mathematics

MDPI

Article

Predicting PM10 Concentrations Using Evolutionary Deep Neural Network and Satellite-Derived Aerosol Optical Depth

Yasser Ebrahimian Ghajari [1], Mehrdad Kaveh [2] and Diego Martín [2,*]

[1] Faculty of Civil Engineering, Babol Noshirvani University of Technology, Babol 47148-71167, Iran; y.ebrahimian@nit.ac.ir
[2] ETSI de Telecomunicación, Universidad Politécnica de Madrid, Av. Complutense 30, 28040 Madrid, Spain; m.kaveh11@email.kntu.ac.ir
* Correspondence: diego.martin.de.andres@upm.es

Abstract: Predicting particulate matter with a diameter of 10 µm (PM10) is crucial due to its impact on human health and the environment. Today, aerosol optical depth (AOD) offers high resolution and wide coverage, making it a viable way to estimate PM concentrations. Recent years have also witnessed in-creasing promise in refining air quality predictions via deep neural network (DNN) models, out-performing other techniques. However, learning the weights and biases of the DNN is a task classified as an NP-hard problem. Current approaches such as gradient-based methods exhibit significant limitations, such as the risk of becoming ensnared in local minimal within multi-objective loss functions, substantial computational requirements, and the requirement for continuous objective functions. To tackle these challenges, this paper introduces a novel approach that combines the binary gray wolf optimizer (BGWO) with DNN to improve the optimization of models for air pollution prediction. The BGWO algorithm, inspired by the behavior of gray wolves, is used to optimize both the weight and bias of the DNN. In the proposed BGWO, a novel sigmoid function is proposed as a transfer function to adjust the position of the wolves. This study gathers meteorological data, topographic information, PM10 pollution data, and satellite images. Data preparation includes tasks such as noise removal and handling missing data. The proposed approach is evaluated through cross-validation using metrics such as correlation rate, R square, root-mean-square error (RMSE), and accuracy. The effectiveness of the BGWO-DNN framework is compared to seven other machine learning (ML) models. The experimental evaluation of the BGWO-DNN method using air pollution data shows its superior performance compared with traditional ML techniques. The BGWO-DNN, CapSA-DNN, and BBO-DNN models achieved the lowest RMSE values of 16.28, 19.26, and 20.74, respectively. Conversely, the SVM-Linear and GBM algorithms displayed the highest levels of error, yielding RMSE values of 36.82 and 32.50, respectively. The BGWO-DNN algorithm secured the highest R^2 (88.21%) and accuracy (93.17%) values, signifying its superior performance compared with other models. Additionally, the correlation between predicted and actual values shows that the proposed model surpasses the performance of other ML techniques. This paper also observes relatively stable pollution levels during spring and summer, contrasting with significant fluctuations during autumn and winter.

Keywords: PM10; air pollution; remote sensing; aerosol optical depth; deep neural network; novel binary gray wolf optimizer

MSC: 97P80; 68T07; 62H11

check for updates

Citation: Ghajari, Y.E.; Kaveh, M.; Martín, D. Predicting PM10 Concentrations Using Evolutionary Deep Neural Network and Satellite-Derived Aerosol Optical Depth. *Mathematics* **2023**, *11*, 4145. https://doi.org/10.3390/math11194145

Academic Editor: Mario Muñoz Organero

Received: 1 September 2023
Revised: 24 September 2023
Accepted: 27 September 2023
Published: 30 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, specifically the presence of particulate matter with a diameter of 10 µm or less (PM10), poses a significant threat to public health and the environment [1]. PM10 particles are small enough to be inhaled deep into the lungs, leading to a range of health

problems [2]. From exacerbating respiratory issues such as asthma and bronchitis to causing heart problems and even premature deaths, PM10 pollution directly impacts human health. Predicting PM10 pollution levels equips us with valuable insights into potential health risks, enabling timely interventions such as issuing health advisories, adjusting outdoor activities, and prescribing preventative measures to vulnerable populations [3].

PM10 pollution not only impacts human health but also the environment by settling on soil and water, disrupting ecosystems and natural processes. Predicting PM10 levels helps understand their distribution and sources, aiding in targeted strategies for emission reduction [4,5]. In urban areas, where PM10 is concentrated from vehicular and industrial emissions, predicting pollution is crucial for urban planning, guiding decisions such as school and hospital locations. Businesses can use forecasts to implement pollution controls, ensuring compliance and avoiding penalties [6]. Regulatory bodies can set emission limits based on predictive models, while the public gains real-time pollution information for informed outdoor activities. This awareness encourages cleaner air initiatives through advocacy and government action.

The integration of aerosol optical depth (AOD) data has emerged as a transformative approach to predict and mitigate PM10 concentrations [7–9]. Traditional ground-based monitoring stations have spatial limitations, whereas AOD obtained from satellite remote sensing (RS) offers real-time data across large areas, overcoming this drawback. AOD focuses on how particles scatter or absorb sunlight, providing insights into aerosol concentrations such as PM10 [10–12]. By combining AOD with ground measurements, a holistic understanding of air quality is achieved. Satellite-based RS captures AOD data accurately and frequently, offering dynamic and comprehensive air quality representations, crucial for prediction and management. The correlation between AOD and PM10 properties allows the development of predictive algorithms for more accurate forecasts. Overall, integrating AOD data marks a pivotal advancement in combating air pollution [13].

Based on the literature review, numerous studies have employed a combination of ML techniques and satellite imagery to forecast air pollution. Lee et al. [10] utilized a ML technique to gauge ground-level PM using data from the geostationary ocean color imager AOD. Their approach, employing random forest, demonstrated remarkable accuracy. They then employed these PM estimates within the Weather Research and Forecasting-Chemistry/three-dimensional variational data assimilation (DA) system to generate analysis fields. By initializing the model with these updated analyses, significant reduction in analysis errors and enhancement in forecast skill were observed. The predictions for PM10 exhibited substantial improvements for a 24 h forecast window, whereas PM2.5 predictions showed enhancement for up to 6 h.

Tuygun and Elbir [11] conducted a study in Turkey spanning 2008 to 2019, estimating intra-daily PM10 levels at 213 monitoring sites using satellite-derived AOD data from moderate resolution imaging spectroradiometer (MODIS). They employed the RF methodology, integrating AOD data from Terra satellite, meteorological information, MERRA-2 aerosol diagnostics, and supplementary variables to develop a prediction model. The RF model demonstrated moderately favorable performance countrywide, with a correlation coefficient of 0.72 and low root-mean-square error (RMSE), surpassing previous research. Particularly in coastal areas, individual monitoring sites showed improved PM10 predictions, achieving R^2 as high as 0.90. However, in sparsely populated regions with limited monitoring stations, the model displayed signs of overfitting.

Numerous investigations have centered on the estimation of air pollutants such as PM, encompassing PM2.5 and PM10, often employing AOD. However, due to the limited precision of AOD imagery, it remains impractical to utilize satellite-derived AOD data for estimating PM10 in smaller urban areas. In response to this challenge, Imani [12] took a novel approach by utilizing the level 1 product of MODIS. In this devised technique, a combination of a DNN and a RF model was employed, leveraging the first and second bands of MODIS data to predict PM values. The findings highlighted the remarkable

efficacy of the proposed models when compared with several cutting-edge methods for PM estimation [12].

You et al. [13] conducted an investigation in Xi'an to establish a connection between AOD and PM10. Initially, they observed weak correlation between satellite-derived AOD and PM10 data, especially noticeable in spring and summer. These seasons exhibited high AOD but low ground-level PM10, potentially due to factors such as agricultural burning and dust storms complicating the relationship. To address this, they refined their analysis by excluding data linked to biomass burning and elevated aerosol layers, resulting in a stronger AOD-PM10 correlation. They further improved estimations using a geographically weighted regression (GWR) model that incorporated AOD and meteorological parameters. Including meteorological factors significantly enhanced the model's performance. Cross-validation yielded an R^2 of 0.77 and an RMSE of 16.91 $\mu g/m^3$.

The spatial and temporal variability of PM10 concentrations demands sophisticated modeling techniques that can capture these nuances accurately. As such, traditional prediction models often struggle to provide the required level of precision. On the other hand, the integration of deep neural networks (DNNs) has emerged as a transformative tool for predicting PM10 concentrations, offering unprecedented advantages in accuracy, complexity handling, and predictive capabilities [12]. DNNs possess a remarkable ability to uncover intricate patterns and relationships within complex datasets. As PM10 concentrations are influenced by a multitude of variables including meteorological conditions, emission sources, and geographical features, DNNs can capture non-linear correlations that conventional models struggle to discern. This leads to higher prediction accuracy and a more comprehensive understanding of the factors influencing air pollution levels [14–16].

The rapid advancements in computational power and data availability have propelled the application of DNNs across various domains [17–19]. Leveraging these advancements, researchers and practitioners are motivated to harness the capabilities of DNNs to revolutionize the prediction of PM10 concentrations, pushing the boundaries of our understanding and predictive accuracy [20,21]. However, the optimization of DNNs to enhance their performance and accuracy remains an uncharted territory. A fundamental challenge in the optimization of DNNs lies in the fine-tuning of their weights and biases throughout the training process. Traditional models often converge precisely to the endpoints of the gradient path, a phenomenon that can result in convergence to local minima and a slowdown in the learning speed.

