IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Carbon Emission Prediction through the Harmonization of Extreme Learning Machine and INFO Algorithm

Afi Kekeli Feda[1], Oluwatayomi Rereloluwa Adegboye[2], Ephraim Bonah Agyekum[3], Abdurrahman Shuaibu Hassan[4], Salah Kamel[5]

[1]Management Information System Department, Faculty of Economics and Administrative Sciences, European University of Lefke, Lefke, Northern Cyprus TR-10 Mersin, 99010, Turkey
[2]Management Information Systems, University of Mediterranean Karpasia, Mersin-10, Turkey
[3]Department of Nuclear and Renewable Energy, Ural Federal University named after the first President of Russia Boris Yeltsin, 620002, 19 Mira Street, Ekaterinburg, Russia
[4] Electrical Engineering Department, School of Engineering Technology, Soroti university Uganda
[5]Electrical Engineering Department, Faculty of Engineering, Aswan University, 81542 Aswan, Egypt

Corresponding author:

ABSTRACT This research introduces a novel optimization algorithm, weIghted meaN oF vectOrs (INFO), integrated with the Extreme Learning Machine (ELM) to enhance the predictive capabilities of the model for carbon dioxide (CO2) emissions. INFO optimizes ELM's weight and bias. In six classic test problems and CEC 2019 functions INFO demonstrated notable strengths in achieving optimal solutions for various functions. The proposed hybrid model, ELM-INFO, exhibits superior performance in forecasting CO2 emissions, as substantiated by rigorous evaluation metrics. Notably, it achieves a superior R2 value of 0.9742, alongside minimal values in Root Mean Squared Error (RMSE) at 0.01937, Mean Squared Error (MSE) at 0.00037, Mean Absolute Error (MAE) at 0.0136, and Mean Absolute Percentage Error (MAPE) at 0.0060. These outcomes underscore the robustness of ELM-INFO in accurately predicting CO2 emissions within the testing dataset. Additionally, economic growth is the most significant element, as indicated by ELM-INFO's permutation significance analysis, which causes the model's MSE to increase by 19%. Trade openness and technological innovation come next, each adding 7.6% and 8.1% to the model's MSE increase, respectively. According to ELM-INFO's performance, it's a powerful tool for developing ecologically sound policies that improve environmental resilience and sustainability.

INDEX TERMS **Artificial neural network, Carbon Emission Prediction, Convergence acceleration, Extreme Learning Machine, Metaheuristic algorithms,**

I. INTRODUCTION

The environment is vital to our continued existence on Earth, as we all know, and changes in it can either benefit or harm humans. One such alteration to the physical environment that has serious ramifications for human survival is climate change. Global ecosystems are universally threatened by climate change, which also impacts various facets of human life. Climate change endangers food security [1], disrupts environmental equilibrium [2], increases the frequency of natural disasters [3], gives rise to new diseases, and exacerbates water scarcity [4]. Moreover, it places stress on public health systems, introduces socioeconomic challenges, triggers unemployment, and induces migration. It is noteworthy that there is a unanimous acknowledgment among scholars and experts regarding the critical significance of the problem of climate change. A strong body of international scientific evidence supports the reality of this phenomenon. If recent trends in global warming continue, temperatures will increase, ocean levels will rise, and severe weather events like storms, heat waves, droughts, floods, and cyclones will become more common. These events could result in the persistence of devastating weather-related phenomena already observed, such as hurricanes (like Katrina and Rita in the USA), tsunamis, typhoons, flooding, particularly in the Asian Continent, wildfires, particularly in Australia and the USA[5]. Climate Change (CC) refers to human-induced changes in the average weather, encompassing variables like temperature, humidity,

precipitation, cloud cover, and wind patterns, as well as variations in the occurrence or intensity of these circumstances. Research indicates that the main cause of climate change is the emission of greenhouse gases, unequivocally triggering global warming that have led to a significant rise in the global surface temperature, reaching 1.1°C above the baseline of 1850-1900 within the timeframe of 2011-2020 [6]. Mitigating climate change involves transforming global energy production systems to curtail Green House Gas (GHG) production, necessitating a multidimensional approach. First an accurate modeling and prediction at individual, community, and country levels. Second, the incorporation of data science and analysis to enhance the understanding of emission trends and facilitate reduction strategies

