

## Article

# Optimizing Long Short-Term Memory Network for Air Pollution Prediction Using a Novel Binary Chimp Optimization Algorithm

Sahba Baniasadi <sup>1</sup>, Reza Salehi <sup>2</sup>, Sepehr Soltani <sup>1</sup> , Diego Martín <sup>2,\*</sup> , Parmida Pourmand <sup>3</sup>  and Ehsan Ghafourian <sup>4</sup>

<sup>1</sup> Department of Industrial Engineering, College of Engineering, University of Houston, Houston, TX 77204, USA; sbaniasa@cougar.net.uh.edu (S.B.); ssoltani2@uh.edu (S.S.)

<sup>2</sup> ETSI de Telecomunicación, Universidad Politécnica de Madrid, Av. Complutense 30, 28040 Madrid, Spain; reza.s2083@gmail.com

<sup>3</sup> Department of Management, FH Wien der WKW University of Applied Sciences for Management & Communication, 1180 Vienna, Austria; parmida.pourmand@edu.fh-wien.ac.at

<sup>4</sup> Department of Computer Science, Iowa State University, Ames, IA 50011, USA; ehsang@iastate.edu

\* Correspondence: diego.martin.de.andres@upm.es

**Abstract:** Elevated levels of fine particulate matter (PM<sub>2.5</sub>) in the atmosphere present substantial risks to human health and welfare. The accurate assessment of PM<sub>2.5</sub> concentrations plays a pivotal role in facilitating prompt responses by pertinent regulatory bodies to mitigate air pollution. Additionally, it furnishes indispensable information for epidemiological studies concentrating on PM<sub>2.5</sub> exposure. In recent years, predictive models based on deep learning (DL) have offered promise in improving the accuracy and efficiency of air quality forecasts when compared to other approaches. Long short-term memory (LSTM) networks have proven to be effective in time series forecasting tasks, including air pollution prediction. However, optimizing LSTM models for enhanced accuracy and efficiency remains an ongoing research area. In this paper, we propose a novel approach that integrates the novel binary chimp optimization algorithm (BChOA) with LSTM networks to optimize air pollution prediction models. The proposed BChOA, inspired by the social behavior of chimpanzees, provides a powerful optimization technique to fine-tune the LSTM architecture and optimize its parameters. The evaluation of the results is performed using cross-validation methods such as the coefficient of determination ( $R^2$ ), accuracy, the root mean square error (RMSE), and receiver operating characteristic (ROC) curve. Additionally, the performance of the BChOA-LSTM model is compared against eight DL architectures. Experimental evaluations using real-world air pollution data demonstrate the superior performance of the proposed BChOA-based LSTM model compared to traditional LSTM models and other optimization algorithms. The BChOA-LSTM model achieved the highest accuracy of 96.41% on the validation datasets, making it the most successful approach. The results show that the BChOA-LSTM architecture performs better than the other architectures in terms of the  $R^2$  convergence curve, RMSE, and accuracy.

**Keywords:** air pollution; particulate matter (PM<sub>2.5</sub>); long short-term memory; novel binary chimp optimization algorithm



**Citation:** Baniasadi, S.; Salehi, R.; Soltani, S.; Martín, D.; Pourmand, P.; Ghafourian, E. Optimizing Long Short-Term Memory Network for Air Pollution Prediction Using a Novel Binary Chimp Optimization Algorithm. *Electronics* **2023**, *12*, 3985. <https://doi.org/10.3390/electronics12183985>

Academic Editors: Mattias O’Nils and Irida Shallari

Received: 11 August 2023

Revised: 11 September 2023

Accepted: 18 September 2023

Published: 21 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Air pollution has emerged as one of the most pressing environmental and public health challenges of our time. As urbanization, industrialization, and transportation continue to grow, so does the release of harmful pollutants into the atmosphere [1]. The consequences of air pollution are far-reaching, impacting both the environment and human health. From causing respiratory illnesses and cardiovascular diseases to contributing to climate change, the detrimental effects of air pollution are undeniable [2]. In this context, the accurate

forecasting of air pollution has become an imperative endeavor that is vital for safeguarding our planet and promoting the well-being of its inhabitants [3].

The motivation to advance air pollution forecasting goes beyond reactive measures; it embodies the spirit of preventive action. Armed with reliable forecasts, communities can engage in targeted public awareness campaigns, encouraging people to adopt eco-friendly practices and reduce their carbon footprint. By raising awareness and fostering a sense of environmental responsibility, forecasting air pollution fosters a culture of sustainable living where individuals become active participants in preserving the quality of the air they breathe [2,3].

Furthermore, forecasting air pollution is a scientific pursuit driven by technological advancements and innovative methodologies [4]. Researchers, environmentalists, and data scientists are continually refining atmospheric models, integrating machine learning (ML) algorithms, and harnessing the power of satellite technology to enhance the accuracy of predictions. This multidisciplinary collaboration serves not only to improve forecasting capabilities, but also to deepen our understanding of the intricate interplay between human activities, weather patterns, and air quality [5].

In recent years, deep learning (DL) has emerged as a revolutionary approach in various fields, and now, it holds the promise of transforming air pollution forecasting into a powerful tool for environmental protection [6]. The importance, necessity, and motivation behind forecasting air pollution using DL lie in its potential to revolutionize our understanding of atmospheric dynamics, improve prediction accuracy, and empower decision makers with intelligent insights. At the core of the DL paradigm lies the ability to analyze vast and complex datasets, revealing hidden patterns and correlations that traditional forecasting methods may overlook. By leveraging neural networks and sophisticated algorithms, DL models can assimilate diverse sources of information, including meteorological data, satellite observations, and emission inventories, to create a comprehensive understanding of air quality dynamics. As a result, these models offer a more nuanced and precise representation of air pollution levels, enabling a proactive rather than reactive approach to environmental management [4–8].

Long short-term memory (LSTM) networks have demonstrated promising results in various time series forecasting tasks, including air pollution prediction [8–10]. However, optimizing LSTM models to enhance their performance and accuracy remains an ongoing research area. One of the key weaknesses of LSTM networks lies in the optimization of network weights and biases during training [9–13]. Like other DL models, LSTM networks rely on optimization algorithms to adjust their parameters (weights and biases) in order to minimize the difference between the predicted outputs and actual targets. The most commonly used optimization algorithm in DL is gradient-based optimization, such as stochastic gradient descent (SGD). Gradient-based optimization possesses some disadvantages [14–17].

LSTM networks are susceptible to vanishing and exploding gradient problems [7]. During back propagation, gradients are propagated through time, and in deep networks or for long sequences, these gradients can become extremely small (vanish) or extremely large (explode). This hampers the learning process as the network struggles to adjust its parameters effectively. Gradient-based optimization can lead to slow convergence, especially in cases where the LSTM network has a large number of parameters and is dealing with complex sequential patterns. Optimization algorithms can sometimes become trapped in local minima, where they find suboptimal solutions instead of the global optimum. This can lead to a subpar LSTM model that fails to generalize well on unseen data. LSTM networks, with their recurrent connections, require substantial memory and computational resources. Gradient-based optimization adds to this overhead, making it more challenging to train large LSTM models on limited hardware [7,10–13].

In recent years, nature-inspired optimization algorithms have gained popularity in solving complex optimization problems [18–25]. One such algorithm, the binary chimp optimization algorithm (BChOA), has shown potential in addressing optimization chal-

allenges in various fields [20]. The main objective of this paper is to optimize LSTM networks for air pollution prediction using the innovative BChOA. The BChOA draws inspiration from the social behavior of chimpanzees and their ability to solve complex problems through cooperation and collaboration. By mimicking the behavior of chimpanzee communities, the BChOA explores the search space more effectively, leading to improved optimization results.

This novel approach aims to improve the accuracy and efficiency of air pollution prediction models by fine-tuning the LSTM architecture and optimizing its parameters. The proposed methodology involves integrating the BChOA into the training process of LSTM networks. The algorithm optimizes the network's weight values and hyper-parameters to enhance its ability to capture and predict air pollution patterns accurately. This paper addresses the limitations of existing air pollution prediction models by leveraging the power of the BChOA. To evaluate the performance of the proposed approach, extensive experiments will be conducted using real-world air pollution data from various monitoring stations. The results will be compared with traditional LSTM and DL models and other optimization algorithms to assess the effectiveness of the BChOA-based optimization. The major contributions of this paper can be summarized as follows:

- This paper introduces a novel BChOA aimed at fine-tuning the optimization parameters of LSTM models to enable more precise and dependable air pollution predictions.
- In the proposed BChOA, a novel approach for updating the positions of the chimpanzees is introduced. In the proposed BChOA, the equation for position updating is formulated as Equation (10). To accomplish this, a new sigmoid function, serving as the transfer function, is employed.
- In this paper, the data regarding the concentration of PM<sub>2.5</sub> pollutants were obtained for the period between 2006 and 2016. Meteorological data for a period of 10 years were obtained. These data encompass various parameters such as the maximum temperature, minimum temperature, pressure, wind speed, wind direction, and air humidity. The data were collected on a daily basis. To prepare and refine the meteorological and air pollution data, this paper employs the Fourier series method and utilizes the Savitzky–Golay filter to eliminate noise from the data.
- The evaluation of the results is performed using cross-validation methods such as the coefficient of determination ( $R^2$ ), accuracy, the root mean square error (RMSE), and receiver operating characteristic (ROC) curve. Additionally, the performance of the BChOA-LSTM model is compared against eight DL architectures.
- The simulation results show that the proposed algorithm has a better performance compared with other algorithms. The BChOA optimizes the values of the weights and biases in the LSTM network, enabling the LSTM network to better capture and represent the underlying patterns and dependencies in the data.