In response to this challenge, contemporary research is turning to meta-heuristic algorithms, transcending the limitations of traditional gradient-based optimization methods [19]. These algorithms hold the promise of optimizing complex problems, while also addressing the shortcomings of conventional approaches [22–24]. By introducing a balanced and objective optimization approach, meta-heuristic algorithms dynamically adapt the weights and structure of the network. This transformative synergy amplifies both the pace and quality of DNN training and prediction. For this purpose, this paper utilizes the innovative binary gray wolf optimizer (BGWO) to optimize the DNN architecture. Just as wolves collaborate to optimize their pack's efficiency, the BGWO algorithm facilitates the optimization of the DNN's parameters. This innovative approach not only enhances the predictive capabilities of DNN but also showcases the potential for drawing inspiration from nature to solve intricate scientific challenges.

Kaya et al. [20] introduced a novel hybrid meta-heuristic algorithm for optimizing deep neural network weights, applied to early sepsis diagnosis. This algorithm aims to achieve global optimization using both particle swarm optimization (PSO) and the human mental search (HMS) algorithm for local search. Their approach was compared with other algorithms and demonstrated better reliability, durability, and adaptability. They integrated this algorithm with a deep neural network to form HMS-PSO-DNN, focusing on predicting sepsis in a dataset of 640 patients aged 18 to 60. The HMS-PSO-DNN model showed superior performance with a mean squared error (MSE) of 0.22 in 30 independent runs, indicating enhanced accuracy and robustness.

Khan et al. [21] introduced a DNN model tailored for software effort estimation, incorporating nature-inspired meta-heuristic algorithms, GWO, and strawberry algorithm (SB), to enhance the optimization process. Their goal was to investigate the benefits of these algorithms in optimizing deep learning (DL) models for software effort estimation. They assessed the algorithms using nine benchmark functions of varying dimensions, revealing that GWO exhibited greater accuracy in estimation compared with SB. The proposed GWDNNSB model, utilizing meta-heuristic algorithms for initializing weights and selecting learning rates, outperformed existing DNN-based software effort estimation methods. Training DL models is complex, and the trend is to use meta-heuristic techniques to fine-tune their parameters, finding a balance between exploration and exploitation for optimal results. To tackle these challenges, this paper introduces a new approach called BGWO for training DNNs.

This paper introduces an innovative strategy by combining the newly developed BGWO technique with DNN to amplify the optimization of models for predicting air pollution. Drawing inspiration from the social behavior of gray wolves, the BGWO algorithm offers a robust approach for refining DNN parameters. By integrating AOD data into the DNN, a comprehensive model is created, capturing variations in PM10 levels both locally and regionally. In the initial stages, information encompassing wind direction, minimum and maximum temperatures, air humidity, air pressure, topography, air pollution measurements from ground stations, and satellite images are gathered. To harness these data for predictive algorithms, a meticulous process of data preparation and enhancement is executed. This phase involves tasks such as eliminating noise, addressing missing data, and normalizing the data. Furthermore, the effectiveness of the BGWO-DNN framework is benchmarked against seven distinct machine learning (ML) models. Through experimental evaluations using air pollution data, the proposed BGWO-DNN technique demonstrates its superiority in performance over various ML approaches.

1.1. Paper Contributions

The key contributions of this paper can be outlined as follows:

- This paper employs the fusion of AOD and meteorological data within an evolutionary DNN architecture to enhance the accuracy of air pollution forecasting.
- This paper presents an innovative BGWO algorithm designed to enhance the refinement of optimization parameters in DNN models, leading to more accurate and reliable air pollution predictions. In the introduced BGWO framework, a fresh method for adjusting the positions of the wolves is put forth. This involves the utilization of a novel sigmoid function as the transfer function.
- The efficacy of the BGWO-DNN is benchmarked against seven distinct ML models: capuchin search algorithm (CapSA), biogeography-based optimization (BBO), PSO, random forest (RF), support vector machine with radial basis function kernel (SVM-RBF), linear support vector machine (SVM-linear), and gradient boosting model (GBM).
- The practical evaluation of the proposed BGWO-DNN using air pollution data showcases its superior performance compared to conventional ML approaches. The BGWO algorithm enhances the optimization process of weights and biases within the DNN framework, leading to an enhanced capability of the DNN to precisely apprehend and illustrate the underlying patterns and correlations present in the data.

1.2. Paper Questions

In this paper, our aim is to address the following questions:

1. How can ML algorithms be utilized to design and implement an air pollution prediction system?
2. How does the combination of the BGWO with DNN contribute to improving the optimization of models for air pollution prediction, and how does it address the NP-hard problem associated with learning the weights and biases of the DNN?

3. Which ML method is considered the most effective for estimating PM10 concentration?
4. During which seasons of the year does air pollution tend to be more pronounced?

1.3. Paper Organization

The rest of this paper is organized as follows: Section 2 describes the materials and methods of the paper, including the dataset, data preprocessing, study area, the proposed BGWO algorithm, and proposed BGWO-DNN model. Section 3 presents performance evaluations of the proposed algorithm for predicting PM10 concentrations in comparison with seven competitive algorithms, and finally, we present our conclusions for this paper in Section 4.

2. Materials and Methods

In this section, we will start by presenting the suggested BGWO algorithm and optimized DNN architecture. Subsequently, we will investigate both the case study and the data employed for the anticipation of air pollution.

2.1. Novel Binary GWO

GWO is a nature-inspired optimization algorithm that draws inspiration from the social hierarchy and hunting behavior of grey wolves in the wild [25]. It was introduced as a meta-heuristic algorithm for solving optimization problems. The GWO algorithm is designed to mimic the social interactions and hunting dynamics of wolf packs to search for optimal solutions in complex search spaces. In the GWO algorithm, a population of wolves represents potential solutions to the optimization problem. The algorithm simulates the social hierarchy among wolves, including alpha, beta, delta, and omega wolves, which are considered leaders in the pack (Figure 1). These leaders guide the exploration of the search space by adjusting their positions based on their fitness values and the positions of other wolves.

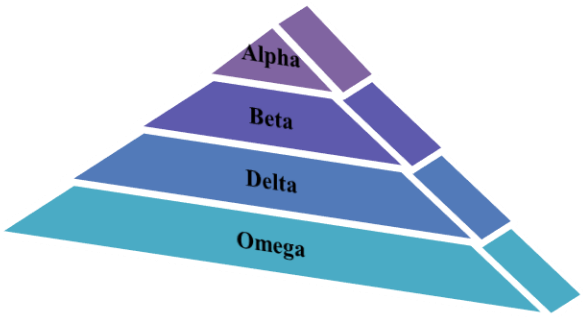


Figure 1. Gray wolf hierarchy.

In a wolf pack, the leaders are a male and a female known as alphas. These alphas hold the primary responsibility for decisions related to hunting, resting locations, wake-up times, and similar matters. Nevertheless, a form of democratic behavior has also been observed, wherein an alpha occasionally follows the lead of other group members. The designation of the alpha wolf as the dominant wolf stems from the necessity of group compliance with their directives. Consequently, the alpha wolves have the exclusive privilege of selecting a mate within the group. In the hierarchy of gray wolves, the beta position holds the second rank. Beta wolves are subordinate individuals that assist the alpha in decision-making and group endeavors. A beta wolf can be of either gender and often stands as the prime contender to succeed the alpha if the latter ages or passes away. The beta wolf is required to demonstrate deference to the alpha while also exerting authority over the lower-ranking wolves. Essentially, the beta serves as an advisor to the alpha and undertakes the role of organizing and coordinating group activities.

Among gray wolves, the omega wolf holds the lowest rank and assumes the role of the most submissive individual. The omega's primary role is obedience to the more

dominant wolves within the pack. They are typically the last to access food and must adhere to the instructions of higher-ranking members. While the omega's role might appear insignificant, it has been observed that the absence of an omega can lead to internal conflicts and issues within the entire group. This is because the omega's presence helps regulate the dominance structure and overall satisfaction of the pack by releasing tension and preventing other wolves from challenging the hierarchy. In addition to the alpha, beta, and omega positions, there exists another role referred to as the subordinate or delta wolf in some sources. Subordinate wolves report to alphas and betas, while also exerting dominance over omegas. This category encompasses various roles such as scouts, guards, elders, hunters, and caretakers. Scouts are responsible for patrolling the territory's borders and alerting the group to potential dangers. Guards ensure the group's safety and security. Elders are experienced wolves who have previously held alpha or beta positions. Predators assist alphas and betas during hunts and in procuring food for the group. Lastly, caretakers are tasked with tending to weak, sick, and injured members. The GWO algorithm involves three main steps in each iteration:

- **Initialization:** The positions of the alpha, beta, and delta wolves are initially set randomly within the search space, representing potential solutions.
- **Search Phase:** Other wolves in the pack adjust their positions based on the positions of the alpha, beta, and delta wolves. This adjustment is influenced by the social hierarchy and the concept of exploration and exploitation. The algorithm aims to strike a balance between exploring new areas of the search space and exploiting promising regions.
- **Update:** The alpha, beta, and delta wolves' positions are updated based on their fitness values and the positions of other wolves. This update helps refine the positions of the leaders, guiding the search towards better solutions.

Below, the mathematical formulation of the GWO algorithm is introduced. In the GWO, the process of hunting (optimization) is directed by α, β, and δ, while ω wolves trail behind these three collectives. As previously stated, gray wolves encircle their prey during hunts. To create a mathematical representation of this encirclement behavior, Equations (1)–(4) are presented:

$$\vec{D} = \left| \vec{C}.\vec{X}_p(t) - \vec{X}(t) \right| \tag{1}$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \bullet \vec{D} \tag{2}$$

$$\vec{A} = 2\vec{a} \bullet \vec{r}_1 - \vec{a} \tag{3}$$

$$\vec{C} = 2 \bullet \vec{r}_2 \tag{4}$$

where $\vec{A}$ and $\vec{C}$ are coefficient vectors, $\vec{X}_p$ is hunting position vector, $\vec{X}$ is the position vector of a gray wolf, the $\vec{a}$ vector is linearly reduced from 2 to 0 during the repetition, $\vec{r}_1$ and $\vec{r}_2$ are random vectors in the interval [0, 1].

