine (RBM). Kuang et al. [29] utilized the RNN algorithm for sentiment classification.

Some of the works related to deep learning methods and metaheuristic are described as follows: Sharifi et al. [30] suggested a novel model depends on LSTM and Metaheuristic algorithms. Chaotic Dolphin Swarm Optimization algorithm is employed to enhance the LSTM. Dirik [31] proposed a hybrid model based on Interval type-2 fuzzy Logic System (IT2FLS) and PSO for electrocardiogram (ECG) prediction. The effectiveness of the NTS (Neutrosophic Time Series) simulation approach was examined by Edalatpanah et al. [32]. It heavily depends on the universe of discourse and the appropriate intervals that are chosen. The Quantum Optimization Algorithm (QOA), GA, and PSO have been selected for this investigation. To provide a multitude of them, Rajabi Moshtaghi et al. [33] concentrated on meta-heuristic techniques.

The existing methods majorly focus on the air quality prediction depending on the entire dataset, and prediction has been performed based on the DL approaches. The major issues of these approaches are that it significantly affects predict outcomes due to irrelevant number of features, and missing attributes presented in the dataset. The existing feature selection methods had easily support to lowest number of samples, and these methods doesn't majorly focus on the weather attributes. In this work, air quality prediction is performed based on the weather and air attributes. In addition irrelevant number of features presented in the dataset is reduced by CCMTSO.

3. PROPOSED METHODOLOGY

In this paper, Circle Chaotic Map Tuna Swarm Optimization (CCMTSO), and FCNN-LSTM has been introduced for feature selection, and air quality prediction. In addition to a few selected parameters from CCMTSO, the FCNN-LSTM model also collects historical data from the target station and neighboring stations. The suggested approach beats competitors in terms of the precision of $PM_{2.5}$ prediction and efficiency in extracting spatiotemporal properties.

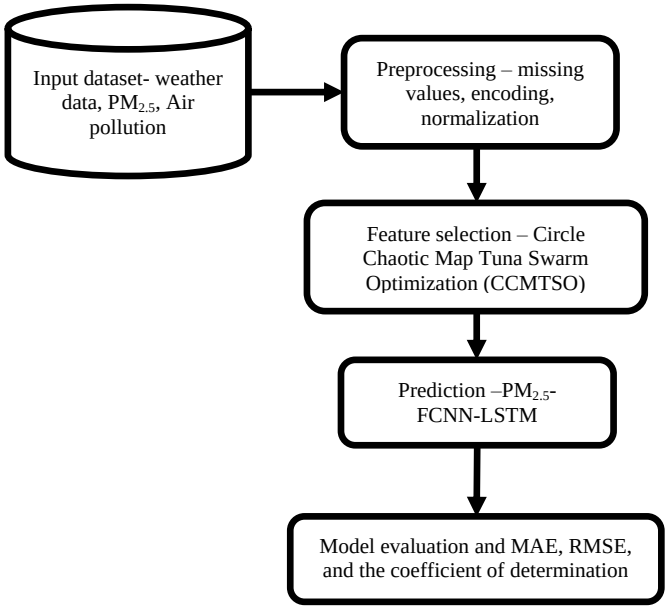

Figure 1: Flow diagram of proposed prediction system

The air quality and air pollution levels at 12 different sites are shown in the dataset. The Beijing Municipal Environmental Monitoring Center's air quality data was made accessible. The China Meteorological Administration's nearest meteorological station's data is compared with the air quality data for each given location. March 1, 2013, to February 28, 2017, are included in the time frame. It is a product of the UCI Machine Learning Repository [34].

3.1. Data preprocessing

35064 records total in this collection include several attributes in each station. From 2013 through 2017, recordings were produced. Date, $PM_{2.5}$, PM_{10}, NO_2, SO_2, CO, O_3, Dew Point, Atmospheric Pressure, Combined Wind Direction, Cumulated Wind Speed, Hours of Snow, and Hours of Rain are included. But owing to machine failure or other unavoidable circumstances, meteorological and air quality monitoring equipment would result in data leakage. Data mining may be affected to some extent by the presence of such missing variables. The mean or median value is the two most often used methods for replacing missing field values in time-independent (non-chronological) data. A time

series, however, is an exception to this rule. There are many different imputation strategies used to tackle issues with partial data. For all simulated missing value percentages, linear interpolation predicts PM_{10}hourly monitoring data optimally. In the processed data set, linear spline imputation was used to fill in less than 4% of the missing values. Without altering the interpolation values at other sites, the SL(x) equation (1) may adjust to local anomalies. Following is an explanation of the equation (1) for the spline linear interpolation function

$$SL(x) = f(x_{i-1})\frac{x-x_i}{x_{i-1}-x_i} + f(x_i)\frac{x-x_{i-1}}{x_i-x_{i-1}}\ x \in [x_{i-1}, x_i], i = 1,2,3, \dots, n \quad (1)$$

where x is a variable that is not a constant, $x_0, x_1, \dots x_n$SL(x), the linear spline's known values that interpolate f at these places.

3.2. Encoding categorical variables

In this experiment, atmospheric activity is measured by the wind factor. The speed of the wind has an impact on the pollution concentration [35], Particularly critical in influencing the $PM_{2.5}$concentrations is the direction of the wind.5. There are 16 values allowed for the category data on wind direction. Convert a number of azimuth degrees to each cardinal wind direction. When moving clockwise, the value of North increased by 22.5 from its initial value of zero.