## 2. Literature Review

Utilizing ML algorithms to predict PM<sub>2.5</sub> levels represents a novel approach within the realm of air pollution investigation. In a notable study, Harishkumar et al. [1] harnessed predictive ML models to anticipate the concentration of particulate matter in the atmosphere using data gathered from monitoring Taiwan's air quality spanning the years 2012 to 2017. These newly developed models were pitted against conventional counterparts, revealing superior predictive capabilities. Particularly, the Gradient Boosting regression model outshone its counterparts in terms of predictive accuracy and overall performance. Tian et al. [2] assessed the effectiveness of six different ML models in predicting PM<sub>2.5</sub> levels in the Pearl River Delta (PRD) region from August 2014 to December 2019. They employed a diverse set of data sources, including meteorological information, vegetation data, topographical details, and points of interest (POIs), to ensure accurate daily PM<sub>2.5</sub> concentration estimations. The findings indicate that, overall, the random forest (RF) model outperformed the other models in terms of prediction accuracy. On the other hand, the

generalized additive model (GAM) exhibited the least favorable performance, followed by the support vector machine (SVM) model.

Xayasouk et al. [3] created predictive frameworks for estimating fine PM levels through the utilization of LSTM and deep auto-encoder (DAE) techniques. The outcomes of these models were assessed using the RMSE metric. These models were then employed to analyze hourly air quality information from 25 monitoring stations situated in Seoul, South Korea. The data spanned from 1 January 2015 to 31 December 2018. The findings demonstrate that the suggested models proficiently anticipated the fine PM concentrations, with the LSTM model exhibiting a marginally superior performance. Naz et al. [4] conducted a comparative examination of distinct DL-driven one-step prediction models aimed at forecasting five different air pollutants, including LSTM, gated recurrent unit (GRU), and a statistical model. To empirically assess their performance, they employed a publicly accessible dataset obtained from an air quality monitoring station located in the central area of Belfast, Northern Ireland. The findings indicate that the DL models consistently outperformed the statistical models, demonstrating a minimal RMSE of 0.59. Moreover, the DL approach exhibited the highest R-squared score of 0.856.

Shu et al. [5] introduced a novel approach known as the discrete wavelet and convolution-based auto-encoder (DW-CAE) model. This innovative model combines the principles of DL and signal processing by leveraging the discrete wavelet transform to extract both high- and low-frequency characteristics from the target sequence. The DW-CAE was applied to predict outcomes in the Beijing PM<sub>2.5</sub> dataset and the Yining air pollution dataset. The R-squared values for each variable exceed 93%, indicating strong predictive capability across all six air pollutants in the comprehensive prediction. By observing the studies mentioned above, it is evident that the significance and effectiveness of DL algorithms in forecasting air pollution have been acknowledged. However, a foundational challenge lies in fine-tuning the hyper-parameters of these DL algorithms. Presently, meta-heuristic algorithms [26] are employed to fine-tune network parameters [27–29].

Ghandourah et al. [27] introduced an improved artificial neural network (ANN) to forecast the displacement in composite pipes subjected to impact from a drop weight with varying velocities. They incorporated the Jaya algorithm and an enhanced version known as E-Jaya to optimize the ANN's training and prediction capabilities. The outcomes demonstrate that the E-Jaya algorithm significantly outperformed the original algorithm in terms of training and prediction accuracy. Aghakhani et al. [10] introduced an innovative hybrid algorithm based on artificial bee colonies. This algorithm was employed to enhance the architecture of a deep convolutional neural network (DCNN) with the goal of improving the detection capabilities of backscatter communication systems. The outcomes of their study demonstrate a noteworthy enhancement in detecting backscattered signals compared to previous methodologies.

In the study by Baniasadi et al. [11], they introduced an original approach called neighborhood search-based particle swarm optimization (NSBPSO) to effectively fine-tune the parameters of a DCNN designed for intrusion detection within IoT systems. The NSBPSO-DCNN framework demonstrated superior performance in terms of accuracy, sensitivity, and specificity across both the training and testing datasets. Remarkably, the NSBPSO-DCNN model achieved accuracy rates of 99.41% and 98.86% on the testing and training datasets, respectively. Sadeghi et al. [12] designed an innovative DL framework aimed at improving the classification of X-ray images related to COVID-19. They introduced a new approach called the multi-habitat migration artificial bee colony (MHMABC) algorithm to effectively train the DCNN. The performance of the MHMABC-DCNN model outperformed the other models. Teaching DL models is a computationally complex task, and the rising trend involves utilizing meta-heuristic methods to refine their parameters. Adapting a balance between exploration and exploitation to address intricate optimization challenges poses a difficulty. In response to these hurdles, this paper presents an innovative BChOA approach for training an LSTM network.

### 3. Proposed BChOA

In this section, we first present the standard version of the ChOA algorithm. Following that, we will explain the concepts of the improved binary algorithm.

#### 3.1. Standard ChOA

The ChOA is a meta-heuristic optimization algorithm inspired by the behavior of chimpanzees in the search for food and resources. The ChOA was introduced by Khishe and Mosavi in 2020 [26]. It was proposed as a nature-inspired algorithm for solving optimization problems. The ChOA mimics the foraging behavior of chimps, incorporating their social interaction and learning mechanisms. The ChOA divides the hunting process into four main phases: driving, blocking, chasing, and attacking. The algorithm begins by generating a random population of chimps to initiate the optimization process. These chimps are then randomly classified into four distinct groups: barrier, attacker, driver, and chaser. Each group plays a specific role in simulating the hunting behavior of chimps.

- **Driver Chimps:** The driver chimps closely follow the prey without attempting to reach it directly. Their purpose is to track the movements of the prey and gather information about its location.
- **Barrier Chimps:** Barrier chimps position themselves strategically, typically in trees, to create obstacles that block the progress of the prey. They act as barriers to divert the prey from reaching certain areas.
- **Chaser Chimps:** Chaser chimps are quick and agile, moving swiftly to catch up with the prey. Their primary objective is to pursue the prey closely and increase the chances of capturing it.
- **Attacker Chimps:** Attacker chimps analyze the behavior of the prey and predict its potential escape routes. They strategically position themselves in a way that forces the prey back towards the chasers, increasing the likelihood of a successful capture.

By emulating these distinct roles within the chimp population, the ChOA aims to effectively search for the optimal solution to the given optimization problem. The behaviors exhibited by each group during the hunting phases contribute to the exploration and exploitation of the search space. Equations (1)–(5) represent the formulations for driving and chasing the prey.

$$\mathbf{d} = \left| \mathbf{c} \cdot \mathbf{X}_{prey}(t) - \mathbf{m} \cdot \mathbf{X}_{chimp}(t) \right| \quad (1)$$

$$\mathbf{X}_{chimp}(t+1) = \mathbf{X}_{prey}(t) - \mathbf{a} \cdot \mathbf{d} \quad (2)$$

$$\mathbf{a} = 2 \cdot \mathbf{f} \cdot \mathbf{r}_1 - \mathbf{f} \quad (3)$$

$$\mathbf{c} = 2 \cdot \mathbf{r}_2 \quad (4)$$

$$\mathbf{m} = \text{Chaotic\_value} \quad (5)$$

where  $\mathbf{X}_{chimp}(t)$  denotes the chimp's position vector;  $\mathbf{X}_{prey}(t)$  is the prey's position vector;  $\mathbf{a}$ ,  $\mathbf{c}$ , and  $\mathbf{m}$  are the coefficient vectors;  $t$  presents the current iteration;  $\mathbf{r}_1$  and  $\mathbf{r}_2$  are the random vectors  $\in [0, 1]$ ;  $\mathbf{f}$  is the dynamic vector  $\in [0, 2.5]$ ; and  $\mathbf{m}$  denotes a chaotic vector. Figure 1 provides a visualization of the position vector in three dimensions and demonstrates the impacts of Equations (1) and (2). The figure also illustrates the presence of multiple neighboring positions. From the figure, it is evident that a chimp positioned at  $(X, Y, Z)$  has the ability to modify its location relative to the prey's position. By considering its current location and adjusting the values of the vectors  $\mathbf{a}$  and  $\mathbf{c}$ , the chimp can explore different positions surrounding the most suitable agent. In Figure 1  $(X^*, Y^*, Z^*)$  is the new position.

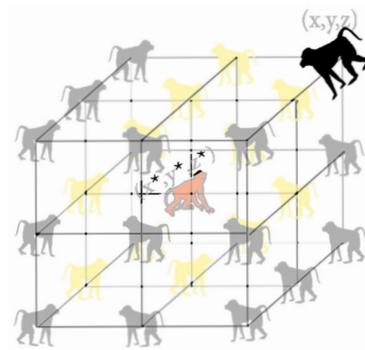


Figure 1. The position vectors and their possible next locations.