To visualize the outcomes of the equations, we present Equations (1) and (2) in conjunction with a two-dimensional location vector and several potential neighborhoods, as depicted in Figure 2. As illustrated in the diagram, a gray wolf positioned at (X, Y) has the capability to adjust its coordinates based on the prey's location (X^*, Y^*). In Figure 2 (X^*, Y^*) is the new position.

To mathematically simulate the hunting conduct of gray wolves, we make the assumption that alpha (the most optimal solution), beta, and delta possess superior awareness regarding the potential whereabouts of the prey. As a result, we preserve the three finest solutions attained and compel the remaining search agents (including omega wolves) to adjust their positions in alignment with the positions of the top-performing search agents. Equations (5)–(7) are introduced to address this matter.

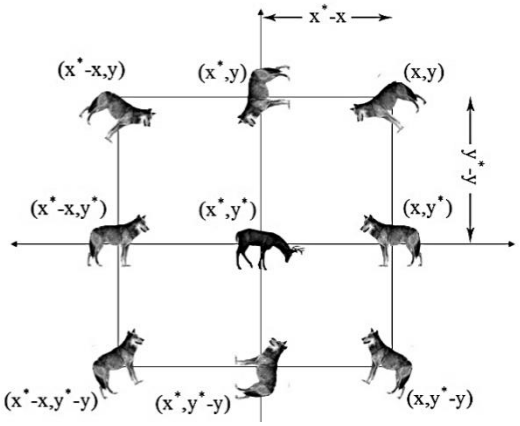


Figure 2. Two−dimensional location vectors and their next possible position.

$$\vec{D}_\alpha = \left|\vec{C}_1 \bullet \vec{X}_\alpha - \vec{X}\right|, \ \vec{D}_\beta = \left|\vec{C}_2 \bullet \vec{X}_\beta - \vec{X}\right|, \ \vec{D}_\delta = \left|\vec{C}_3 \bullet \vec{X}_\delta - \vec{X}\right| \tag{5}$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \bullet (\vec{D}_\alpha), \ \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \bullet (\vec{D}_\beta), \ \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \bullet (\vec{D}_\delta) \tag{6}$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \tag{7}$$

Figure 3 illustrates the procedure for adjusting the position of a search agent within a two-dimensional search space, based on the positions of alpha, beta, and delta. Evidently, the resulting position lies within a randomly determined location enclosed by a circle, the dimensions of which are determined by the positions of alpha, beta, and delta. In essence, the alpha, beta, and delta estimate the prey's location, prompting the remaining wolves to sporadically update their positions within the vicinity surrounding the prey.

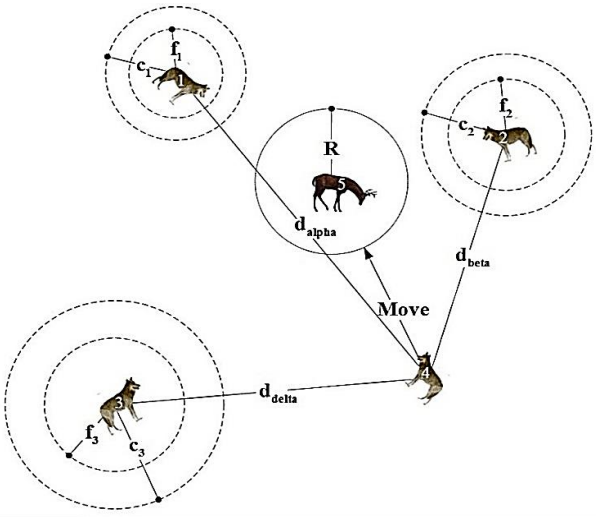


Figure 3. Updating the position in the GWO algorithm.

The motivation for developing a new variant of the binary GWO stems from the growing need for efficient and adaptable optimization algorithms in various fields of science, engineering, and industry. The original GWO algorithm, inspired by the social hunting behavior of gray wolves, has demonstrated its effectiveness in solving continuous optimization problems. However, many real-world problems involve discrete variables, which the original GWO is not directly designed to handle. Therefore, there is a clear

motivation to extend the capabilities of the GWO algorithm to address discrete optimization problems by creating a new binary version.

Furthermore, optimization algorithms play a pivotal role in tackling complex, high-dimensional problems that are often computationally expensive and time-consuming to solve. As such, researchers and practitioners are continually seeking enhancements to existing algorithms or novel approaches to improve optimization efficiency and effectiveness. By developing a new binary GWO, researchers aim to tap into the algorithm's potential for solving discrete optimization problems, offering a powerful tool for addressing a wider range of practical applications. This motivation arises from the recognition that a binary GWO could provide a valuable addition to the arsenal of optimization techniques, particularly in domains where binary decision variables are prevalent.

The transfer function in binary meta-heuristic algorithms plays a critical role in facilitating the transition from a continuous search space to a discrete one, where binary decision variables are used. In these algorithms, such as the BGWO, the transfer function acts as a bridge that enables the algorithm to make informed decisions about flipping the binary variables, i.e., switching between 0 and 1. This transition is essential because traditional optimization algorithms often deal with continuous variables, and adapting them to handle discrete variables requires a mechanism to guide the exploration and exploitation of the solution space effectively.

The transfer function serves as a decision-making mechanism that influences the probability of changing the values of binary variables in a solution. It takes into account various factors, such as the current state of the algorithm, the fitness value of solutions, and possibly randomization, to determine whether a specific binary variable should be flipped. The design of an effective transfer function is crucial, as it impacts how the algorithm explores the solution space, balances exploration and exploitation, and navigates through different candidate solutions.

In essence, the transfer function acts as a control mechanism that guides the algorithm's search process by modulating the transitions between binary values. It helps strike a balance between maintaining diversity in the search process (exploration) and converging towards promising solutions (exploitation). The goal is to optimize the binary decision variables in a way that leads to improved solutions while efficiently navigating the discrete solution space. The transfer function's formulation and tuning are essential aspects of designing a successful binary meta-heuristic algorithm, as they directly influence its search performance and convergence behavior.

In numerous research endeavors, scientists have employed binary algorithms to address optimization challenges [26–28]. In their work, Mirjalili and Hashim [26] introduced a binary variant of the magnetic optimization algorithm (MOA), incorporating both V-shaped and S-shaped (sigmoid) transfer functions. The findings demonstrated that BMOA outperformed PSO and genetic algorithm (GA) in terms of accuracy and speed when seeking global minimums. They also underscored the widespread use of sigmoid functions in developing binary versions of meta-heuristic algorithms. A comprehensive review of the literature reveals the significant and influential role that sigmoid functions play in shaping transfer functions. Consequently, in this paper, we leverage both established sigmoid functions (S-shaped) and a novel binary approach for adapting GWO to binary optimization. Here are some of the main and important advantages of using sigmoid transfer functions in this context:

- Smooth Transformation: Sigmoid functions offer a smooth and continuous transformation of input values. This smoothness aids in the convergence of optimization algorithms, as it enables gradual adjustments to the solutions being explored. This can help prevent abrupt and erratic changes in the search space, leading to more stable optimization processes.
- Non-Linearity: This non-linearity can help algorithms explore diverse regions of the search space and escape local optima. It enables the algorithm to adapt and

respond to different types of fitness landscapes, including those with complex and irregular shapes.

- Sigmoid Shaping: Sigmoid functions can shape the transfer functions effectively. They can map the real-valued outputs of the optimization algorithm to binary values (0 or 1) in a controlled and gradual manner. This ensures that the binary versions of metaheuristic algorithms maintain their effectiveness while working with discrete solutions.
- Compatibility: Sigmoid functions are compatible with various optimization algorithms, making them a valuable choice for adapting algorithms such as GWO to binary optimization.
- Transfer Function Design: Researchers have developed a deep understanding of how to design sigmoid transfer functions to suit different optimization scenarios. This knowledge base has led to the development of novel sigmoid functions specifically tailored for binary optimization, further enhancing their effectiveness.
- Empirical Success: The empirical success of using sigmoid functions in binary optimization is well-documented in the literature.

In this paper, a novel method for revising the wolves' positions is presented. The novel strategy put forth in the BGWO involves the formulation of the position update equation, which can be found in Equation (8). This is achieved by utilizing a sigmoid function as the transfer mechanism, represented by Equation (9):

$$X_d^{t+1} = \begin{cases} 1 & if\ sigmoid\left(\frac{X_1+X_2+X_3}{3}\right) \geq R \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$Sigmoid\ (x) = \frac{1}{1+\varphi e^{-\theta(x-\sigma)}} \tag{9}$$

where X_d^{t+1} is the updated binary position; $Sigmoid\ (x)$ is the novel transfer functions; θ is a threshold number $\in$ (17, 18, 19, 20, 21); σ is a random number $\in [0.39,\ 0.58]$; φ is a random number $\in [0.91,\ 1]$; R is a random number $\in [0,\ 1]$. Algorithm 1 shows the Pseudo code of the proposed BGWO.

Algorithm 1: The Pseudo code of the proposed BGWO.