3.3. Normalization

Normalize $PM_{2.5}$concentration estimates using the Min-Max normalization, as described in equation (2), to increase forecast accuracy,

$$x = \frac{x-min}{max-min}. \quad (2)$$

The terms "min and max" stand for the characteristics' minimum and maximum values.

3.4. Feature selection by Circle Chaotic Map Tuna Swarm Optimization (CCMTSO)

The selection of features is a crucial stage in machine learning applications and may be accomplished in a number of ways. The tuna is the primary predator in water environments. The parabolic foraging strategy and the spiral foraging strategy are two effective predatory strategies used by the tuna swarm. Each tuna will closely follow the preceding one to select the best air prediction features when the swarm utilizes parabolic foraging. The tuna swarm encloses its target in a parabola. When the tuna swarm used the spiral foraging technique, it directed prey into areas of shallow water, grouping into spiral formations. Based on the precision of the air quality forecast, prey is more likely to be caught.

Population Initialization: There are NP is the population size of tunas in a tuna swarm has been considered based on the amount of samples in the air quality dataset. The TSO algorithm creates the first swarm in the feature selection search space at random during the swarm initiation phase. Initializing tuna individuals requires solving equation (3) [36],

$$X_i^{int} = rand.(ub - lb) + lb = \left[x_i^1, \ldots . x_i^j\right], i = 1, \ldots, NP, j = 1,2, \ldots ., Dim \quad (3)$$

where $X_i^{int}$is the i[th] tuna, ub and lb are feature selection, tuna exploration upper and lower limitations are being evaluated. From 0 to 1, the random variable "Rand" has a uniform distribution. In specifically, each person, X_i^{int}indicates a potential TSO solution in the tuna swarm. Dim-dimensional numbers make up each individual tuna.

Parabolic Foraging Strategy: Accuracy will serve as a benchmark for the tuna swarm as it continues to pursue prey in search of its greatest attributes. Each tuna mimics the behavior of the one before it during predation, forming a parabola to encircle the target. Additionally, the tuna swarm employs a spiral foraging technique. Equation (4-5) describes the tuna swarm's parabolic foraging behavior under the assumption that there is a 50% chance that it will choose either method [36],

$$X_i^{t+1} = \{X_{best}^t + rand.(X_{best}^t - X_i^T) + TF.pw^2.(X_{best}^t - X_i^t), if\ rand < 0.5TF.pw^2.X_i^t, if\ rand \geq 0.5 \quad (4)$$

$$pw = \left(1 - \frac{t}{t_{max}}\right)^{\left(\frac{t}{t_{max}}\right)} \quad (5)$$

Here, maximum number of iterations can be represented as t_{max} and the current iteration is t. That may be selected as the maximum for a feature, in this case, *t*. The random value of TF is either 1 or -1.

Spiral Foraging Strategy: The majority of tuna cannot decide which direction to swim in while pursuing their prey (accuracy), but a tiny percentage of fish may direct the swarm. The neighbouring tuna will follow them (features) when a small group of tuna starts to search for their prey (accuracy). In order to successfully capture its victim, the whole tuna swarm will eventually create a spiral configuration. When a spiral foraging strategy is used by the tuna swarm, features will communicate with one another to choose which feature will follow samples or other nearby features. The most advantageous trait may not always be able to successfully direct the swarm to acquire prey. A random feature in the swarm will then be chosen by the tuna to follow. The spiral foraging method is mathematically represented by the following equation (6) [36],

$$X_i^{t+1} = \begin{cases} \alpha_1.(X_{rand}^t + \tau.|X_{rand}^t - X_i^t| + \alpha_2.X_i^t), if\ rand < \frac{t}{t_{max}}, i = 1 \\ \alpha_1.(X_{rand}^t + \tau.|X_{rand}^t - X_i^t| + \alpha_2.X_{i-1}^t), if\ rand < \frac{t}{t_{max}}\ \ i = 2,3, \ldots, NP \\ \alpha_1.(X_{best}^t + \tau.|X_{best}^t - X_i^t| + \alpha_2.X_i^t), if\ rand \geq \frac{t}{t_{max}}, i = 1 \\ \alpha_1.(X_{best}^t + \tau.|X_{best}^t - X_i^t| + \alpha_2.X_{i-1}^t), if\ rand \geq \frac{t}{t_{max}}, i = 2,3, \ldots, NP \end{cases} \quad (6)$$

where X_i^{t+1}denotes the i[th]tuna in the t + 1 iteration. The current best feature is X_{best}^t. $X_{rand}^t$is a spot in the tuna swarm that was arbitrarily chosen as a reference. 1 is the trend weight coefficient used to direct each tuna to swim to the ideal feature or a set of randomly chosen neighbouring features α_2 is the tuna swimming to the feature in front of it controlled by the trend weight coefficient. τ determines the distance between the tuna and either a randomly chosen reference feature or the ideal feature. According to their mathematical calculation methodology [36],

$$\alpha_1 = a + (1 - a).\frac{t}{t_{max}} \quad (7)$$

$$\alpha_2 = (1-a) - (1-a).\frac{t}{t_{max}} \tag{8}$$

$$\tau = e^{bl}.cos(2\pi b) \tag{9}$$

$$l = e^{3coscos\left(\left(\left(t_{max}+\frac{1}{t}\right)-1\right)\pi\right)} \tag{10}$$

where b is a random number equally distributed between [0, 1] and c is the tuna degree. In the iterative TSO algorithm process, each tuna will alternate between using the parabolic and spiral foraging strategies. The population of tuna individuals (features) is continuously updated while the TSO algorithm runs until the number of iterations approaches a predefined value. Last but not least, the TSO algorithm delivers the population's best characteristic together with its best value.