The initial step of locating the prey during the hunting phase involves the collaboration of the blocker, driver, and chaser chimps in identifying its position. Subsequently, the position of the prey is determined through calculations performed by the barrier, attacker, chaser, and driver chimps, with the remaining chimpanzees adjusting their positions based on the prey. Equations (6)–(8) encapsulate the formulations for these phases.

$$\begin{cases} d_{Attacher} = |c_1 \cdot X_{Attacher} - m_1 \cdot X| \\ d_{Barrier} = |c_2 \cdot X_{Barrier} - m_2 \cdot X| \\ d_{Chaser} = |c_3 \cdot X_{Chaser} - m_3 \cdot X| \\ d_{Driver} = |c_4 \cdot X_{Driver} - m_4 \cdot X| \end{cases} \quad (6)$$

$$\begin{cases} X_1 = X_{Attacher} - a_1(d_{Attacher}) \\ X_2 = X_{Barrier} - a_2(d_{Barrier}) \\ X_3 = X_{Chaser} - a_3(d_{Chaser}) \\ X_4 = X_{Driver} - a_4(d_{Driver}) \end{cases} \quad (7)$$

$$X(t + 1) = \frac{X_1 + X_2 + X_3 + X_4}{4} \quad (8)$$

where  $X_{Attacher}$  presents the best search agent,  $X_{Barrier}$  is the second-best search agent,  $X_{Chaser}$  denotes the third-best search agent,  $X_{Driver}$  is the fourth-best search agent, and  $X(t + 1)$  is the updated position of each chimp (Figure 2).

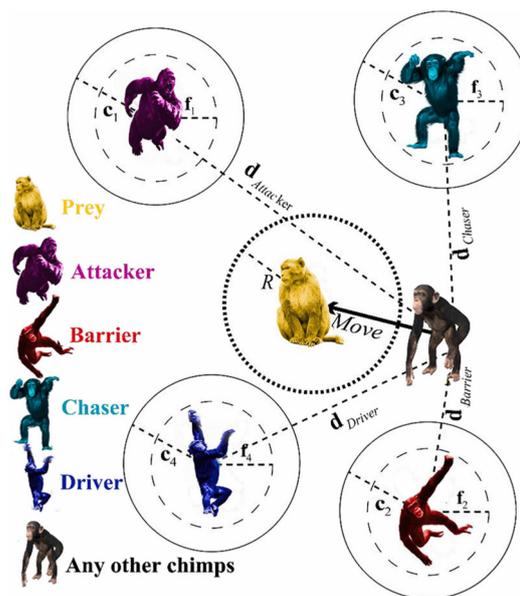


Figure 2. Position updating in ChOA.

Furthermore, in order to facilitate the exploration phase, the  $a$  parameter is introduced, where values greater than 1 or less than  $-1$  cause the chimps and preys to diverge. Conversely, parameter values between  $+1$  and  $-1$  aid in converging the chimps and preys, enhancing exploitation. Additionally, the parameter,  $c$ , plays a role in promoting the exploration process within the algorithm. Figure 3 demonstrates that the inequality condition compels the chimps to engage in attacking the prey.

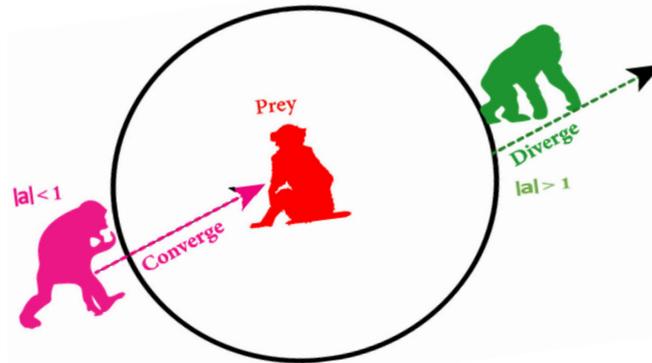


Figure 3. Position updating mechanism of chimps and effects of  $|a|$  on it.

Moreover, it is worth noting that all chimps engage in attacking the prey, and are driven by the desire for social rights (sexual incentives), regardless of their specific roles in the hunting process. To model this social behavior, chaotic maps are employed, as indicated in Equation (9).

$$X_{chimp}(t + 1) = \begin{cases} X_{prey}(t) - a \cdot d & \text{if } \mu < 0.5 \\ \text{Chaotic\_value} & \text{if } \mu \geq 0.5 \end{cases} \quad (9)$$

where  $\mu$  is the random number  $\in [0, 1]$ . In the continuous version of the ChOA, chimpanzees constantly change their positions at any point in space.

### 3.2. Novel BChOA

The motivation for developing a new BChOA stems from the need for efficient optimization techniques that can handle binary decision variables. Many real-world problems can be modeled using binary variables, such as binary-coded integer variables or binary flags representing the presence or absence of certain features or constraints. Traditional optimization algorithms are primarily designed for continuous variables, and their application to binary optimization problems often leads to suboptimal results or high computational complexity. The binary ChOA aims to address these limitations by specifically targeting binary optimization problems.

The effectiveness of binary meta-heuristic algorithms depends on the particular problem they are applied to. This means that not all binary meta-heuristic algorithms work equally well for every engineering optimization problem. Furthermore, the No Free Lunch (NFL) theorem suggests that there is no one-size-fits-all solution, implying that no single binary meta-heuristic algorithm can universally excel at all engineering optimization problems. Consequently, there are still many unresolved engineering optimization problems that could benefit from the development of novel binary meta-heuristic algorithms. As a result, researchers have maintained a strong interest in creating new binary algorithms specifically tailored to address discrete problems [30].

Binary encoding simplifies the representation of variables, particularly in optimization problems where variables can take on discrete values. By representing variables in a binary format, BChOA eliminates the need for continuous parameter tuning, making it easier to apply to various problem domains. BChOA’s binary encoding often leads to reduced computational complexity compared to ChOA, which relies on continuous variables. This reduction in complexity can result in a faster convergence and a lower computational over-

head, making the BChOA more convenient for solving optimization problems, especially those with large solution spaces. BChOAs are often more amenable to parallelization due to the discrete nature of binary variables. This means that the BChOA can take advantage of modern parallel computing architectures, further enhancing its convenience in solving complex and computationally intensive optimization problems [30–32].

Operators employed in meta-heuristic methods that deal with binary variables are limited to shifting 0 to 1 and 1 to 0, only allowing for movement towards closer or farther corners of the hypercube. Consequently, in the design of BChOA, the equation responsible for updating the positions needs to be adjusted. To achieve this, a transfer function becomes essential, as it maps the continuous space onto a discrete one. This transfer function plays a crucial role in determining the probability of switching the elements of the position vector from 0 to 1 or vice versa. The transfer function serves as a guide for exploring and exploiting the search space, controlling the movement of the search agents (symbolized as chimps) as they transition between different solutions. Its functionality relies on evaluating the fitness value of each solution and considering the current state of the algorithm to determine the likelihood of selecting a particular solution as the next candidate solution.

In numerous research endeavors, scientists have employed the creation of binary algorithms to tackle optimization challenges. Mirjalili and Hashim [30] presented a binary version of the magnetic optimization algorithm (MOA), utilizing both V-shaped and S-shaped transfer functions. To assess the BMOA’s performance, they evaluated it against PSO and a genetic algorithm (GA) on four benchmark functions. The results indicate that the BMOA was more accurate and faster in finding global minimums compared to the PSO and GA. These studies collectively highlight the advantages of using binary algorithms to address discrete problems. They also emphasize the use of transfer functions as a common approach in developing binary versions of meta-heuristic algorithms, as demonstrated in the studies mentioned. For reference, Table 1 provides a list of various binary algorithms and their associated transfer function types.

**Table 1.** A selection of binary algorithms and their transfer functions in the literature review.

Binary Algorithms	Transfer Functions	Equation
BMOA [30]	V-shaped	$T(x) =  \text{Tanh}(x) $
Binary Grey Wolf Optimizer [31]	S-shaped (sigmoid)	$T(x) = \frac{1}{1+e^{(-10x-0.5)}}$
Binary Gravitational Search Algorithms [32]	V-shaped	$T(x) =  \text{Tanh}(x) $

Based on the information provided in Table 1, it can be concluded that the transfer function holds paramount significance within binary algorithms. Consequently, in this paper, we employ both well-established transfer functions (S-shaped) and an innovative binary approach for the binary adaptation of the ChOA. In this section, a novel approach for updating the positions of the chimpanzees is introduced. In the proposed BChOA, the equation for position updating is formulated as Equation (10). To accomplish this, a sigmoid function, serving as the transfer function, is employed as depicted in Equation (11):

$$X_d^{t+1} = \begin{cases} 1 & \text{if } \text{sigmoid} \left( \frac{X_1+X_2+X_3+X_4}{4} \right) \geq R \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$\text{Sigmoid} (x) = \frac{1}{1 + e^{-\mu(x-\lambda)}} \tag{11}$$

where  $X_d^{t+1}$  is the updated binary position at the iteration  $t$ ;  $\text{Sigmoid} (x)$  represents the S-shaped functions;  $\mu$  is a threshold number  $\in (18, 19, 20, 21, 22)$ ;  $\lambda$  is a random number  $\in [0.45, 0.65]$ ;  $R$  is a random number  $\in [0, 1]$ ; and  $X_1, X_2, X_3,$  and  $X_4$  denote the chimpanzees’ movements towards the attacker, barrier, chaser, and driver chimps, respectively. The fundamental procedures of the BChOA are depicted in Figure 4. This illustration

highlights that the BChOA introduces distinct elements when compared to its continuous counterpart, the ChOA. These distinctions primarily involve the integration of a novel transfer function and a modified approach to updating positions.