Input:
n: Number of grey wolves,
N: Number of iterations.
Output:
X_α: Optimal grey wolf binary position,
$F\ (X_\alpha)$: Best fitness value.
Initialize a population of n wolves' positions at random $\in$ [0, 1].
Find the $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ solutions based on fitness.
While stopping criteria not met **do**
  for **wolf** $_i \in$ pack **do**
    Update **wolf** $_i$ position to a binary position according to Equation (8).
  end
  Update $\boldsymbol{\alpha}$, A, C.
  Evaluate the positions of individual wolves.
  Update $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$.
End

2.2. Optimized DNN

DNNs represent a remarkable breakthrough in the realm of artificial intelligence, standing as the pinnacle of ML achievements. The concept of DNNs originates from the broader field of artificial neural networks, which draws inspiration from the structure and function of biological neurons in the brain [29]. The history of DNNs dates back to the 1940s when the initial ideas were proposed. However, practical implementation

and training of deep networks were limited by computational resources and vanishing gradient problems. The term "DL" gained prominence in the 2000s, and breakthroughs in the mid-2010s, particularly with the advent of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), propelled DNNs into various applications such as image recognition, natural language processing, and more [30]. The immense power of DNNs becomes vividly apparent when delving into their myriad applications across diverse domains [31].

The distinguishing feature of DNNs is their depth, comprising multiple layers that allow them to autonomously learn hierarchical representations from raw input data. This depth fosters the extraction of increasingly abstract features, enabling DNNs to encapsulate intricate nuances present in various datasets. What sets DNNs apart is their ability to automatically learn features directly from raw data, bypassing the need for intricate manual feature engineering. The layers within DNNs act as hierarchical feature extractors, systematically refining representations of input data, thereby capturing complex relationships that were previously challenging to discern. This endows DNNs with a remarkable capacity for generalization, where they can apply their learned knowledge to previously unseen examples, translating into superior predictive performance [32].

DNNs have found significant utility in the domain of time series analysis, representing a powerful tool for extracting intricate temporal patterns and trends from sequential data. Time series data are characterized by their sequential nature, where each data point is associated with a specific time stamp. These types of data arise in various fields, including finance, economics, weather forecasting, signal processing, and more. DNNs offer a robust framework to effectively model and analyze time series data due to their ability to capture both short-term and long-term dependencies within the sequences. One of the primary strengths of DNNs in time series analysis lies in their capacity to learn complex temporal relationships that might not be apparent through traditional statistical methods. DNN architectures are designed to handle sequential data by maintaining an internal memory or hidden state that retains information about previous time steps. This makes them particularly adept at capturing patterns that unfold over time, such as trends, seasonality, and temporal dependencies [29–32].

However, applying DNNs to time series data comes with challenges. Proper preprocessing, normalization, and handling of missing data are essential to ensure optimal performance. Overfitting remains a concern, as time series data are often limited, and the network can memorize noise if not regulated effectively. Furthermore, selecting appropriate network architectures and hyper-parameters is crucial to achieve the desired balance between model complexity and generalization.

The primary contribution of the study lies in employing the improved BGWO technique for training DNN. This innovative approach treats the weights and biases as optimization variables and leverages the BGWO algorithm, resulting in a more streamlined and effective DNN training process. This has the potential to elevate performance across a spectrum of tasks and problem domains. Conventionally, DNNs have been trained utilizing the back-propagation (BP) algorithm, which fine-tunes network weights and biases based on error gradients. However, the research advocates a transition from the conventional BP method to the novel BGWO algorithm.

Through the application of the BGWO algorithm, the research effectively illustrates the optimized updating of the weight and bias vector within the DNN framework. Figure 4 visually presents the architecture of the novel BGWO-DNN paradigm, showcasing the integration of the BGWO algorithm into the DNN training process. This innovative approach marks a shift from traditional training methods and showcases how the BGWO algorithm can enhance the efficiency and efficacy of weight and bias adjustments within the DNN structure.

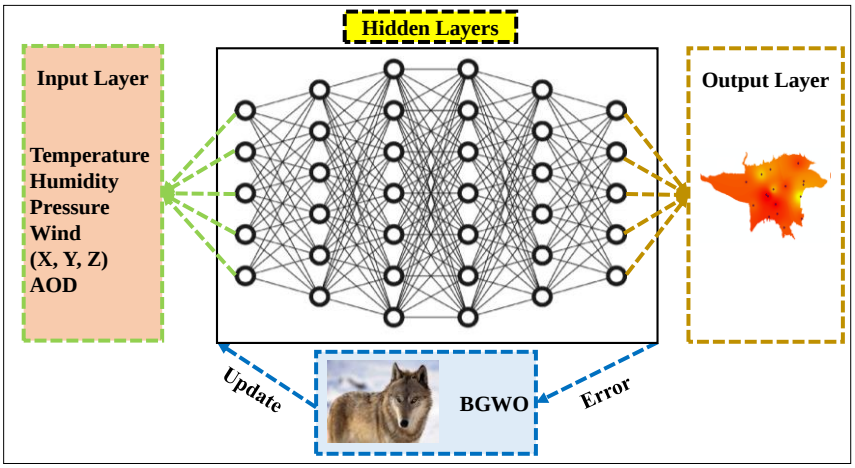


Figure 4. The structure of the proposed BGWO-DNN.

This paper employs BGWO for the training of a DNN. In the proposed methodology, BGWO is employed to optimize the parameters, including the weights and biases, of the DNN. In the context of BGWO modeling, a key goal is to create a solution inspired by the behavior of wolves. Figure 5 provides an illustration of the wolf structure within the BGWO framework.

Wolf definition	W_1	W_1	…	W_n	B_1	B_1	…	B_m

Figure 5. Wolf definition in proposed BGWO-DNN.

In Figure 6, a schematic and conceptual representation of BGWO operators is provided. As depicted in Figure 3, wolves ultimately update their preferred positions based on alpha, beta, and delta wolves, thereby enhancing the value of the objective function.

Alpha	0.21	0.32	0.62	0.71	…	0.24	0.53	0.91	0.26	RMSE = 23.52
Beta	0.09	0.63	0.29	0.81	…	0.37	0.63	0.76	0.14	RMSE = 26.19
Delta	0.82	0.33	0.27	0.49	…	0.92	0.74	0.51	0.19	RMSE = 32.17
New position	0.21	0.63	0.28	0.71	…	0.94	0.53	0.51	0.14	RMSE = 21.81

Figure 6. An example of updating the position in the BGWO algorithm.

2.3. Case Study

Tehran, the capital city of Iran, is renowned for its unique blend of cultural heritage and modern urban development. However, it faces significant challenges in terms of air pollution, largely attributed to its geographic location, topography, and meteorological conditions. Surrounded by mountains on three sides, Tehran experiences temperature inversions during the colder months, trapping pollutants close to the ground. This is exacerbated by high levels of vehicle emissions, industrial activity, and urban sprawl. The city's air quality has been a major concern, with particulate matter and smog frequently exceeding safe levels. Efforts to combat this issue include improving public transportation, promoting clean energy sources, and implementing stricter emission regulations.

The population of Tehran has grown exponentially over the years, making it one of the most populous cities in the Middle East. The city's rapid urbanization has led to increased demand for resources, energy, and infrastructure, putting additional stress on the environment. High population density contributes to traffic congestion and subsequently

worsens air quality. Furthermore, the city's population growth has led to the expansion of residential and commercial areas, often encroaching on green spaces and agricultural land. Sustainable urban planning and efficient resource management are crucial to address the challenges arising from the city's growing population [33].

Tehran's meteorological conditions and climate play a significant role in its air pollution dynamics. The city experiences a semi-arid climate with hot, dry summers and cold winters. The surrounding mountains influence the city's weather patterns, contributing to temperature inversions during colder months. These inversions trap pollutants in the lower atmosphere, leading to episodes of severe smog. Tehran's climate and topography underscore the need for tailored air quality management strategies that account for specific meteorological conditions. Balancing the demand for energy and comfort with the need to reduce pollution remains a complex challenge, particularly given the seasonal variations in weather patterns.

Human activities, including transportation, industry, and energy production, significantly contribute to Tehran's air pollution woes. The city's heavy reliance on private vehicles, coupled with insufficient public transportation options, results in high emissions of pollutants such as nitrogen oxides and volatile organic compounds. Industrial zones within and around the city release pollutants into the air, further deteriorating air quality. The use of outdated technologies and inadequate emission control measures in industries exacerbate the problem. While efforts are being made to transition to cleaner energy sources and improve emission regulations, a comprehensive approach to addressing human activities and their impact on air quality remains essential for the well-being of Tehran's residents and the environment. Figure 7 shows the geographical location of the study area [34].

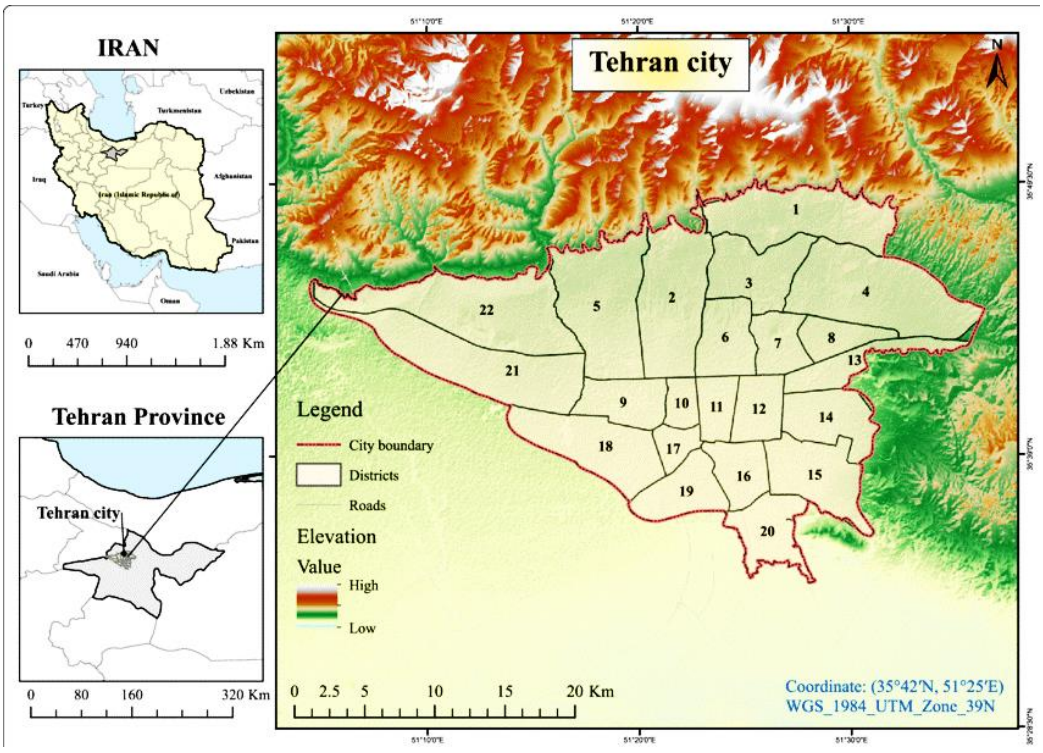


Figure 7. The study area.