The CCMTSO algorithm, depends on the Circle Chaotic Map and Adaptive Weight, has been presented. The variety of the swarm may be increased by employing the Circle chaotic map for population initiation. To change the tuna-following weight coefficient, the CCMTSO introduces a nonlinear adaptive weight operator. Iterative processes in CCMTSO have a well-balanced interaction between local exploitation and global exploration. Although chaos exhibits random behavior, it follows certain principles, allowing the TSO feature search space to show more states with this operator [37]. Since the position of the tuna is created at random during the tuna algorithm's starting phase, it is straightforward to get the initial tuna to congregate in one place.

Circle chaotic map: There is a little discrepancy between tuna and the features that tuna chose because the first tuna swarm covers just part of the feature search area. The Circle chaotic mapping technique's mathematical description was improved to provide a more uniform chaotic value distribution. Circle chaotic map's mathematical modelling looks like the following [37]:

$$x_{i+1} = mod\left(3.85x_i + 0.4 - \left(\frac{0.7}{3.85\pi}\right) sin\, sin(3.85\pi x_i), 1\right) \tag{11}$$

where x_i is the ith chaotic particle and x_{i+1} is the (i + 1)th chaotic particle. To improve TSO and provide more consistent candidate feature solutions, the circle chaotic map operator has been developed. The population diversity of TSO may be considerably increased by evenly distributing the algorithm's feature search space's initial tuna individuals.

Nonlinear Adaptive Weight: It's vital to recognize the way of striking a balance between your exploration capacity and the algorithm's exploration capacity. The TSO algorithm heavily relies on weight parameters. The degree to which tuna characteristics follow the ideal individual to forage is determined by the weight factors 1 and 2 in Equations (7) and (8) when the spiral foraging method is used by the tuna. Two enhanced nonlinear weight parameters$\alpha_{1i}$and $\alpha_{2i}(t)$in this article are presented. These are their mathematical representations [37]:

$$\alpha_{1i}(t) = \alpha_{1ini} - \left(\alpha_{1ini} - \alpha_{1fin}\right).sin\, sin\left(\frac{t}{\mu.T_{max}} - \pi\right) \tag{12}$$

$$\alpha_{2i}(t) = \alpha_{2ini} - (\alpha_{2ini} - \alpha_{2fin}).\sin\left(\frac{t}{\mu.T_{max}} - \pi\right) \tag{13}$$

where $\mu = 2$, α_{1ini} indicates α_1, α_{1fin} indicates α_1, α_{2ini} indicates α_2, and α_{2fin} indicates α_2 ultimate value. In the experiment, $T_{max} = 100$, α_{1i}&α_{2i} is the improved version of $\alpha_1$and α_2.

New spiral foraging strategy: A novel nonlinear weight parameter is used in the spiral foraging method $pw_i$is suggested to enhance TSO's performance. Here is its mathematical representation [37],

$$pw_i(t) = pw_{ini} - \left(pw_{ini} - pw_{fin}\right).sin\ sin\left(\frac{t}{\mu.T_{max}} - \pi\right) \tag{14}$$

where pw_{ini}shows the value of pw at start, and $pw_{fin}$is the final value that p. The enhanced weight parameter is comparablepw_i^2combined with the original weight parameter pw^2 .The following is the nonlinear adaptive weight strategy and enhanced spiral foraging method's mathematical model [37],

$$X_i^{t+1} = \begin{cases} \alpha_{1i}.(X_{rand}^t + \tau.|X_{rand}^t - X_i^t| + \alpha_{2i}.X_i^t),\ if\ rand < \frac{t}{t_{max}}, i = 1 \\ \alpha_{1i}.(X_{rand}^t + \tau.|X_{rand}^t - X_i^t| + \alpha_{2i}.X_{i-1}^t),\ if\ rand < \frac{t}{t_{max}}\ \ i = 2,3,\dots,NP \\ \alpha_{1i}.(X_{best}^t + \tau.|X_{best}^t - X_i^t| + \alpha_{2i}.X_i^t),\ if\ rand \geq \frac{t}{t_{max}}, i = 1 \\ \alpha_{1i}.(X_{best}^t + \tau.|X_{best}^t - X_i^t| + \alpha_{2i}.X_{i-1}^t),\ \ if\ rand \geq \frac{t}{t_{max}}, i = 2,3,\dots,NP \end{cases} \tag{15}$$

New parabolic foraging strategy: Based on a nonlinear adaptive weight system, this is the mathematical model of improved parabolic foraging [37],

$$X_i^{t+1} = \{X_{best}^t + \alpha.(X_{best}^t - X_i^T) + TF.pw^2.(X_{best}^t - X_i^t), if\ rand < 0.5TF.pw^2.X_i^t, if\ rand \geq 0.5 \tag{16}$$

ALGORITHM 1. PSEUDOCODE OF CCMTSO ALGORITHM

Initialization: Fix the parameters NP, Dim, a, z and TMax.