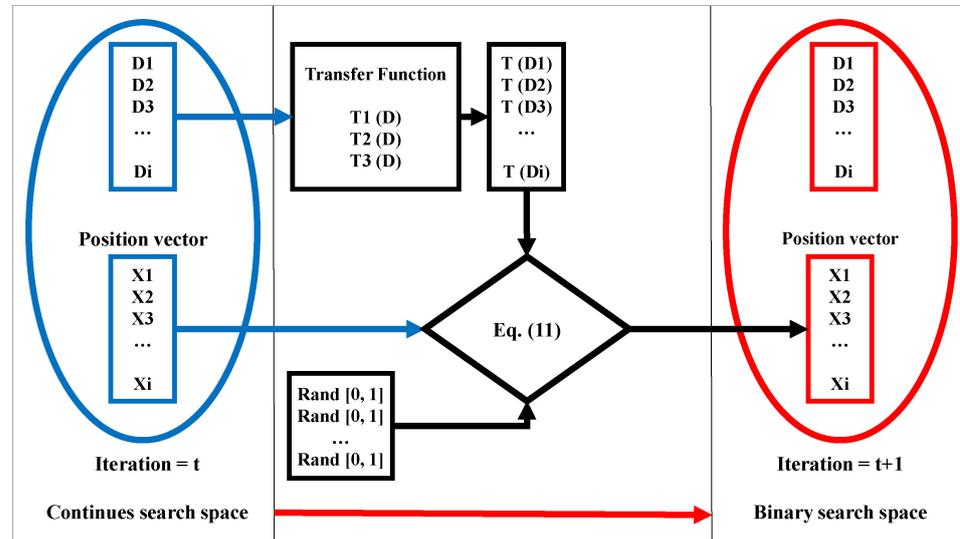


Figure 4. The structure of the proposed BChOA.

This transfer function serves as a critical bridge between the continuous search space and the binary search space. It facilitates the transformation of continuous variables into binary values, which is a fundamental requirement for addressing optimization problems involving binary decision variables. This transfer function plays a pivotal role in guiding the algorithm’s exploration and exploitation processes, effectively controlling the probability of transitioning between binary states (0 and 1). By introducing this function, the BChOA harnesses the benefits of binary encoding while maintaining the continuous representation for enhanced adaptability.

Unlike continuous optimization algorithms, where solutions evolve smoothly within a continuous domain, binary optimization algorithms must navigate the discrete landscape of binary variables. The position update method in the BChOA accounts for this binary nature, enabling the algorithm to efficiently explore the binary solution space. This modification contributes to reduced computational complexity and faster convergence rates, particularly when dealing with large solution spaces or problems with binary constraints. These innovations collectively empower the BChOA to excel in tackling optimization problems that involve binary decision variables, offering a promising alternative to traditional continuous optimization algorithms.

#### 4. Research Method

In this section, we will begin by introducing the proposed hybrid architecture. Following that, we will explore the study area and the data that are utilized for predicting air pollution.

##### 4.1. Evolutionary LSTM Network

LSTM networks have gained significant popularity and have been widely applied in various fields as one of the most promising DL techniques. An LSTM network is a type of recurrent neural network (RNN) architecture that is designed to effectively process and model sequences of data. It was introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997 and has since become one of the most popular and powerful models for sequential data analysis, such as natural language processing, speech recognition, and time series prediction [33,34].

LSTM networks are specifically designed to address the limitations of traditional RNNs, which struggle to capture long-term dependencies in sequences. Traditional RNNs suffer from the “vanishing gradient” problem, where the gradients diminish exponentially as they propagate back in time, making it difficult for the network to retain information from distant past steps. In contrast, LSTM networks are capable of learning and remembering information over longer time spans, making them more effective in modeling sequences with long-term dependencies [8,34].

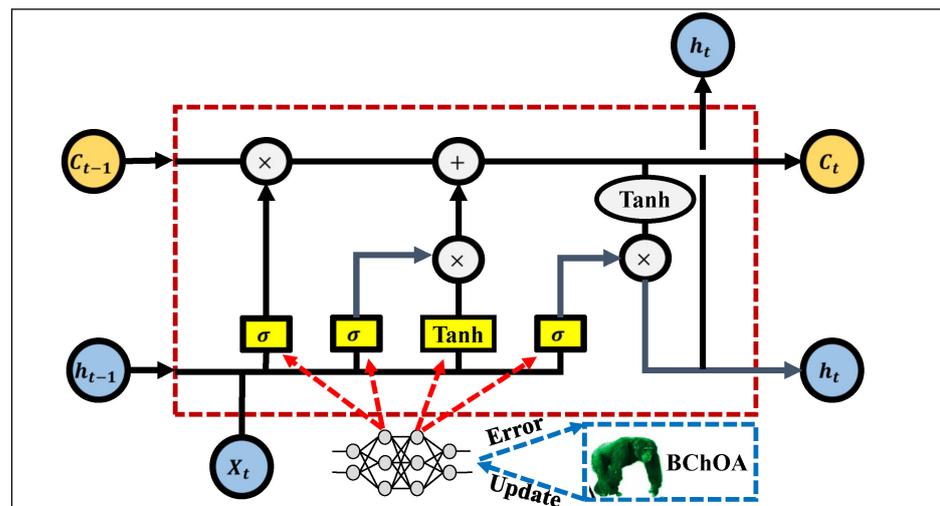
LSTM networks have been successfully applied in fields such as natural language processing (NLP), speech recognition, time series analysis, image captioning, machine translation, sentiment analysis, and many others. Their ability to handle sequential and temporal data makes them particularly effective in tasks that involve processing and understanding sequences. In NLP, LSTM networks have been used for tasks like text classification, named entity recognition, sentiment analysis, machine translation, language modeling, and text generation. They excel in capturing the contextual information and long-range dependencies that are present in natural language sequences. In the field of speech recognition, LSTM networks have been employed to model acoustic features and predict phonemes or words from audio signals. Their ability to capture long-term dependencies helps to improve the accuracy of speech recognition systems [7,33–35].

LSTM networks have also been applied in time series analysis, where they are used to forecast future values based on historical data. They have been used in financial forecasting, stock market prediction, energy load forecasting, and various other time-dependent prediction tasks. Overall, the success of LSTM networks can be attributed to their ability to capture long-term dependencies, handle sequential data, and model complex relationships. Their versatility and effectiveness in a wide range of tasks have made them a popular choice in the DL community [36].

The key component of an LSTM network is its memory cell, which is responsible for storing and updating information over time. The memory cell consists of three main parts: an input gate, a forget gate, and an output gate. These gates are learned through training and control the flow of information into, out of, and within the cell. When processing a sequence, the LSTM network takes the current element in the sequence, along with the previous hidden state and memory cell state, as an input [37]. The input gate determines how much of the current input should be stored in the memory cell. The forget gate decides what information to discard from the memory cell, allowing for the network to forget irrelevant information. The output gate determines how much of the memory cell’s content should be output as the current hidden state. By using these gates, LSTM networks can selectively store and retrieve information over multiple time steps, enabling them to capture long-term dependencies in sequences. This makes them particularly useful for tasks that involve understanding and generating sequences of data.

Overall, this paper’s main contribution is the application of the enhanced BChOA to train LSTM networks. By treating the weights and biases as optimization parameters and using the BChOA, this paper demonstrates a more efficient and effective training method for LSTM networks, potentially leading to enhanced performance in various tasks and problem domains. Traditionally, LSTM networks have been trained using the backpropagation (BP) algorithm, which adjusts the weights and biases of the network based on the gradient of the error with respect to these parameters. However, this paper proposes a shift from the conventional BP algorithm to the BChOA.

By utilizing the BChOA, this paper shows that the vector of weights and biases in the LSTM network can be efficiently updated. This updating process is performed in a way that satisfies the specific requirements of the problem being addressed. The BChOA optimizes the values of these parameters, enabling the LSTM network to better capture and represent the underlying patterns and dependencies in the data. Figure 5 shows the structure of the proposed BChOA-LSTM.



**Figure 5.** The structure of the proposed BChOA-LSTM.

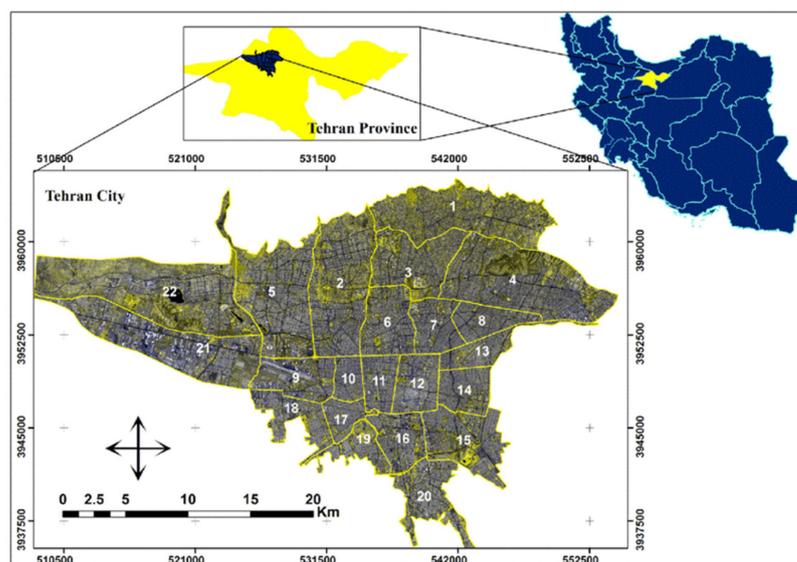
#### 4.2. Study Area and Dataset

As the capital of the country, Tehran holds great significance as the most important metropolis and serves as the political and commercial center of Iran. It is home to over 20% of the country's population. The climate of Tehran is influenced by its geographical location. While the northern areas of Tehran, which are close to the mountains, experience milder and more humid conditions, the weather in other parts of the city tends to be hot and dry, with slightly cold winters. The presence of the Alborz mountain range acts as a barrier, hindering the entry of many air masses into Tehran, similar to a dam. As a result, the city exhibits a relatively dry climate. However, being surrounded by mountains on three sides also creates a challenge when it comes to air pollution. The mountains trap pollutants within the city, preventing them from dissipating easily. Additionally, the excessive use of vehicles and the expansion of industries contribute significantly to the air pollution problem in Tehran.