Based on world health organization (WHO) PM10 standards, there are prescribed annual average limits. For instance, the WHO has set an annual average PM10 limit at approximately 20 ($\mu g/m^3$). Tehran, on the other hand, exhibits significantly elevated annual average PM10 levels, pointing to severe air quality issues in the city. Furthermore, throughout various seasons of the year, Tehran experiences PM10 levels that surpass the WHO-established thresholds.

Air pollution represents a phenomenon that encounters influence from a multitude of factors. Achieving precise forecasts demands a precise identification of these pivotal parameters that contribute to air pollution. Broadly, there are four categories of parameters that wield significant influence over air pollution. These encompass data related to pollutant concentration, RS, which incorporates AOD, meteorological conditions, as well as spatial characteristics. These influential parameters can essentially be categorized into two groups: spatial data and temporal data. Temporal data pertain to information that undergoes rapid fluctuations within brief moments and short time spans [35,36]. Refer to Figure 8 for an illustration depicting the segregation of data into the two aforementioned clusters: spatial data and temporal data.

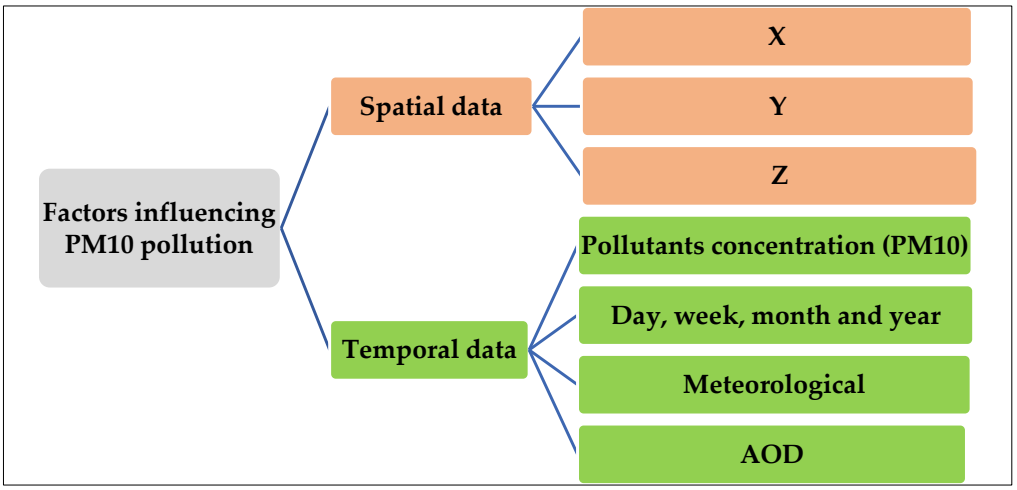


Figure 8. Factors influencing air pollution.

In this paper, information pertaining to the levels of PM10 particulate matter pollutants was acquired from the Air Quality Control Company of Tehran Municipality. The company consistently measures and documents the concentrations of these pollutants on a daily schedule. Over a span of 10 years, meteorological data were sourced from the Meteorological Research Center of Tehran Province. This dataset encompasses diverse factors including peak temperature, lowest temperature, atmospheric pressure, wind direction, and humidity. The data were collected on a daily cadence. This study also incorporates AOD data obtained from the MODIS sensor. The MODIS sensor, positioned on NASA's Terra platform, represents a prevalent RS technology that was deployed in 1999, entering Earth's orbit [37–39]. Subsequently, this equipment was integrated into the Aqua satellite's configuration in 2002 and dispatched into orbit. This instrument is endowed with the capability to capture data across 36 distinct spectral bands, spanning from wavelengths of 0.4 microns to 4.4 microns, while offering variable spatial resolution (2 bands at 250 m, 5 bands at 500 m, and 29 bands at 1 km).

In this paper, we primarily used MODIS data due to their high-resolution AOD measurements, which are essential for precise PM10 concentration estimation, especially in urban areas with intricate air quality patterns. Additionally, MODIS offers global coverage, making it versatile for models addressing spatial variability in PM10 on a larger scale. Its wide availability and established reliability for air quality applications further supported its selection as our data source.

To employ the data for the execution of prediction algorithms, the initial step involves preparing and refining the data. During this phase, tasks such as eliminating noise, identifying missing data (employing techniques such as Spline), and normalizing the data have been undertaken. Initially, the application of the Savitzky-Golay filter aids in the removal of erratic signals and abrupt fluctuations. However, the challenge of absent data persists, as the Savitzky-Golay filter assumes a signal value of zero in the absence of data points. To address this issue, the research employs Spline functions to effectively compensate for

the gaps in the data. Figure 9 displays the outcome of applying the Spline method to the time series of minimum temperatures. As depicted in the figure, the Spline method adeptly predicts the time intervals and patterns of the data for the instances of daily missing data, effectively memorizing and estimating the trends.

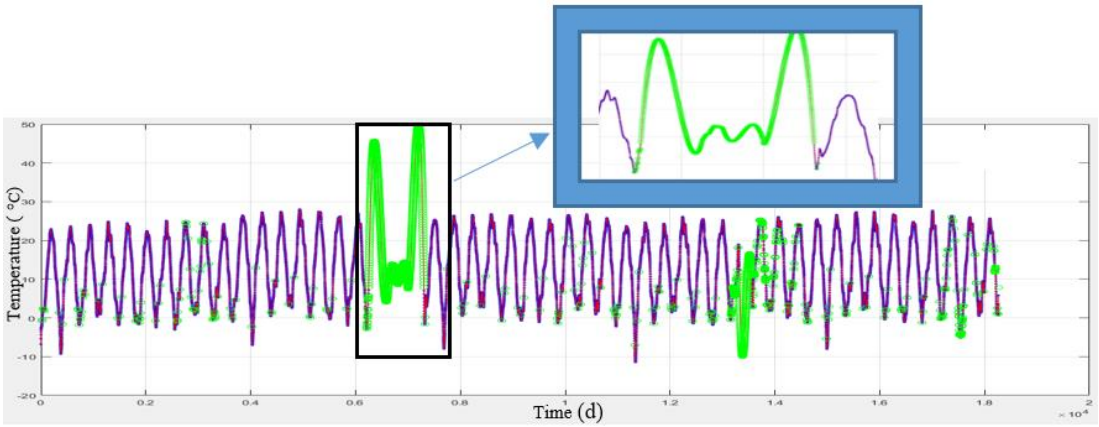


Figure 9. Outcome of applying the Spline method to the time series of minimum temperatures.

3. Results

This section analyzes the effectiveness of the BGWO-DNN approach that has been put forward. The evaluation is carried out using a dataset related to air pollution in Tehran city. A comparative study is conducted between the proposed algorithm and seven other algorithms, namely CapSA, BBO, PSO, RF, SVM_RBF, SVM-Linear, and GBM. The implementation of all these algorithms takes place within the R Studio software environment. For further insight, Table 1 provides details about the calibration parameters associated with the optimization algorithms.

Calibrating the parameters of meta-heuristic algorithms is a critical step in ensuring their optimal performance, but it is a process that requires careful and thoughtful attention. Essentially, this involves finding the right combination of parameter values that enable the algorithm to function effectively. However, before assessing how well the algorithm performs, it is crucial to identify these optimal parameter settings. In the context of this research paper, a method known as trial-and-error is utilized for parameter calibration. This means that we systematically vary each parameter individually, trying out different values for each while keeping all other factors constant. For instance, if an algorithm has multiple parameters such as learning rates, convergence thresholds, or population sizes, each of these parameters would be experimented with to understand their effects on the algorithm's behavior.

To measure the success of these parameter settings, a fitness function is employed. This function acts as a benchmark to evaluate the algorithm's performance for each combination of parameter values. While the range of possible values for each parameter is typically broad, due to practical limitations, we must select a subset of instances where they varied the parameters and present those results in a more manageable format. This condensed selection of instances is what's showcased in Table 1. This table serves as a snapshot of the trial-and-error process, showing which parameter values led to better or worse algorithm performance in those specific instances.

This paper employed a method called cross-validation to evaluate the results it obtained. This method involves partitioning the dataset into subsets to assess the model's performance multiple times, ensuring a comprehensive evaluation. The assessment was based on four key criteria: R Square (R^2), accuracy, correlation rate, and RMSE.

Table 1. Parameter setting of algorithms through the trial and error method.