Initialize the feature position of tuna Xi (i = 1, 2, . . . , NP) by circle chaotic map (equation (11)

Counter t = 0

while $T < T_{Max}$ do

Compute the fitness value of all tuna

Update the feature position and value of the best tuna X_{best}^t

Nonlinear Adaptive Weight α_{1i}, α_{2i}updating by equations (12-13)

New spiral foraging strategy pw_i by equation (14)

if (rand < z) then

Update $X_i^{t+1}$by circle chaotic map(equation (11))

else if (rand ≥ z) then

if (rand < 0.5) then

Update $X_i^{t+1}$by new spiral foraging strategy (equation (15))

else if (rand ≥ 0.5) then

Update $X_i^{t+1}$by new parabolic foraging strategy (equation (16))

t = t + 1

return the best fitness value f (X_{best}) and the best tuna X_{best}

3.5. $PM_{2.5}$ predictions by FCNN-LSTM classifier

A type of RNN called LSTM can keep track of information from previous events in order to anticipate what will happen in the future. Since LSTM have longer memory, they can learn from delayed inputs. LSTM have output, input, and a forget gate that deletes obsolete data. The sigmoid function has 0 -1 range, supports these three analog gates. Figure 2 below shows these three sigmoid gates. Hyperparameters of LSTM classifier has been tuned by Adam Optimizer. Overfitting in LSTM can be handled by using two strategies: Dropout and Weight Decay. Dropout randomly excludes some nodes during training, which helps prevent overfitting by changing the network topology. Weight Decay evaluates smaller weights between neurons, which also helps prevent overfitting. A horizontal line that traverses the cell indicates the state of the cell [38].

Input gate $i_t = \sigma\left(W^{(it)}\underline{x_t} + W^{(it)}h_{t-1}\right)$ (17)

Forgot gate $f_t = \sigma\left(W^{(if)}\underline{x_t} + W^{(hf)}h_{t-1}\right)$ (18)

Output gate $i_t = \sigma\left(W^{(io)}\underline{x_t} + W^{(ho)}h_{t-1}\right)$ (19)

Process Input $\tilde{C_t} = tanh\left(W^{(i\underline{c})}\underline{x_t} + W^{(h\underline{c})}h_{t-1}\right)$ (20)

Cell update$C_t = f_t * C_{t-1} + i_t * \tilde{C_t}$ (21)

Output:$y_t = h_t = O_t + i_t * tanh(C_t)$ (22)

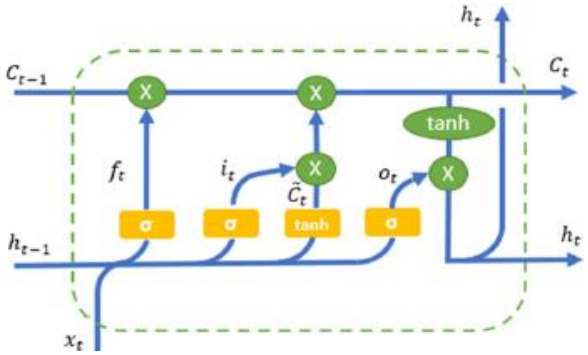

Figure 2: Architecture of LSTM Cell

Fully Convolutional Neural Network (FCNN): In this work, Fully Convolutional Neural Network (FCN) has been introduced for air quality prediction. In a convent, each layer of data is represented as a one-dimensional array of a specific size. The initial layer pertains to the data. Convents are constructed based on the principle of translation invariance. The fundamental constituents of the system, namely convolution, pooling, and activation functions, function by processing localized input data with i and j dimensions (samples, and attributes). In writing x_{ij} for the data vector at i, j (samples and attributes) in a particular layer, and y_{ij} for the layer after that, these functions generate the outputs y_{ij} using equation (23),

$$y_{ij} = f_{ks}(\{X_{si+\delta i,sj+\delta j}\}0 \leq \delta i, \delta j \leq k \quad (23)$$

When the stride or subsampling factor s, the kernel size k, and the layer type f_{ks}. A spatial maximum, matrix multiplication for convolution or average pooling, element wise nonlinearity for activation functions, etc. for max pooling. According to the transformation rule, kernel size and stride, during composition, this functional form is maintained.

$$f_{ks} \circ g_{k's'} = (f \circ g)_{k'} + (k-1)s', ss' \tag{24}$$

A net with only these layers calculates a nonlinear filter or entirely convolutional network, while a standard deep net calculates a generalized nonlinear function. An FCNN automatically processes any size input to produce an identically sized output. An FCN-composed loss function specifies a job. If the last layer's dimensions are added together to form the loss function,

$$\ell(x;\theta) = \sum_{ij} \ell'(x_{ij};\theta) \tag{25}$$

Each of its components' gradients will be added together to form its gradient. As a consequence, stochastic gradient descent on l will equal stochastic gradient descent l' when calculated on all data using a minibatch to capture all of the final layer receptive fields.