Air pollution in the city of Tehran is primarily of anthropogenic origin, meaning it is caused by human activities. One of the major contributors to air pollution in the city is vehicular emissions. The high volume of vehicles on the roads significantly contributes to the city's air pollution levels. At times, the level of air pollution in Tehran exceeds the standard limits, leading to the complete shutdown of the city in severe cases. Given the severity of the issue, it becomes crucial to forecast and model air pollution in Tehran. By doing so, necessary measures can be implemented to control pollution effectively, and areas facing hazardous levels of pollution can be identified for targeted interventions. Furthermore, informing the public, particularly individuals with pre-existing health conditions, can help to prevent the spread of diseases associated with air pollution. To achieve these objectives, the entire city of Tehran has been selected as a study area for predicting air pollution levels and implementing appropriate measures. This comprehensive approach aims to tackle the complex issue of air pollution and its detrimental effects on public health and the environment. Figure 6 shows the geographical location of Tehran.

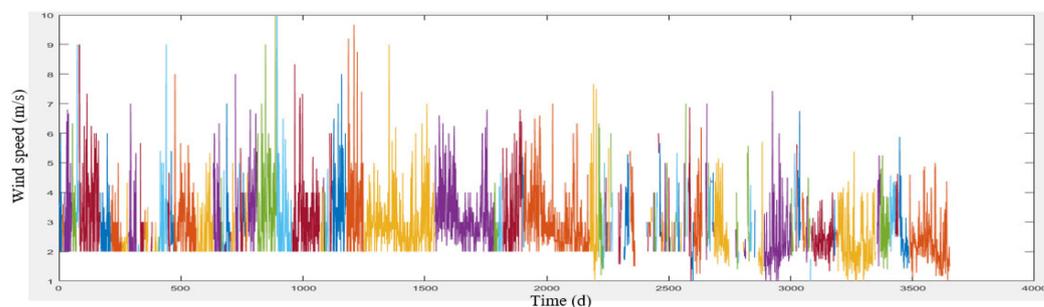
Air pollution is a phenomenon that is influenced by various factors. To make accurate forecasts, it is essential to correctly identify the parameters that affect air pollution. Generally, there are three main categories of parameters that have a significant impact on air pollution. These categories include pollutant concentration data, meteorological data, and spatial parameters. In this paper, data regarding the concentration of  $PM_{2.5}$  pollutants were obtained from the Tehran Municipality Air Quality Control Company for the period between 2006 and 2016. The concentration of these pollutants was measured and recorded on a daily basis by the company. Meteorological data for a period of 10 years were obtained from the Meteorological Research Center of Tehran Province. These data encompass various parameters such as the maximum temperature, minimum temperature,

pressure, wind speed, wind direction, and air humidity. The data were collected on a daily basis. This paper also employed geographic information system (GIS) techniques to incorporate spatial and descriptive parameters into our modeling process. This involved the integration of geographical features ( $X$ ,  $Y$ ,  $Z$ ) and characteristics of the study area, such as pollutant concentration data ( $PM_{2.5}$ ) and meteorological data. These spatial data layers were used to create additional input features for our predictive model. Furthermore, we employed a GIS analysis to visualize and interpolate maps.



**Figure 6.** The geographical location of the 22 districts of Tehran.

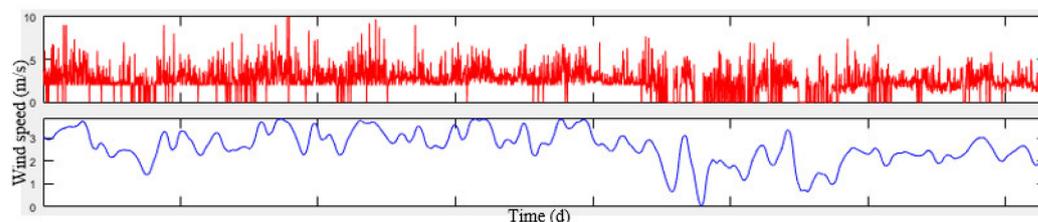
Overall, the preparation and refinement of meteorological and air pollution data are essential for developing reliable and effective air pollution forecasting models. By analyzing and refining meteorological data, researchers can identify correlations and patterns between specific weather parameters and air pollution levels. This enables the models to make more precise predictions of how air pollutants disperse and interact with the atmosphere. Proper data preparation and refinement allow for the validation and evaluation of air pollution forecasting models. Figure 7 illustrates the time series of the wind speed parameter at the Chitgar station as an example over the course of the ten-year period (with time measured in days).



**Figure 7.** The time series of the wind speed parameter.

In Figure 7, the change in the color of the lines indicates the presence of missing data. It is evident from the figure that neither the Fourier series nor the spline alone can adequately fit the data points due to errors and the lack of a suitable curve-fitting structure. Therefore, prior to fitting a curve, it is necessary to remove the noise, which, in this research, was achieved using the Savitzky–Golay filter. This filter was chosen because it is particularly effective in analyzing irregular or rapidly changing waves, as evident from the shape of the

sine and wavelet waves. Figure 8 displays the time series with the noise removed using the Savitzky–Golay filter. As can be observed, the filter provides a suitable fit for points with sudden changes in the data.



**Figure 8.** Refined wind speed image with Savitzky–Golay filter.

As depicted in Figure 8, the Savitzky–Golay filter successfully eliminated the irregular signals and sudden changes in the data. However, the issue of missing data in the signal still persists, as the Savitzky–Golay filter considers the signal as zero in the places where data are not available. To address this, the Fourier series and spline functions were employed in this research to compensate for the missing data. These mathematical techniques, such as the Fourier series and spline functions, aim to approximate the missing values based on the surrounding data points. By fitting curves or functions to the available data, they provide estimates for the missing portions of the signal, thereby mitigating the impact of the data gaps. By incorporating these techniques, researchers can enhance the continuity and completeness of the wind speed signal, enabling a more comprehensive analysis and interpretation of the data.

## 5. Simulation Results

This section assesses the effectiveness of the BChOA-LSTM method. To gauge its performance, five established and cutting-edge algorithms, namely ChOA, genetic algorithm (GA), ant colony optimization (ACO), black widow optimization (BWO), and improved crow search algorithm (I-CSA), were employed. Additionally, the performance of the BChOA-LSTM method was compared against three ML architectures: standard LSTM, RNN, and ANN. All of these algorithms were implemented in MATLAB.

Calibrating the parameters of meta-heuristic algorithms is essential but requires careful consideration. Hence, it is important to determine the best combination of parameters before assessing the algorithm’s performance. In this paper, a trial-and-error method was employed to adjust the parameter calibration. Each parameter was tested with different values while keeping the other variables unchanged. The fitness function was used as the main measure for evaluating and fine-tuning the algorithm’s parameters. Although a wide range of values were tested for each calibration parameter, only a selected set of instances is chosen and presented in Table 2.

**Table 2.** Parameter calibration achieved through the trial-and-error approach.

Algorithm	Parameter	Value
BChOA	a	[−1, 1]
	f	Linearly decreased from 2 to 0
	Threshold number ( $\mu$ )	19
	The probability range ( $\lambda$ )	0.52
	Number of scout bees (population size)	120
	Iteration	300

Table 2. Cont.

Algorithm	Parameter	Value
ChOA	a	$[-1, 1]$
	f	Linearly decreased from 2 to 0
	Number of scout bees (population size)	120
	Iteration	300
I-CSA	Flight length (fl)	2
	Awareness probability (AP)	0.1
	Population size	120
	Iteration	300
BWO	Procreate rate (PP)	0.65
	Mutation rate (PM)	0.21
	Cannibalism rate (CR)	0.47
	Population size	120
ACO	Iteration	300
	$\beta$	2
	$\rho, \alpha$	0.11
	$q_0$	0.89
	Population size	120
GA	Iteration	300
	Elitism percent	12%
	Mutation rate	0.13
	Crossover rate	0.92
	Population size	120
	Iteration	300

Interpolation is necessary to determine the pollutant values at the user position. In this paper, after data preparation, the Kriging interpolation method was employed to model the air pollution of Tehran city, and the interpolation map is presented in Figure 9.

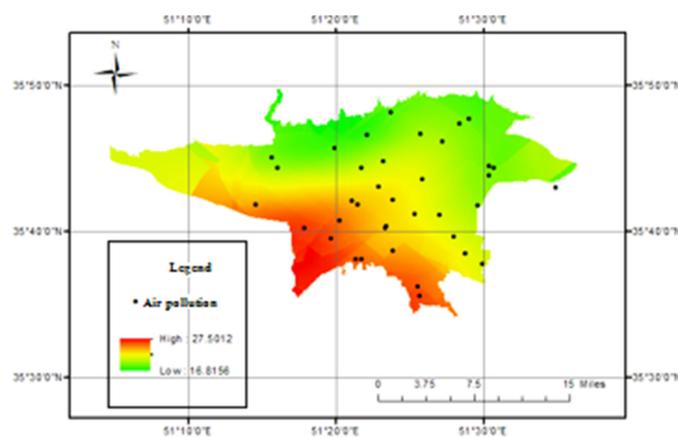


Figure 9. The PM<sub>2.5</sub> pollutant level map in Tehran using the Kriging interpolation method.