Algorithm	Parameter	Value
BGWO	C	0.7
	A	0.3
	A	[0, 2]
	θ	20
	σ	0.49
	φ	0.98
	Population size	150
	Iteration	300
CapSA	Velocity control constants	1.00
	Inertia parameter	0.64
	Balance and elasticity factors	0.73, 9
	Population size	150
	Iteration	300
BBO	The probability range for migrating	[0, 1]
	Elitism percent	9%
	Mutation rate	0.13
	Population size	150
	Iteration	300
PSO	The inertial movement rate (α)	0.12
	Movement toward the best personal experience rate	0.66
	Movement toward the best global experience rate	0.92
	Population size	150
	Iteration	300
DNN	Number of hidden layers	{6, 7, 8}
	Number of neurons in hidden layers	{10, 25, 55}
	Learning rate	0.21
	Momentum	0.32
	Activation	Linear and Tanh
	Optimizer	SGD and BGWO
GBM	Number of estimators	100
	Learning rate	1.09
	Regularization parameters	0.07
	Maximum depth	11
RF	Number of estimators	100
	Maximum depth of trees	10
	Minimum samples per split	5
SVM	C (regularization parameter)	10
	Kernel type	Linear and RBF
	Gamma	0.002
	Iteration	300

Table 2 displays the outcomes of various evolutionary designs created for predicting air pollution. The information presented in the table distinctly indicates that the BGWO-DNN architecture outperforms the alternative designs in terms of R^2, accuracy, RMSE, and correlation rate. As shown in Table 2, the BGWO-DNN, CapSA-DNN, and BBO-DNN algorithms achieved the lowest RMSE values of 16.28, 19.26, and 20.74, respectively. These algorithms demonstrated a notably enhanced ability to accurately predict pollutant concentrations. Conversely, the SVM-Linear and GBM algorithms displayed the highest levels of error, yielding RMSE values of 36.82 and 32.50, respectively. The elevated R^2 values indicate superior algorithmic performance. Among these, the BGWO-DNN algorithm secured the highest R^2 value at 82.42%, signifying its superior performance compared to other models. Conversely, SVM-Linear and GBM algorithms exhibited poorer performance. Additionally, the correlation value and coefficient depict the relationship between estimated and measured values.

Table 2. The results of proposed models.

Models	**Criteria**			
	Correlation (%)	**Accuracy (%)**	**R^2 (%)**	**RMSE**
BGWO-DNN	92.25	93.17	88.21	16.28
CapSA-DNN	90.19	88.73	85.24	19.26
BBO-DNN	88.62	86.12	83.65	20.74
PSO-DNN	87.06	82.10	82.95	22.16
RF	84.49	80.21	74.09	24.25
SVM_RBF	84.28	79.19	71.41	25.30
SVM-Linear	65.51	72.85	42.27	36.82
GBM	72.06	73.46	52.19	32.50

In Figure 10, a visual representation of the discrepancies between the values predicted by the models and the actual values observed at the monitoring station is presented using error bar charts. These charts provide a clear depiction of how much variance exists between the estimated values generated by the models and the real measurements taken at the station. Additionally, the correlation between these predicted and actual values is showcased alongside the error bars. The key findings derived from this visual representation are significant. The study demonstrates that the proposed DNNs, SVM-RBF, and RF algorithms surpass the performance of other ML techniques. This is particularly evident in their ability to more accurately predict PM10 pollutant concentrations. The pronounced superiority of these algorithms is highlighted by their consistently lower discrepancies between predictions and actual measurements, as indicated by the error bar charts. This outcome accentuates the efficacy of proposed DNNs, SVM-RBF, and RF algorithms in enhancing the precision of PM10 concentration estimations. Such empirical evidence can significantly strengthen the argument for the adoption of these advanced algorithms in environmental monitoring and prediction applications.

Figure 11 display a visual comparison of various architectures. These architectures have been ordered based on their performance, with BGWO-DNN attaining the highest rank, trailed by CapSA-DNN, BBO-DNN, PSO-DNN, RF, SVM-RBF, GBM, and SVM-Linear. These results indicate the effective training of the proposed architectures through the utilization of meta-heuristic algorithms. In essence, the algorithms employed for training these architectures successfully optimized their operational efficiency. Moreover, the precision of these architectures remained consistent across diverse hybrid DL structures in both testing and training datasets. This consistency suggests that the meta-heuristic algorithms integrated into the training process produced dependable and uniform accuracy across various models and datasets.

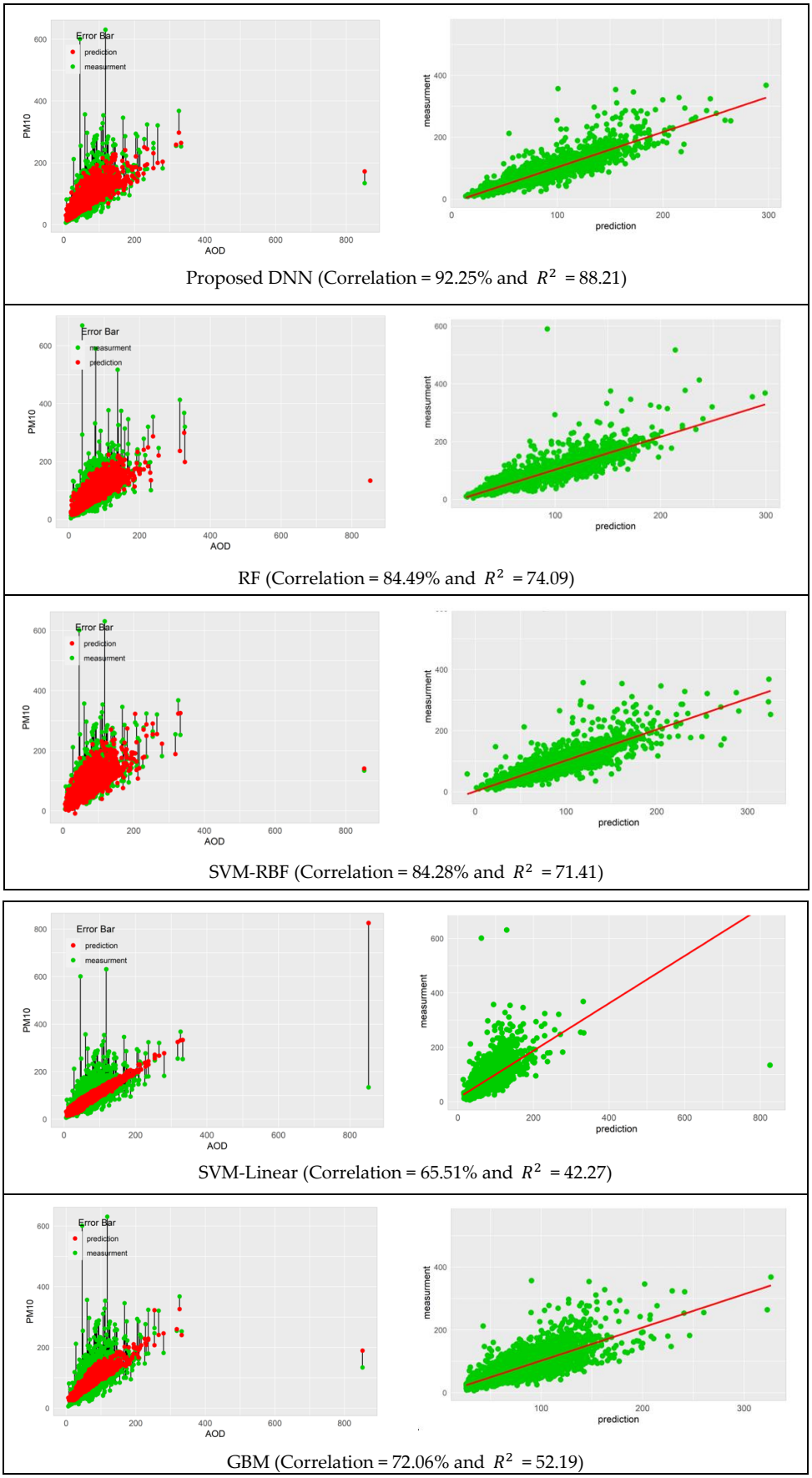


Figure 10. Error bar charts of algorithms.

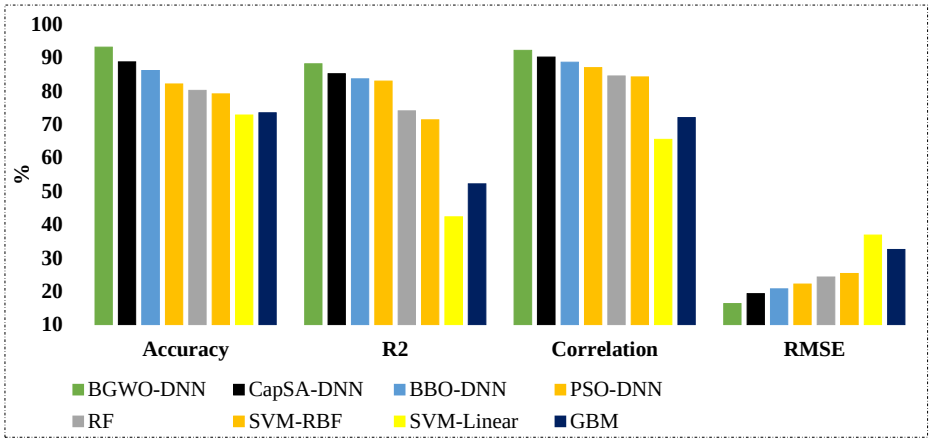


Figure 11. A visual comparison of various architectures.