FCNN-LSTM: In this design, the output layer retains the LSTM cell rather than merely a 1-D vector since it incorporates a convolution procedure and an input size of information. At each gate of a traditional LSTM, a convolution operation takes the role of matrix multiplication. The FCNN and LSTM Network's capabilities are combined by the ConvLSTM architecture. Long-term Recurrent Convolutional Network (LRCN) model is known as Convolutional LSTM (FCNN-LSTM). The initial stage of this model involves the utilization of convolutional layers to extract fundamental features from the input data. These extracted features are then transformed into a one-dimensional tensor through the process of flattening. This enables them to serve as input for the subsequent phase of the model. Data must finally be reformed into its original form before being sent through the last concealed layer. Figure 3 displays the FCNN-LSTM's architectural layout.

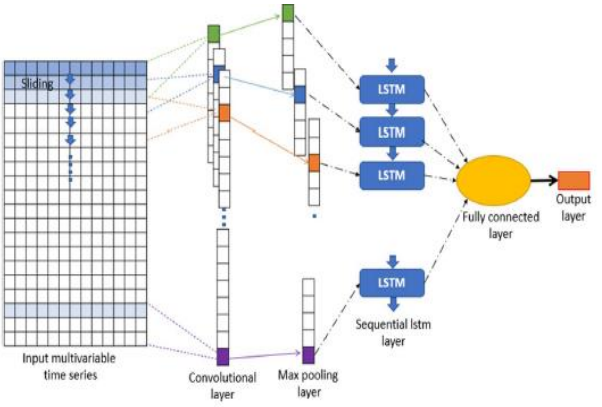

Figure 3: Structural Design of FCNN-LSTM Model

The structural design of FCNN-LSTM, initially features selected from the CCMTSO is given as input to the input time series dataset. From the selected features then feature

vectors are formed at the convolution layer. Then pooling layer is used to perform pooling operation between the feature vectors to make final prediction. The pooling operation performed by pooling layer is given as input the LSTM layer. The feature a vector from pooling layer is again trained using the LSTM, and then results of these layers are combined by fully connected layer. Results combined by the fully connected layer are produced at the output layer. It has been used for air quality prediction. Totally 75% of samples are employed for training, and rest of samples will be testing.

4. RESULTS AND DISCUSSION

Models' evaluation score: Using the training set, the network is trained until convergence when the framework's structure is introduced. This article uses the R^2, RMSE, and the MAE to assess the framework's efficiency.

MAE: MAE, which may more accurately reflect the current state of prediction error, is the arithmetic mean of all samples' deviations between actual and model prediction values. The following is the solution to equation (26):

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{26}$$

RMSE: The RMS of all errors is known as the RMSE. It could be an accurate reflection of the forecast error. Below is a display of equation (27),

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{27}$$

R^2: The percentage of all dependent variable fluctuations that the independent variable can account for by way of the regression connection is represented by the coefficient of determination. The independent variable could be able to better explain the dependent variable. R^2 value that is closer to 1. The formula for equation (28) is given below,

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{28}$$

The sample size, n, is used in these equations (26-28), $y_i$and $\hat{y}_i$reflect, respectively, the actual value and the expected value at the moment; $\underline{y_i}$the average of all genuine values is shown. Table 1 presents a summary of the MAE, RMSE, and R^2 values pertaining to the model's predictions of airborne $PM_{2.5}$concentrations. Among the models examined, it was observed that the model labeled 128 showed the lowest RMSE values. A further finding from these data is that the CNN-LSTM with a batch size of 128 is better accuracy across all measures. Figures 4, 5, 6 shows the results comparison of MAE, RMSE, and R^2 in 24, 32, 64, and 128 Batch size for deep learning methods. These numbers are within the range of $PM_{2.5}$ concentrations that are expected and actual.

K-fold CV (Cross-Validation) is a method used in this study to assess predictive frameworks. The dataset is split up into folds, or subsets, of size k=10. A distinct fold is used as the validation set for each of the k training and evaluation rounds of the framework. To determine the final effectiveness of the approach, performance metrics from each fold in Table 1 are averaged.

Table 1: Results of different models VS. Evaluation metrics

Batch/ models	MAE				
	GRU	**CNN**	**Bi-LSTM**	**CNN-GRU**	**FCNN-LSTM**
24	10.736	9.689	9.265	8.765	7.895
32	10.263	9.425	8.781	8.414	7.536
64	10.892	9.541	9.354	8.562	7.455
128	10.164	9.246	9.157	8.215	7.214
Batch/ models	**RMSE**				
	GRU	**CNN**	**Bi-LSTM**	**CNN-GRU**	**FCNN-LSTM**
24	18.764	16.952	17.597	15.496	14.546
32	17.263	15.449	16.976	15.282	14.128
64	17.652	16.495	17.262	16.834	14.908
128	18.256	15.793	17.845	15.961	13.437
Batch/ models	**R_2**				
	GRU	**CNN**	**Bi-LSTM**	**CNN-GRU**	**FCNN-LSTM**
24	0.902	0.911	0.922	0.928	0.932
32	0.909	0.922	0.931	0.942	0.949
64	0.916	0.931	0.939	0.948	0.954
128	0.924	0.939	0.946	0.955	0.961