Kriging is a geostatistical interpolation method used to estimate values at unmeasured locations based on known data points. It is commonly used in various fields, including environmental science, geology, and spatial data analysis. The method takes into account

the spatial correlation or variability of the data to make predictions. In Kriging, the idea is to create a weighted average of the known data points, where the weights are determined based on the spatial distance and correlation between the points. Kriging is considered advantageous over some other interpolation methods because it incorporates both the spatial information and the spatial autocorrelation structure of the data.

In this paper, the evaluation of the results was performed using the cross-validation method and three parameters: the coefficient of determination ( $R^2$ ), accuracy, and the root mean square error (RMSE). The coefficient of determination measures the correlation between the observed values and the calculated values, ranging from 0 to 1. A value of one indicates a perfect correlation, while a value of zero indicates no correlation between the observed and calculated values. Equations (12)–(14) can be used to calculate the  $R^2$ , RMSE, and accuracy.

$$R^2 = \left[ \frac{1}{N} \frac{\sum_{i=1}^N [(P_i - \bar{P})(O_i - \bar{O})]}{\sigma_p \sigma_o} \right]^2 \quad (12)$$

$$RMSE = \left( \frac{1}{N} \sum_{i=1}^N [P_i - O_i]^2 \right)^{\frac{1}{2}} \quad (13)$$

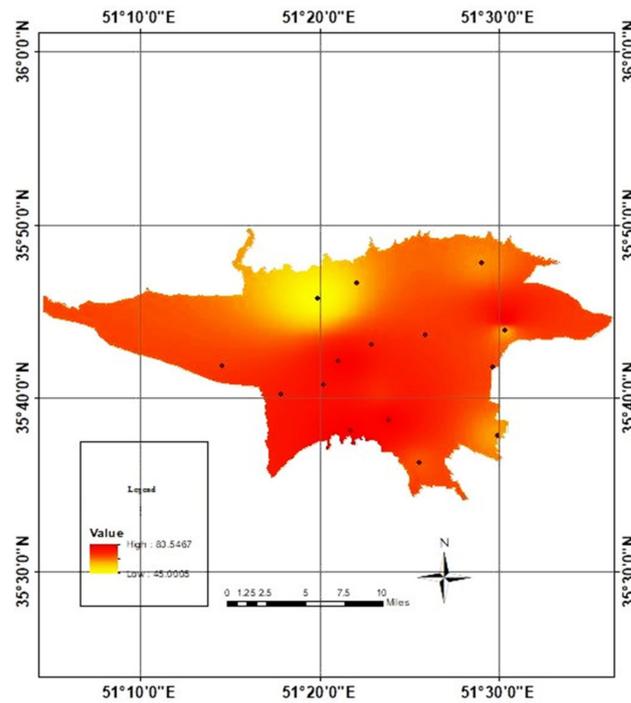
$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (14)$$

where  $N$  is the number of observations,  $O_i$  is the observed parameter,  $P_i$  is the calculated parameter,  $\bar{O}$  is the average observations parameter,  $\bar{P}$  is the average calculation parameter,  $\sigma_o$  is the standard deviation of observations,  $\sigma_p$  is the standard deviation of calculations,  $TP$  = true positive,  $FN$  = false negative,  $TN$  = true negative, and  $FP$  = false positive.

Table 3 presents the  $R^2$  values and accuracy results of different evolutionary architectures developed to forecast air pollution. The data in the table clearly show that the BChOA-LSTM architecture performs better than the other architectures in terms of both  $R^2$  and accuracy, not only in the training set, but also in the validation set. The BChOA-LSTM architecture achieved accuracies of 96.41% and 98.19% in the testing and training sets, respectively. By stating that the BChOA-LSTM architecture has the best  $R^2$  value, it means that this particular architecture demonstrates the highest level of accuracy in capturing and explaining the variances in the air pollution data. In other words, it provides the best fit to the actual data points and has the most reliable predictive power among all the architectures being evaluated. Figure 10 illustrates the generated map of the  $PM_{2.5}$  concentration using the proposed BChOA-LSTM architecture in December 2009. As depicted in the figure, it is evident that the northern region of Tehran has experienced relatively healthier air quality.

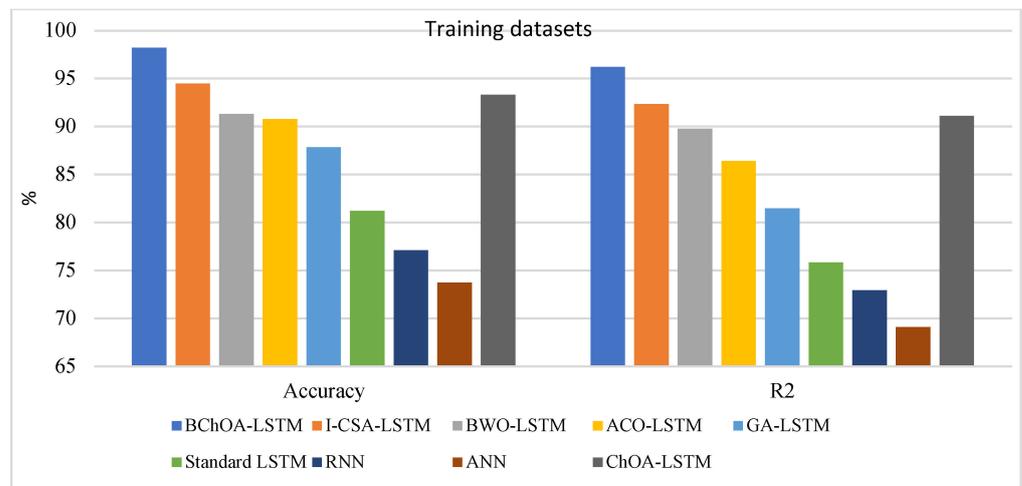
**Table 3.** The results of proposed architectures in the testing and training datasets.

Architectures	Training		Testing	
	$R^2$ (%)	Accuracy (%)	$R^2$ (%)	Accuracy (%)
BChOA-LSTM	96.19	98.19	94.31	96.41
ChOA-LSTM	91.08	93.29	90.06	90.98
I-CSA-LSTM	92.34	94.46	90.24	91.28
BWO-LSTM	89.76	91.29	86.95	87.76
ACO-LSTM	86.39	90.76	83.60	85.19
GA-LSTM	81.46	87.84	78.27	82.39
Standard LSTM	75.82	81.19	71.53	76.18
RNN	72.91	77.08	67.73	74.46
ANN	69.08	73.71	63.81	70.73

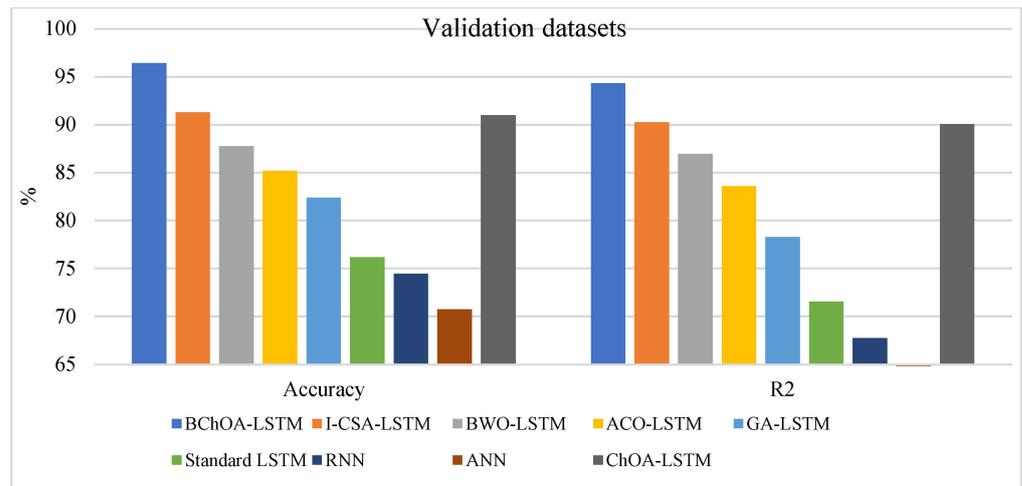


**Figure 10.** The generated map of PM<sub>2.5</sub> concentration using the proposed BChOA-LSTM model.

Figures 11 and 12 present a comparison of different architectures in the training and validation datasets. The architectures were ranked in order of performance, with the BChOA-LSTM architecture being the highest ranked, followed by I-CSA-LSTM, ChOA-LSTM, BWO-LSTM, ACO-LSTM, GA-LSTM, Standard LSTM, RNN, and ANN. These findings indicate that the suggested architectures were effectively trained using meta-heuristic algorithms. In other words, the algorithms used to train these architectures successfully optimized their performance. Additionally, the accuracy of these architectures remained consistent across different hybrid DL architectures in both the testing and training datasets. This suggests that the meta-heuristic algorithms employed during training yielded reliable and consistent accuracy across various models and datasets.