Figure 12 illustrates the convergence progression of the algorithms based on the RMSE criterion. It is clear that the BGWO-DNN framework surpasses the alternative designs, highlighting the efficacy of this approach for the specified problem. The BGWO-DNN configuration outperforms competing designs, underscoring the effectiveness of this strategy for the given issue. Through the integration of the BGWO algorithm, the outcomes showcase proficient enhancements to the weight and bias vector of the DNN. The BGWO algorithm optimizes parameter values, guiding the DNN to more effectively capture and represent the underlying data patterns and relationships. As depicted in Figure 13, the BGWO-DNN architecture displays quicker convergence in comparison to the other architectures. At epoch = 110, the BGWO-DNN architecture achieves nearly the minimum RMSE, whereas the other architectures still exhibit higher RMSE values. Furthermore, the BGWO-DNN architecture demonstrates impressive consistency and rapid convergence as the epoch count increases.

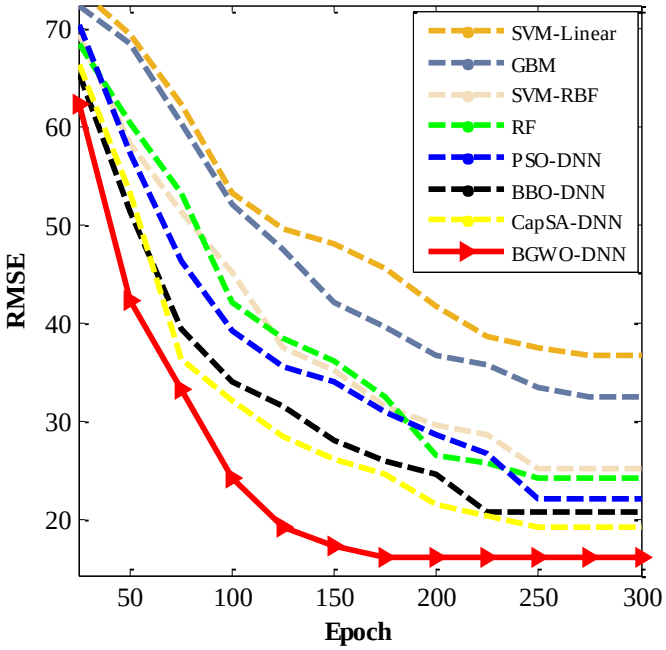
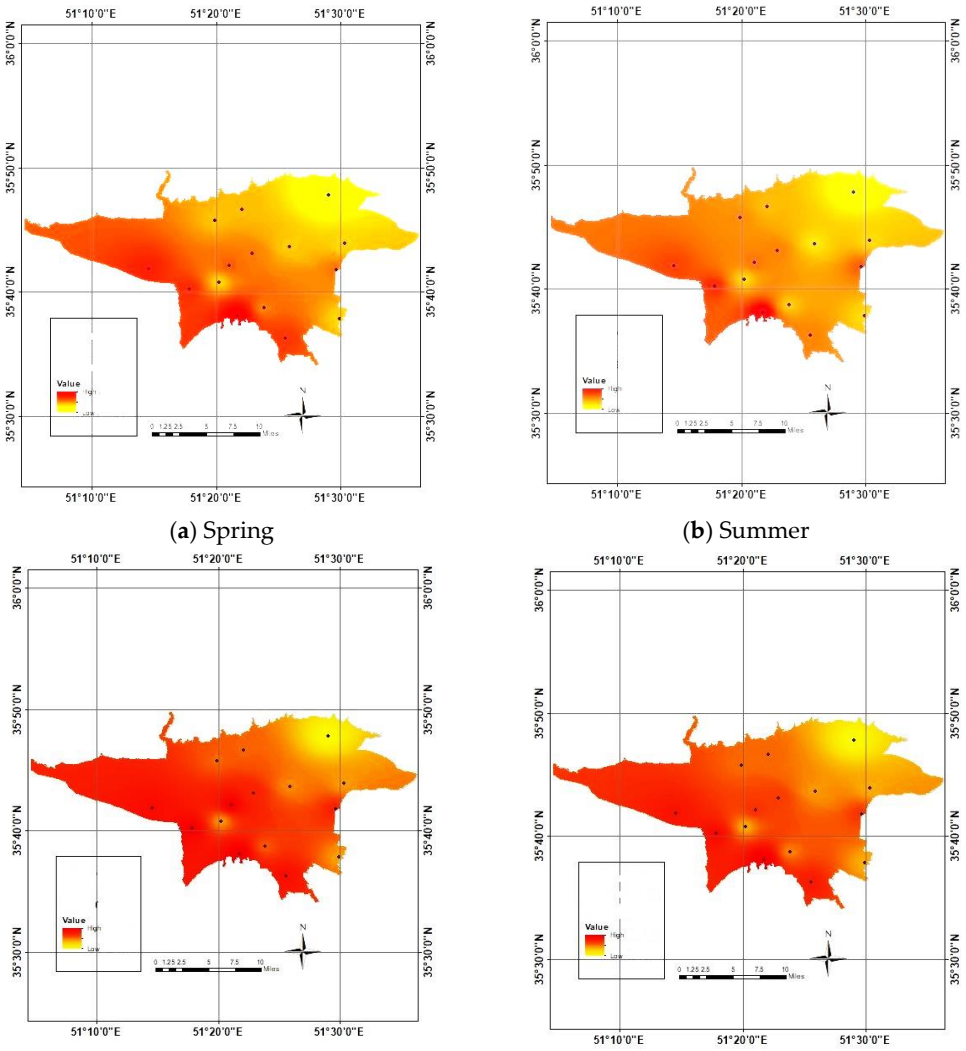


Figure 12. The convergence curve of the algorithms based on the RMSE criterion.



(**a**) Spring (**b**) Summer

(**c**) Autumn (**d**) Winter

Figure 13. The PM10 concentration maps using the proposed BGWO-DNN.

Figure 13a portrays the PM10 pollutant concentration during the spring of 2020, while Figure 13b illustrates the concentration of PM10 pollutant in the winter of 2020. Upon analyzing the outcomes from the autumn and winter of 2020, the statistical figures present a certain degree of resemblance to previous years, as the pollutant concentration attains relatively elevated levels. Nevertheless, a significant reduction became apparent in the spring of 2020, which can be attributed to the emergence of the Coronavirus and subsequent traffic restrictions implemented in Tehran. Regarding the distribution of pollution, a consistent pattern prevails across all months, with the northeastern part of Tehran identified as having the least pollution, whereas the southern and southwestern regions exhibit the highest concentrations of pollutants.

The results indicate that irrespective of the prevailing conditions in the region, pollutant concentration is at its lowest during the spring season. As we progress from spring and summer toward the colder autumn and winter seasons, the concentration gradually rises. Conversely, with the onset of spring again, the concentration decreases once more. Notably, during the spring of 2020, due to a significant reduction in urban traffic, the average PM10 concentration reached an unprecedented level in the past decade.

4. Conclusions

Accurately predicting PM10 concentrations is crucial for both human health and the environment. AOD data are promising due to their high resolution and wide spatio-temporal

coverage. This study presents a novel approach that combines the BGWO algorithm with the DNN model to enhance air pollution prediction models. Proposed BGWO, inspired by the behavior of gray wolves, optimizes DNN architecture and parameters. Data preparation involved noise reduction, handling missing data, and normalization. Evaluation included cross-validation using various metrics like correlation, R square, accuracy, and RMSE. The BGWO-DNN approach's effectiveness was compared against seven different ML models.

Empirical testing of the suggested BGWO-DNN technique, utilizing real-world air pollution data, showcased its superiority over conventional ML approaches in terms of performance. The study's insights unveiled intriguing patterns in air pollution concentration fluctuations across different seasons. While spring and summer demonstrated minimal concentration variations, autumn and winter exhibited significant fluctuations. This highlights the nuanced interplay between environmental factors and the intricate dynamics of air quality. In essence, the research underscores the potential of advanced computational techniques, such as the BGWO-DNN framework, to revolutionize our ability to predict and understand air pollution dynamics, thus contributing to more effective environmental management strategies.

In the following discussion, we delve into future works and challenges in this field. Through the amalgamation of satellite data and ground-based information, a wider expanse of pollutant coverage is achieved across both temporal and spatial dimensions. This amalgamated data serves as a valuable resource for authorities, researchers, and the general populace, enabling them to institute essential enhancements in air quality supervision and the mitigation of pollution. It is essential to recognize, however, that the precision of these forecasts hinges upon the fidelity of the satellite and ground data, as well as the efficacy of the employed models, demanding an ongoing drive for refinement.

In our future research, we recognize the significance of exploring alternative remote sensing data sources, including those from the Copernicus program and unmanned aerial vehicles (UAVs). These sources offer unique advantages and insights, while their integration presents intriguing challenges. Our objective is to comprehensively assess the strengths and weaknesses of each data source for air quality prediction. This will entail a thorough examination of integration challenges, encompassing data fusion, alignment, and consistency to ensure the robustness of our predictive models. We will conduct a detailed analysis of the pros and cons of each data source, accounting for factors such as spatial resolution, temporal coverage, data availability, and suitability across diverse geographical regions. Additionally, we will investigate how the choice of data source influences our methodology and model performance, providing valuable insights for both researchers and practitioners in the field of air quality prediction.

In future research, there is potential to narrow the focus on predicting how industrial emissions affect air quality. This can involve delving into industrial location data, refining emissions dispersion modeling techniques, and incorporating traffic data. Expanding the scope of the study to forecast the presence of additional pollutants such as PM2.5, Nitrogen Dioxide (NO_2), Sulfur Dioxide (SO_2), and Ozone (O_3) would undoubtedly constitute a valuable avenue for further investigation.