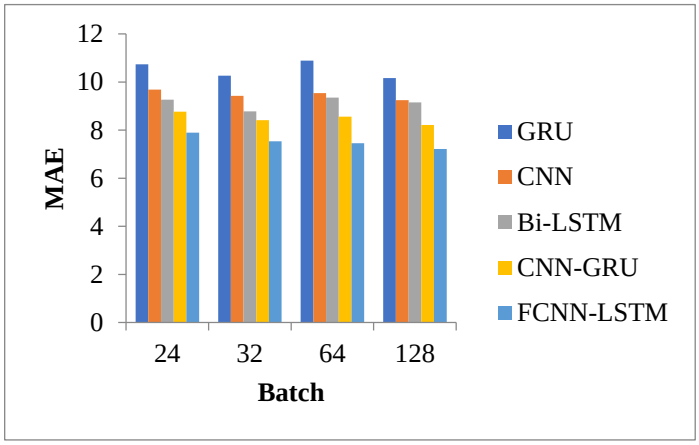

Figure 4. MAE comparison VS. different deep learning models

Figure 4 shows the MAE comparison of different learning methods like GRU, CNN, Bi-LSTM, CNN-GRU, and proposed FCNN-LSTM system. The results are measured by varying different sizes like 24,32, 64, and 128. From the results it shows that the proposed FCNN-LSTM system has lesser MAE of 7.895, 7.536, 7.455, and 7.214 for 24, 32, 64, and 128 batch sizes. The other methods like GRU, CNN, Bi-LSTM, and CNN-

GRU gives higher MAE of 10.164, 9.246, 9.157, and 8.215 for 128 batch size (Refer Table 1).

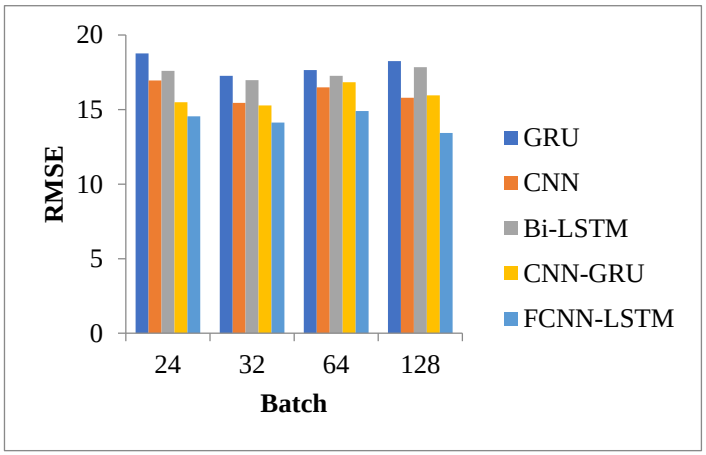

Figure 5. Different deep learning models VS. RMSE comparison

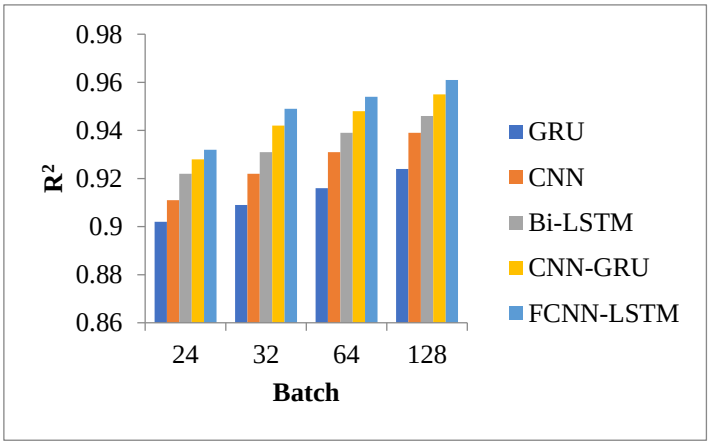

Figure 6: Different deep learning models VS. R^2 comparison

RMSE comparison of different learning methods like GRU, CNN, Bi-LSTM, CNN-GRU, and proposed FCNN-LSTM system are illustrated in figure 5. The results are measured by varying different sizes like 24,32, 64, and 128. FCNN-LSTM system has lesser RMSE of 14.546, 14.128, 14.908, and 13.437 for 24,32, 64, and 128 batch sizes. The other methods like GRU, CNN, Bi-LSTM, and CNN-GRU gives higher RMSE of 18.256, 15.793, 17.845, and 15.961 for 128 batch size (Refer Table 1).

R^2 comparison of different learning methods like GRU, CNN, Bi-LSTM, CNN-GRU, and proposed FCNN-LSTM system by varying different sizes like 24,32, 64, and 128 are illustrated in figure 6. FCNN-LSTM system has lesser R^2 of 0.932, 0.949, 0.954, and 0.961 for 24,32, 64, and 128 batch sizes. The other methods like GRU, CNN, Bi-LSTM, and CNN-GRU gives higher R^2 of 0.924, 0.939, 0.946, and 0.955 for 128 batch size (Refer Table 1).