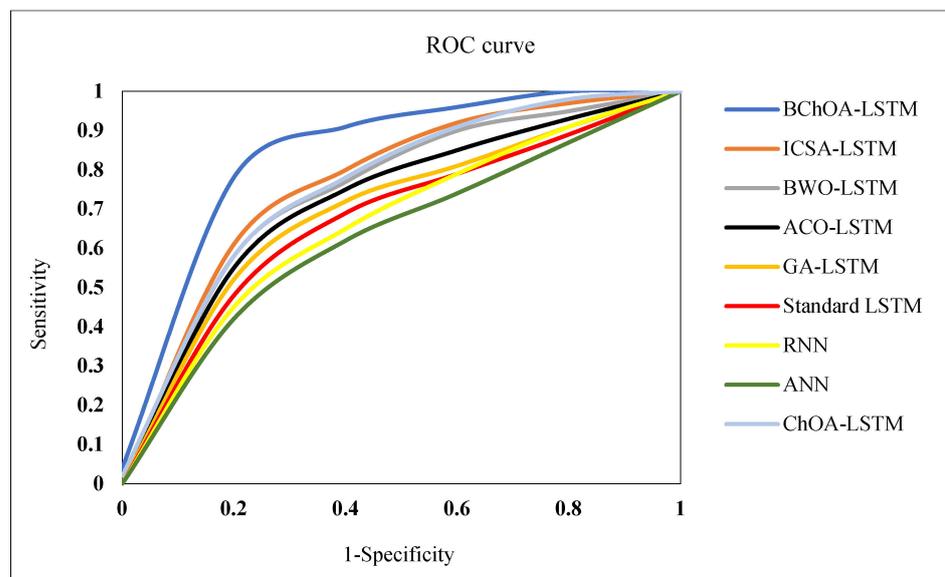
**Figure 11.** A graphical depiction illustrating the comparison of algorithms by utilizing training datasets.



**Figure 12.** A graphical depiction illustrating the comparison of algorithms by utilizing validation datasets.

In summary, the results highlight the strong performance and consistent accuracy of the suggested architectures, which were trained using meta-heuristic algorithms. This indicates that these architectures are capable of achieving high accuracy and maintaining consistency across a range of hybrid DL models.

Figure 13 presents a visual comparison of the ROC curve for various architectures. The ROC curve is a graphical representation that demonstrates the performance of a binary classifier, such as by distinguishing between two classes (e.g., positive and negative), as the discrimination threshold is adjusted. The ROC curve illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1—specificity) across different threshold values. Sensitivity represents the proportion of actual positive instances that were correctly identified, while specificity represents the proportion of actual negative instances that were correctly classified.



**Figure 13.** A visual comparison of the ROC curve for various architectures.

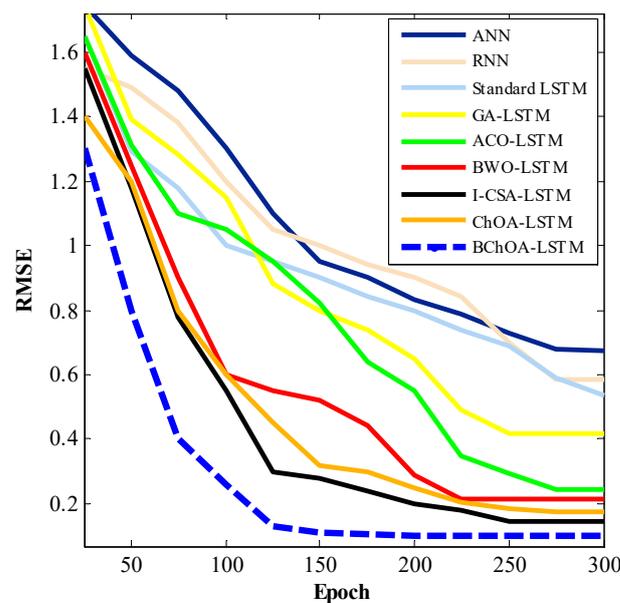
By examining the graph in Figure 13, it becomes apparent that the area under the curve (AUC) for the BChOA-LSTM, a specific architecture, outperforms the other architectures. The AUC is a metric that measures the overall performance of a classifier, indicating the probability that a randomly chosen positive instance will be ranked higher than a randomly

chosen negative instance. In this case, the higher AUC for the BChOA-LSTM architecture suggests that it achieves better classification accuracy and discrimination ability compared to the other architectures being evaluated.

The RMSE criteria are utilized to compare the proposed models listed in Table 4. It is evident that the BChOA-LSTM architecture outperforms the other architectures, signifying the effectiveness of this approach for the given problem. By incorporating the BChOA, the results demonstrate efficient updates to the LSTM network’s vector of weights and biases. The BChOA optimizes the parameter values, leading the LSTM network to better capture and represent the underlying patterns and dependencies within the data. As shown in Figure 14, the BChOA-LSTM architecture exhibits faster convergence compared to the other architectures. At epoch = 130, the BChOA-LSTM architecture achieves nearly the lowest RMSE value, while the other architectures still have higher RMSE values. Moreover, the BChOA-LSTM architecture demonstrates remarkable stability and rapid convergence as the epoch increases.

**Table 4.** The RMSE values of the different algorithms.

Algorithm	RMSE	
	Training Datasets	Validation Datasets
BChOA-LSTM	0.98	1.21
ChOA-LSTM	1.72	2.35
I-CSA-LSTM	1.47	2.19
BWO-LSTM	2.13	3.32
ACO-LSTM	2.43	4.62
GA-LSTM	4.19	6.29
Standard LSTM	5.34	8.85
RNN	5.87	9.18
ANN	6.74	11.52



**Figure 14.** The convergence trend of the algorithms.

In order to make a precise comparison of the algorithms with respect to their stability, we calculated the normalized variance of the optimal cost function across multiple runs for each method, and these values are presented in Table 5. For typical data, the variance

typically falls within the range of 0 to 1, with values closer to 0 indicating a more consistent and stable procedure. As depicted in Table 5, the BChOA-LSTM and ChOA-LSTM algorithms exhibit lower variances compared to the other algorithms, signifying their superior stability in performance.

**Table 5.** The normalized variance of different algorithms.

Algorithm	Run		
	10	15	20
BChOA-LSTM	0.0125	0.0049	0.0007
ChOA-LSTM	0.1056	0.0452	0.0063
I-CSA-LSTM	0.5698	0.2451	0.0452
BWO-LSTM	0.7452	0.4475	0.1236
ACO-LSTM	0.9521	0.8452	0.4596
GA-LSTM	1.1275	0.9521	0.7541
Standard LSTM	2.2514	2.0253	1.5486
RNN	2.3214	2.1210	1.9632
ANN	3.9125	3.0965	2.8652

## 6. Conclusions

Air pollution is a pressing environmental issue that poses severe health risks and ecological challenges worldwide. The accurate prediction of air pollution levels is crucial for implementing effective pollution control measures and public health interventions. This paper presents a novel approach to optimize LSTM networks for air pollution prediction using a novel BChOA. In the proposed BChOA technique, an innovative method was introduced to update the positions of chimpanzees. To achieve this, a new sigmoid function was utilized as the transfer mechanism. The study collected PM<sub>2.5</sub> pollutant concentration data from 2006 to 2016, alongside 10 years of meteorological data, including parameters like maximum and minimum temperatures, pressure, wind speed, wind direction, and air humidity. By combining the strengths of LSTM networks and the BChOA, we aimed to address the challenges in air pollution prediction and provide more accurate and reliable forecasts.

The assessment of outcomes involved employing cross-validation techniques, including metrics like the  $R^2$ , accuracy, RMSE, and ROC curve. Additionally, the performance of the BChOA-LSTM model was contrasted with seven different DL architectures. Through an experimental analysis using real-world air pollution data, the proposed BChOA-LSTM model demonstrated superior performance when compared to the other algorithms. Among these, the BChOA-LSTM model attained the highest accuracy at 96.41% on the validation datasets, marking it as the most successful approach. The findings of this research have significant implications for environmental management, public health, and policy making in combating air pollution. Furthermore, some unresolved issues concerning the BChOA and DL are outlined, motivating further research in this area.

The future of research in the BChOA involves a thorough investigation into optimizing the specific parameters and thresholds embedded within the algorithm's equations. This endeavor could encompass a comprehensive analysis of how variations in these parameters impact the algorithm's convergence rate, solution quality, and computational efficiency. Researchers could explore techniques such as metaheuristic parameter tuning or adaptive parameter adjustment strategies to dynamically adapt the parameters during the optimization process. The application of the BChOA across a wide spectrum of domains represents a compelling avenue for future research. The evolution of DL models is expected to shift towards addressing the challenge of limited labeled data availability. This transition could manifest in an increased emphasis on semi-supervised and unsupervised learning approaches. Future research might explore how the BChOA can be integrated into

these learning paradigms to enhance the utilization of unlabeled data and improve the performance and generalization of DL models.

Future studies could focus specifically on predicting the influence of industrial emissions on air pollution, using emission data, industrial location data, and emission dispersion modeling techniques. Extending the research to predict other pollutants like PM<sub>10</sub>, Nitrogen Dioxide (NO<sub>2</sub>), Sulfur Dioxide (SO<sub>2</sub>), and Ozone (O<sub>3</sub>) would certainly represent a valuable direction for future research. However, it would require a new research initiative accompanied with its own unique set of challenges and considerations.

**Author Contributions:** Conceptualization, S.B., R.S., S.S., D.M., P.P. and E.G.; methodology, S.B., R.S., S.S. and D.M.; software, S.S. and P.P.; validation, S.B. and E.G.; investigation, R.S., S.S. and D.M.; writing—original draft preparation, S.B., P.P. and S.S.; writing—review and editing, S.B., R.S., E.G. and D.M.; supervision, D.M.; funding acquisition, D.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The datasets generated and/or analyzed during the current study are available from the corresponding authors upon reasonable request.