While DNNs offer remarkable capabilities, they are not without drawbacks. Their effectiveness heavily relies on extensive labeled data for training, which may not always be available. Additionally, training complex networks requires substantial computational resources like GPUs. Overfitting, where DNNs learn noise rather than patterns, requires careful regularization and hyper-parameter tuning. Moreover, the complexity of deep networks often leads to limited interpretability, making their decision-making process less transparent. Transfer learning allows models trained for one task to be adapted for related tasks with limited data, addressing data scarcity. Attention mechanisms help networks focus on relevant input data, proving valuable for sequence tasks such as language processing. The revolutionary "Transformer" architecture, seen in models such as BERT and GPT, has reshaped natural language understanding. These advancements demonstrate ongoing efforts to refine DNNs and unlock their potential.

Author Contributions: Conceptualization, Y.E.G. and M.K.; methodology, M.K. and D.M; software, Y.E.G. and M.K.; validation, Y.E.G. and D.M.; investigation, M.K.; data curation, Y.E.G. and M.K.; writing—original draft preparation, Y.E.G., D.M. and M.K.; supervision, Y.E.G. and 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 during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgments: The work described in this paper has been developed within the PRESECREL project (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. Shepelev, V.; Glushkov, A.; Slobodin, I.; Cherkassov, Y. Measuring and Modelling the Concentration of Vehicle-Related PM2. 5 and PM10 Emissions Based on Neural Networks. *Mathematics* **2023**, *11*, 1144. [CrossRef]
2. Park, D.H.; Kim, S.W.; Kim, M.H.; Yeo, H.; Park, S.S.; Nishizawa, T.; Kim, C.H. Impacts of local versus long-range transported aerosols on PM10 concentrations in Seoul, Korea: An estimate based on 11-year PM10 and lidar observations. *Sci. Total Environ.* **2021**, *750*, 141739. [CrossRef] [PubMed]
3. Chen, B.; Song, Z.; Shi, B.; Li, M. An interpretable deep forest model for estimating hourly PM10 concentration in China using Himawari-8 data. *Atmos. Environ.* **2022**, *268*, 118827. [CrossRef]
4. Tadano, Y.D.S.; Bacalhau, E.T.; Casacio, L.; Puchta, E.; Pereira, T.S.; Antonini Alves, T.; Siqueira, H.V. Unorganized machines to estimate the number of hospital admissions due to respiratory diseases caused by PM10 concentration. *Atmosphere* **2021**, *12*, 1345. [CrossRef]
5. Tırınk, S.; Öztürk, B. Evaluation of PM10 concentration by using Mars and XGBOOST algorithms in Iğdır Province of Türkiye. *Int. J. Environ. Sci. Technol.* **2023**, *20*, 5349–5358. [CrossRef]
6. Hong, W.Y.; Koh, D.; Yu, L.E. Development and Evaluation of Statistical Models Based on Machine Learning Techniques for Estimating Particulate Matter (PM2. 5 and PM10) Concentrations. *Int. J. Environ. Res. Public Health* **2022**, *19*, 7728. [CrossRef]
7. Mohammadi, Y.; Zandi, O.; Nasseri, M.; Rashidi, Y. Spatiotemporal modeling of PM10 via committee method with in-situ and large scale information: Coupling of machine learning and statistical methods. *Urban Clim.* **2023**, *49*, 101494. [CrossRef]
8. Hongthong, A.; Nanthapong, K.; Kanabkaew, T. Estimation of Respiratory Disease Burden Attributed to Particulate Matter from Biomass Burning in Northern Thailand Using 1-km Resolution MAIAC-AOD. *Appl. Environ. Res.* **2023**, *45*, 2. [CrossRef]
9. Shao, H.; Li, H.; Jin, S.; Fan, R.; Wang, W.; Liu, B.; Gong, W. Exploring the Conversion Model from Aerosol Extinction Coefficient to PM1, PM2. 5 and PM10 Concentrations. *Remote Sens.* **2023**, *15*, 2742. [CrossRef]
10. Lee, S.; Park, S.; Lee, M.I.; Kim, G.; Im, J.; Song, C.K. Air quality forecasts improved by combining data assimilation and machine learning with satellite AOD. *Geophys. Res. Lett.* **2022**, *49*, e2021GL096066. [CrossRef]
11. Tuna Tuygun, G.; Elbir, T. Estimation of particulate matter concentrations in Türkiye using a random forest model based on satellite AOD retrievals. *Stoch. Environ. Res. Risk Assess.* **2023**, *2023*, 3469–3491. [CrossRef]
12. Imani, M. Concentration Estimation of Air Pollutants (PM2. 5 and PM10) Using MODIS Satellite Data, Deep Neural Network and Random Forest. *Soft Comput. J.* **2023**, *12*, 1–19.
13. You, W.; Zang, Z.; Zhang, L.; Li, Z.; Chen, D.; Zhang, G. Estimating ground-level PM10 concentration in northwestern China using geographically weighted regression based on satellite AOD combined with CALIPSO and MODIS fire count. *Remote Sens. Environ.* **2015**, *168*, 276–285. [CrossRef]
14. Machupalli, R.; Hossain, M.; Mandal, M. Review of ASIC accelerators for deep neural network. *Microprocess. Microsyst.* **2022**, *89*, 104441. [CrossRef]
15. 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]
16. 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]
17. 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]
18. 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]
19. Kaveh, M.; Mesgari, M.S. Application of meta-heuristic algorithms for training neural networks and deep learning architec-tures: A comprehensive review. *Neural Process. Lett.* **2022**, *55*, 4519–4622. [CrossRef] [PubMed]
20. Kaya, U.; Yılmaz, A.; Aşar, S. Sepsis Prediction by Using a Hybrid Metaheuristic Algorithm: A Novel Approach for Optimizing Deep Neural Networks. *Diagnostics* **2023**, *13*, 2023. [CrossRef]

21. Khan, M.S.; Jabeen, F.; Ghouzali, S.; Rehman, Z.; Naz, S.; Abdul, W. Metaheuristic algorithms in optimizing deep neural network model for software effort estimation. *IEEE Access.* **2021**, *9*, 60309–60327. [CrossRef]
22. 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]
23. Kaveh, M.; Aghapour, S.; Martin, 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.
24. 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]
25. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [CrossRef]
26. Mirjalili, S.; Hashim, S.Z.M. BMOA: Binary magnetic optimization algorithm. *Int. J. Mach. Learn. Comput.* **2012**, *2*, 204. [CrossRef]
27. Emary, E.; Zawbaa, H.M.; Hassanien, A.E. Binary grey wolf optimization approaches for feature selection. *Neurocomputing* **2016**, *172*, 371–381. [CrossRef]
28. 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]
29. Montavon, G.; Samek, W.; Müller, K.R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **2018**, *73*, 1–15. [CrossRef]
30. Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.R. Explaining deep neural networks and beyond: A review of methods and applications. *Proc. IEEE* **2021**, *109*, 247–278. [CrossRef]
31. Gawlikowski, J.; Tassi, C.R.N.; Ali, M.; Lee, J.; Humt, M.; Feng, J.; Zhu, X.X. A survey of uncertainty in deep neural networks. *Artif. Intell. Rev.* **2023**, 1–77. [CrossRef]
32. Abdou, M.A. Literature review: Efficient deep neural networks techniques for medical image analysis. *Neural Comput. Appl.* **2022**, *34*, 5791–5812. [CrossRef]
33. Vafa-Arani, H.; Jahani, S.; Dashti, H.; Heydari, J.; Moazen, S. A system dynamics modeling for urban air pollution: A case study of Tehran, Iran. *Transp. Res. Part D Transp. Environ.* **2014**, *31*, 21–36. [CrossRef]
34. Habibi, R.; Alesheikh, A.A.; Mohammadinia, A.; Sharif, M. An assessment of spatial pattern characterization of air pollution: A case study of CO and PM2. 5 in Tehran, Iran. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 270. [CrossRef]
35. Wang, M.; Wang, Y.; Teng, F.; Li, S.; Lin, Y.; Cai, H. Estimation and Analysis of PM2. 5 Concentrations with NPP-VIIRS Nighttime Light Images: A Case Study in the Chang-Zhu-Tan Urban Agglomeration of China. *Int. J. Environ. Res. Public Health* **2022**, *19*, 4306. [CrossRef] [PubMed]
36. Wang, Y.; Wang, M.; Huang, B.; Li, S.; Lin, Y. Estimation and analysis of the nighttime PM2. 5 concentration based on lj1-01 images: A case study in the pearl river delta urban agglomeration of china. *Remote Sens.* **2021**, *13*, 3405. [CrossRef]
37. Vahidi, M.; Aghakhani, S.; Martín, D.; Aminzadeh, H.; Kaveh, M. Optimal Band Selection Using Evolutionary Machine Learning to Improve the Accuracy of Hyper-spectral Images Classification: A Novel Migration-Based Particle Swarm Optimization. *J. Classif.* **2023**, *2023*, 1–36. [CrossRef]
38. Mohammadi, R.; Sahebi, M.R.; Omati, M.; Vahidi, M. Synthetic aperture radar remote sensing classification using the bag of visual words model to land cover studies. *Int. J. Geol. Environ. Eng.* **2018**, *12*, 588–591.
39. Khajehyar, R.; Vahidi, M.; Tripepi, R. Determining Nitrogen Foliar Nutrition of Tissue Culture Shoots of Little-Leaf Mockorange By Using Spectral Imaging. In Proceedings of the 2021 ASHS Annual Conference, Denver, CO, USA, 5–9 August 2021; pp. 1–8.

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.