5. CONCLUSION AND FUTURE WORK

For predicting the $PM_{2.5}$ of air pollutants in the metropolitan region, a hybrid model based on FCNN-LSTM was utilized in this study. The stations' historical data were first examined for feature selection. CCMTSO is based on the swarm, which may be increased by employing the Circle chaotic map for population initiation. A nonlinear adaptive weight operator has also been developed to change the weight coefficient of the tuna-following behavior in CCMTSO. With CCMTSO, the link between international research and local applications is well-balanced for feature selection. FCNN was included in the suggested hybrid framework to efficiently extract the internal and spatial properties of various parameters. LSTM gathered temporal information and created a more accurate and constant prediction result. As a result, more knowledge is gained for predicting air pollution trends. The results also confirm that the GRU, CNN, Bi-LSTM, CNN-GRU, and proposed FCNN-LSTM systems have been experimented using evaluation metrics like MAE, RMSE, and R^2. FCNN-LSTM system can be considered new and superior for air quality prediction over different batch sizes. Future work should focus on improving performance using stacking ensemble with hyperparameter optimization for forecasts.

Conflict of interest statement. No conflicts of interest have been revealed by the author.

Funding. This research received no external funding.

REFERENCES

[1] N.N. Maltareand S. Vahora, “Air Quality Index prediction using machine learning for Ahmedabad city”, *Digital Chemical Engineering*, 7, pp. 100093, 2023.

[2] “Urban population (% of total population), 2018” Available at: https://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS& “Urban Population Change, 2018”, Available at: https://www.un.org/development/desa/pd/

[3] “Nada Osseiran, Christian Lindmeier: 9 out of 10 people worldwide breathe polluted air, but more countries are taking action”, 2018. Available at: https://www.who.int/news/item/02-05-2018-9-out-of-10-people-worldwide-breathe-polluted-air-but-more-countries-are-taking-action

[4] J.A. Ailshire and E.M Crimmins,“Fine particulate matter air pollution and cognitive function among older US adults,” *American journal of epidemiology*, vol. 180, no. 4, pp. 359-366, 2014.

[5] S.AAram, E.A. Nketiah, B.M. Saalidong, H. Wang, A.R Afitiri, A.B. Akotoand P.O. Lartey, “Machine learning-based prediction of air quality index and air quality grade: A comparative analysis,”*International Journal of Environmental Science and Technology*, vol 21, no 2, pp. 1345-1360, 2024.

[6] Y. Li, J. Du, S. Lin, H. He, R. Jiaand W. Liu,“Air pollution increased risk of reproductive system diseases: a 5-year outcome analysis of different pollutants in different seasons, ages,

and genders", *Environmental Science and Pollution Research*, pp. 7312–7321, 2022.
[7] L. Bai, J. Wang, X. Maand H. Lu, "Air pollution forecasts: An overview", *International journal of environmental research and public health*, vol. 15, no. 4, pp. 1-44, 2018.
[8] J. Wang, H. Jiang, Q. Zhou, J. Wuand S. Qin, "China's natural gas production and consumption analysis based on the multicycle Hubbert model and rolling Grey model", *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 1149-1167, 2016.
[9] S. Ameer, M.A. Shah, A. Khan, H. Song, C. Maple, S.U. Islamand M.N. Asghar, "Comparative analysis of machine learning techniques for predicting air quality in smart cities", *IEEE access*, vol. 7, pp. 128325-128338, 2019.
[10] X. Wangand B. Wang, "Research on prediction of environmental aerosol and PM2. 5 based on artificial neural network", *Neural Computing and Applications*, vol. 31, no. 12, pp. 8217-8227, 2019.
[11] X. Feng, Q. Li, Y. Zhu, J. Hou, L. Jinand J. Wang, "Artificial neural networks forecasting of PM2. 5 pollution using air mass trajectory based geographic model and wavelet transformation", *Atmospheric Environment*, vol. 107, pp. 118-128, 2015.
[12] F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo, S. Bianco, S. Di Tommaso, C. Colangeli, G. Rosatelliand P. Di Carlo, "Recursive neural network model for analysis and forecast of PM10 and PM2. 5", *Atmospheric Pollution Research*, vol. 8, no. (4, pp. 652-659, 2017.
[13] A.G. Salman, Y. Heryadi, E. Abdurahmanand W. Suparta, "Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting", *Procedia Computer Science*, vol. 135, pp.89-98, 2018.
[14] Y.T. Tsai, Y.R. Zengand, Y.S. Chang, "Air pollution forecasting using RNN with LSTM", In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress* (DASC/PiCom/DataCom/ CyberSciTech), pp. 1074-1079. IEEE.Athens, Greece, 2018.
[15] S. Du, T. Li, Y. Yangand S.J. Horng, "Deep air quality forecasting using hybrid deep learning framework", *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 6, pp.2412-2424, 2019.
[16] Y. Han, J.C. Lam, V.O. Liand Q. Zhang, "A domain-specific Bayesian deep-learning approach for air pollution forecast", *IEEE Transactions on Big Data*, vol. 8, no. 4, pp.1034-1046, 2020.
[17] E. Hossain, M.A.U. Shariff, M.S. Hossainand K. Andersson, "A novel deep learning approach to predict air quality index", In *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*, pp. 367-381. Singapore: Springer Singapore. Dhaka, Bangladesh, 2020.
[18] A. Heydari, M. Majidi Nezhad, D. Astiaso Garcia, F. Keyniaand L. De Santoli, "Air pollution forecasting application based on deep learning model and optimization algorithm", *Clean Technologies and Environmental Policy*, pp. 607–62, 2022.
[19] A. Bekkar, B. Hssina, S. Douziand K. Douzi, "Air-pollution prediction in smart city, deep learning approach", *Journal of big Data*, vol. 8, pp.1-21, 2021.
[20] A. Dairi, F. Harrou, S. Khadraouiand Y. Sun, "Integrated multiple directed attention-based deep learning for improved air pollution forecasting", *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp.1-15.2021.
[21] Z. Zhang, Y. Zengand K.Yan, "A hybrid deep learning technology for PM 2.5 air quality forecasting", *Environmental Science and Pollution Research*, vol. 28, pp. 39409-39422, 2021.
[22] Y.S. Chang, H.T. Chiao, S. Abimannan, Y.P. Huang, Y.T. Tsaiand K.M Lin, "An LSTM-based aggregated model for air pollution forecasting", *Atmospheric Pollution Research*, vol. 11, no. 8, pp.1451-1463,2020.
[23] D. Kothandaraman, N. Praveena, K. Varadarajkumar, B. Madhav Rao, D. Dhabliya, S.Satla, and W. Abera, "Intelligent forecasting of air quality and pollution prediction using machine