**Acknowledgments:** The work described in this paper has been developed within the project PRESE-CREL (PID2021-124502OB-C43). We would like to acknowledge the financial support of the Ministerio de Ciencia e Investigación (Spain), in relation to the Plan Estatal de Investigación Científica y Técnica y de Innovación 2017-2020.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Harishkumar, K.S.; Yogesh, K.M.; Gad, I. Forecasting air pollution particulate matter (PM<sub>2.5</sub>) using machine learning regression models. *Procedia Comput. Sci.* **2020**, *171*, 2057–2066.
2. Tian, H.; Zhao, Y.; Luo, M.; He, Q.; Han, Y.; Zeng, Z. Estimating PM<sub>2.5</sub> from multisource data: A comparison of different machine learning models in the Pearl River Delta of China. *Urban Clim.* **2021**, *35*, 100740. [[CrossRef](#)]
3. Xayasouk, T.; Lee, H.; Lee, G. Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models. *Sustainability* **2020**, *12*, 2570. [[CrossRef](#)]
4. Naz, F.; McCann, C.; Fahim, M.; Cao, T.V.; Hunter, R.; Viet, N.T.; Duong, T.Q. Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas. *IEEE Access* **2023**, *11*, 64016–64025. [[CrossRef](#)]
5. Shu, Y.; Ding, C.; Tao, L.; Hu, C.; Tie, Z. Air Pollution Prediction Based on Discrete Wavelets and Deep Learning. *Sustainability* **2023**, *15*, 7367. [[CrossRef](#)]
6. Son, R.; Stratoulis, D.; Kim, H.C.; Yoon, J.H. Estimation of surface PM<sub>2.5</sub> concentrations from atmospheric gas species retrieved from tropomi using deep learning: Impacts of fire on air pollution over Thailand. *Atmos. Pollut. Res.* **2023**, *14*, 101875. [[CrossRef](#)]
7. Kaveh, M.; Mesgari, M.S. Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review. *Neural Process. Lett.* **2022**, *55*, 4519–4622. [[CrossRef](#)]
8. Fard, S.S.; Kaveh, M.; Mosavi, M.R.; Ko, S.B. An efficient modeling attack for breaking the security of XOR-Arbiter PUFs by using the fully connected and long-short term memory. *Microprocess. Microsyst.* **2022**, *94*, 104667. [[CrossRef](#)]
9. Gabriel, M.; Auer, T. LSTM Deep Learning Models for Virtual Sensing of Indoor Air Pollutants: A Feasible Alternative to Physical Sensors. *Buildings* **2023**, *13*, 1684. [[CrossRef](#)]
10. Chang, M.; Ku, Y. LSTM model for predicting the daily number of asthma patients in Seoul, South Korea, using meteorological and air pollution data. *Environ. Sci. Pollut. Res.* **2023**, *30*, 37440–37448. [[CrossRef](#)]
11. Najafi, F.; Kaveh, M.; Martín, D.; Reza Mosavi, M. Deep PUF: A highly reliable DRAM PUF-based authentication for IoT networks using deep convolutional neural networks. *Sensors* **2021**, *21*, 2009. [[CrossRef](#)] [[PubMed](#)]
12. Aghakhani, S.; Larijani, A.; Sadeghi, F.; Martín, D.; Shahrakht, A.A. A Novel Hybrid Artificial Bee Colony-Based Deep Convolutional Neural Network to Improve the Detection Performance of Backscatter Communication Systems. *Electronics* **2023**, *12*, 2263. [[CrossRef](#)]
13. Baniasadi, S.; Rostami, O.; Martín, D.; Kaveh, M. A novel deep supervised learning-based approach for intrusion detection in IoT systems. *Sensors* **2022**, *22*, 4459. [[CrossRef](#)] [[PubMed](#)]
14. Sadeghi, F.; Rostami, O.; Yi, M.K.; Hwang, S.O. A Deep Learning Approach for Detecting COVID-19 Using the Chest X-Ray Images. *Comput. Mater. Contin.* **2023**, *75*, 751–768.
15. Lotfy, A.; Kaveh, M.; Martín, D.; Mosavi, M.R. An efficient design of Anderson PUF by utilization of the Xilinx primitives in the SLICEM. *IEEE Access* **2021**, *9*, 23025–23034. [[CrossRef](#)]

16. Aghapour, S.; Kaveh, M.; Mosavi, M.R.; Martín, D. An ultra-lightweight mutual authentication scheme for smart grid two-way communications. *IEEE Access* **2021**, *9*, 74562–74573. [[CrossRef](#)]
17. Miri, S.; Kaveh, M.; Shahhoseini, H.S.; Mosavi, M.R.; Aghapour, S. On the security of ‘an ultra-lightweight and secure scheme for communications of smart meters and neighborhood gateways by utilization of an ARM Cortex-M microcontroller’. *IET Inf. Secur.* **2023**, *17*, 544–551. [[CrossRef](#)]
18. Kaveh, M.; Mesgari, M.S. Hospital site selection using hybrid PSO algorithm-Case study: District 2 of Tehran. *Sci.-Res. Q. Geogr. Data* **2019**, *28*, 7–22.
19. Kaveh, M.; Mesgari, M.S.; Martín, D.; Kaveh, M. TDMBBO: A novel three-dimensional migration model of biogeography-based optimization (case study: Facility planning and benchmark problems). *J. Supercomput.* **2023**, *79*, 9715–9770. [[CrossRef](#)]
20. Wang, J.; Khishe, M.; Kaveh, M.; Mohammadi, H. Binary chimp optimization algorithm (BChOA): A new binary meta-heuristic for solving optimization problems. *Cogn. Comput.* **2021**, *13*, 1297–1316. [[CrossRef](#)]
21. Kaveh, M.; Kaveh, M.; Mesgari, M.S.; Paland, R.S. Multiple criteria decision-making for hospital location-allocation based on improved genetic algorithm. *Appl. Geomat.* **2020**, *12*, 291–306. [[CrossRef](#)]
22. Gharehchopogh, F.S. Quantum-inspired metaheuristic algorithms: Comprehensive survey and classification. *Artif. Intell. Rev.* **2023**, *56*, 5479–5543. [[CrossRef](#)]
23. Kaveh, M.; Mesgari, M.S.; Saeidian, B. Orchard Algorithm (OA): A new meta-heuristic algorithm for solving discrete and continuous optimization problems. *Math. Comput. Simul.* **2023**, *208*, 19–35. [[CrossRef](#)]
24. Azizi, M.; Talatahari, S.; Gandomi, A.H. Fire Hawk Optimizer: A novel metaheuristic algorithm. *Artif. Intell. Rev.* **2023**, *56*, 287–363. [[CrossRef](#)]
25. Kaveh, M.; Mesgari, M.S.; Khosravi, A. Solving the local positioning problem using a four-layer artificial neural network. *Eng. J. Geospat. Inf. Technol.* **2020**, *7*, 21–40.
26. Khishe, M.; Mosavi, M.R. Chimp optimization algorithm. *Expert Syst. Appl.* **2020**, *149*, 113338. [[CrossRef](#)]
27. Ghandourah, E.; Khatir, S.; Banoqitah, E.M.; Alhawsawi, A.M.; Benaissa, B.; Wahab, M.A. Enhanced ANN Predictive Model for Composite Pipes Subjected to Low-Velocity Impact Loads. *Buildings* **2023**, *13*, 973. [[CrossRef](#)]
28. Amoura, N.; Benaissa, B.; Al Ali, M.; Khatir, S. Deep Neural Network and YUKI Algorithm for Inner Damage Characterization Based on Elastic Boundary Displacement. In Proceedings of the International Conference of Steel and Composite for Engineering Structures, Hanoi, Vietnam, 25–27 November 2022; pp. 220–233.
29. Benaissa, B.; Hocine, N.A.; Khatir, S.; Riahi, M.K.; Mirjalili, S. YUKI Algorithm and POD-RBF for Elastostatic and dynamic crack identification. *J. Comput. Sci.* **2021**, *55*, 101451. [[CrossRef](#)]
30. Mirjalili, S.; Hashim, S.Z.M. BMOA: Binary magnetic optimization algorithm. *Int. J. Mach. Learn. Comput.* **2012**, *2*, 204. [[CrossRef](#)]
31. Emary, E.; Zawbaa, H.M.; Hassanien, A.E. Binary grey wolf optimization approaches for feature selection. *Neurocomputing* **2016**, *172*, 371–381. [[CrossRef](#)]
32. Guha, R.; Ghosh, M.; Chakrabarti, A.; Sarkar, R.; Mirjalili, S. Introducing clustering based population in binary gravitational search algorithm for feature selection. *Appl. Soft Comput.* **2020**, *93*, 106341. [[CrossRef](#)]
33. Xin, J.; Zhou, C.; Jiang, Y.; Tang, Q.; Yang, X.; Zhou, J. A signal recovery method for bridge monitoring system using TVFEMD and encoder-decoder aided LSTM. *Measurement* **2023**, *214*, 112797. [[CrossRef](#)]
34. Ranawat, N.S.; Prakash, J.; Miglani, A.; Kankar, P.K. Performance evaluation of LSTM and Bi-LSTM using non-convolutional features for blockage detection in centrifugal pump. *Eng. Appl. Artif. Intell.* **2023**, *122*, 106092. [[CrossRef](#)]
35. Kaveh, M.; Aghapour, S.; Martín, D.; Mosavi, M.R. A secure lightweight signcryption scheme for smart grid communications using reliable physically unclonable function. In Proceedings of the 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Madrid, Spain, 9–12 June 2020; pp. 1–6.
36. Challa, S.K.; Kumar, A.; Semwal, V.B.; Dua, N. An optimized-LSTM and RGB-D sensor-based human gait trajectory generator for bipedal robot walking. *IEEE Sens. J.* **2022**, *22*, 24352–24363. [[CrossRef](#)]
37. Xu, Y.; Hu, C.; Wu, Q.; Jian, S.; Li, Z.; Chen, Y.; Wang, S. Research on particle swarm optimization in LSTM neural networks for rainfall-runoff simulation. *J. Hydrol.* **2022**, *608*, 127553. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.