learning", *Adsorption Science & Technology*, pp. 1-15, 2022.
[24] G. I. Drewiland, R. J. Al-Bahadili, "Air pollution prediction using LSTM deep learning and metaheuristics algorithms", *Measurement: Sensors*, vol. 24, pp.100546, 2022.
[25] Z. Khodaverdian, H. Sadrand, S. A. Edalatpanah, "A shallow deep neural network for selection of migration candidate virtual machines to reduce energy consumption", In *2021 7th International conference on web research* (ICWR), pp. 191-196. IEEE. Tehran, Iran, 2021.
[26] Z. Khodaverdian, H. Sadr, S.A. Edalatpanahand, M. Nazari, "An energy aware resource allocation based on combination of CNN and GRU for virtual machine selection", *Multimedia Tools and Applications*, vol. 83, no. 9, pp.25769-25796, 2024.
[27] H. Abbasimehrand, M. Khodizadeh Nahari, "Improving demand forecasting with LSTM by taking into account the seasonality of data", *Journal of applied research on industrial engineering*, vol. 7, no. 2, pp.177-189, 2020.
[28] A.F. RahmatAbadiand, J. Mohammadzadeh, "Leveraging deep learning techniques on collaborative filtering recommender systems", *arXiv preprint arXiv: 2304.09282*, pp. 1-24, 2023.
[29] M. Kuang, R. Safa, S.A. Edalatpanahand, R.S. Keyser, "A Hybrid Deep Learning Approach for Sentiment Analysis in Product Reviews", *Facta Universitatis, Series: Mechanical Engineering*, vol. 21, no. 3, pp.479-500, 2023.
[30] A.M. Sharifi, K. Khalili Damghani, F. Abdiand, S. Sardar, "A hybrid model for predicting bitcoin price using machine learning and metaheuristic algorithms", *Journal of applied research on industrial engineering*, vol. 9, no. 1, pp.134-150, 2022.
[31] M. Dirik, "Type-2 fuzzy logic controller design optimization using the PSO approach for ECG prediction", *Journal of fuzzy extension and applications*, vol. 3, no. 2, pp.158-168, 2022.
[32] S.A. Edalatpanah, F.S. Hassani, F. Smarandache, A. Sorourkhah, D. Pamucar, B. Cui, "A hybrid time series forecasting method based on neutrosophic logic with applications in financial issues", *Engineering applications of artificial intelligence*, vol. 129, pp.107531, 2024.
[33] H. Rajabi Moshtaghi, A. Toloie Eshlaghyand, M.R. Motadel, "A comprehensive review on meta-heuristic algorithms and their classification with novel approach", *Journal of Applied Research on Industrial Engineering*, vol. 8, no. 1, pp.63-89, 2021.
[34] S. Zhang, B. Guo, A. Dong, J. He, Z. Xuand, S.X Chen, "Cautionary tales on air-quality improvement in Beijing. Proceedings of the Royal Society A: Mathematical", *Physical and Engineering Sciences*, vol. 473, no. 2205, pp. 1-14, 2017.
[35] P. Shi, G. Zhang, F. Kong, D. Chen, C. Azorin-Molina, J.A. Guijarro, "Variability of winter haze over the Beijing-Tianjin-Hebei region tied to wind speed in the lower troposphere and particulate sources", *Atmospheric research*, vol. 215, pp.1-51, 2019.
[36] L. Xie, T. Han, H. Zhou, Z.R. Zhang, B. Han, A. Tang, "Tuna swarm optimization: a novel swarm-based metaheuristic algorithm for global optimization", *Computational intelligence and Neuroscience*, pp.1-22, 2021.
[37] W. Wang, J. Tian, "An improved nonlinear tuna swarm optimization algorithm based on circle chaos map and levy flight operator", *Electronics*, vol. 11, 11, no. 22, 3678, 2022.
[38] K. Smagulovaand A.P. James, "A survey on LSTM memristive neural network architectures and applications", *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313-2324, 2019.