Recent technological progress, specifically in artificial intelligence (AI) and machine learning (ML), provide valuable tools for modeling complex environmental phenomena associated with climate change [7]. The field of ML, succinctly described as empowering computers to make accurate predictions based on past experiences, has witnessed remarkable advancements, particularly fueled by the rapid enhancement of computer storage capacity and processing power. In conjunction with various other disciplines, machine learning methodologies have found extensive applications for modeling and prediction in diverse areas, including the prediction of food safety [8], the diagnosis of COVID-19 through X-ray and CT Images [9], Predicting Match Results in Team Sport [10]; damage detection [11] predicting patient response to therapy [12], predicting solar energy use [13], breast cancer detection [14], Weather Prediction [15], stock market trends [16], visitors' green behavior [17], prediction of organic solid waste treatment [18], predicting the consequences of construction accidents [19], Android malware detection [20], etc. Likewise, ML techniques have been employed in the context of climate change. Hamrani et al. investigated the possibilities of using three types of machine learning (ML) regression models: shallow learning deep learning and classical regression to forecast soil GHG releases from a region used for agriculture [21]. Mardani et al. employed clustering, dimensionality reduction, and prediction machine learning algorithms to establish an effective multi-stage system to estimate carbon dioxide production on the basis of two crucial variables, namely the energy consumption and economic growth [22]. Nguyen et al. suggested a machine learning method to ameliorate the prediction of business carbon emissions for risk assessments by investors. As the optimal emission forecasting method, they presented a two-step architecture that combines forecasts from various base-learners using a Meta-Elastic Net learner [23]. Another example is the work of Bakay and Ağbulut that used deep learning (DL), support vector machine (SVM), and artificial neural network (ANN) techniques to forecast GHG releases from Turkey's electricity generating industry [24]. There are many other research studies which have focused on the employment of Artificial neural networks (ANNs) in the context of climate change. Artificial neural networks (ANNs) are mathematical and information-processing models that are motivated by organic brain systems [25]. Their strong forecasting ability, capacity to represent dynamic and complicated systems, simplicity of use, and parallel structure are only a few of their numerous benefits, which have led to their widespread use in many classification and regression issues. ANNs are formed by a multitude of processing units referred to as "neurons" dispersed throughout several layers. One of the most commonly used neural network topologies is single-hidden-layer feedforward neural networks (SLFN). SLFN are general approximators that can approximate every continuous function, as demonstrated by the literature. Usually, gradient descent techniques like Backpropagation (BP) are used to train SLFNs. Despite their widespread use, gradient descent dependent training methods, like BP, have significant limitations, including sluggish convergence, a high likelihood of getting caught in a local minimum, and a strong dependence on the network's initial weights. Extreme learning machine (ELM), a novel machine learning algorithm was introduced as a means of training single-hidden-layer feed-forward neural networks (SLFNs) with the aim to address the shortcomings of gradient descent training techniques [26]. The input weights and biases, which are ELM learning parameters, are assigned at random and do not require tuning, whereas the output weights are derived analytically via a straightforward generalization reverse procedure. Due to its lack of iteration, ELM can therefore complete training considerably more quickly than standard algorithms. Additionally, ELM can escape several issues related to conventional gradient base methods, including local minimums and learning speed.

Despite its advantages, ELM is confronted with certain limitations [27]. The initial configuration of weights and biases plays a pivotal role in its overall effectiveness. Various strategies have been suggested in scholarly works to address these challenges and enhance the efficacy of ELM networks. A category of strategies that have garnered significant attention involves the utilization of metaheuristic-based methods. Metaheuristic algorithms represent efficient techniques tailored to yield satisfactory or nearly optimal solutions for complex optimization problems. These algorithms systematically guide the search process to explore the solution space effectively, aiming to achieve greater effectiveness. The majority of these algorithms draw inspiration from physical or biological systems. Noteworthy advantages include their problem-independent nature, stochastic guidance in the search stage to discover nearly optimum solutions, and applicability to problems spanning from straightforward searches to intricate scenarios. Shariaty et al., introduced a novel

approach to forecast the compressive strength of concrete with partial cement replacements [28]. By joining extreme learning machine (ELM) with the grey wolf optimizer (GWO) algorithm, the proposed ELM-GWO model outperforms five other well-known machine learning models. The outcomes indicate that this merged algorithm achieves superior performance, providing an efficient alternative for predicting compressive strength, especially in scenarios involving concrete with partial cement replacements. Boriratrit et al. integrated the Jellyfish Search Extreme Learning Machine (JS-ELM) and employed the hybrid model for energy demand prediction [29]. Data on actual electric energy demand in Thailand from 2018 to 2020 were gathered and utilized to assess and contrast the effectiveness of the introduced model against existing forecasting methods. The comprehensive findings indicate that the JS-ELM outperforms other forecasting methods, displaying the lowest root mean square error. Additionally, the JS-ELM exhibits optimal processing time during the course of this experiment. Qiu et al. addressed the challenge of estimating the uniaxial compressive strength (UCS) of rocks, a crucial parameter in engineering projects for disaster mitigation, utilizing extreme learning machines (ELM) optimized by the whale optimization algorithm (WOA) [30]. Performance evaluation reveals that the WOA-ELM model outperforms others, demonstrating smaller relative errors (0.22%, 72.05%, and 11.48%) and residual errors (0.02 and 2.64 MPa). The derived UCS values from WOA-ELM exhibit superior accuracy, indicating its potential for widespread application in estimating UCS for various rock types. In an attempt to predict super capacitor capacity Li et al., proposed an advanced predicting model, merging extreme learning machine (ELM) with the metaheuristic Kalman filter (HKF) algorithm [31]. This HKF-ELM model outperforms common data-driven models for supercapacitor life forecasting, showing significant improvements over traditional ELM, Kalman filtering, and other methods.

Notwithstanding the benefits of ELM stated in existing literature, ELM still suffers from a few limitations. Notably, the performance of ELM is significantly influenced by the initialization of its structure, with the model's effectiveness hinging on the initial weights and biases. The stochastic nature of hidden bias and random input weight selection introduces variability that may lead to suboptimal outcomes, impeding the model's generalization capability. In addressing these challenges, our research introduces a novel solution by leveraging the INFO algorithm (Weighted Mean of Vectors). This choice is motivated by the algorithm's sophisticated weighted mean methodology, encompassing three critical processes: updating rule, vector combining, and local search. This novel approach serves as an innovative strategy to overcome the shortcomings of ELM, particularly in the context of CO2 emission prediction.

The remaining portions of this article are thus structured: Section 2 presents a brief introduction to ELM and INFO. Section 3 details the evaluation of INFO algorithm on test functions. In Section 4, a comprehensive explanation of the proposed ELM-INFO is provided. Section 5 explains the optimization and evaluation of ELM-INFO. Section 6 covers the experiments and results, followed by discussions and the significance of the study. The conclusion and future work are presented in Section 7.

II. Preliminary

A. INFO

The INFO algorithm functions as a population inspired optimization technique, primarily relying on the computation of the weighted mean for a group of vectors within the search region [31]. These vectors serve as representations of possible solutions. The algorithm progresses through three steps, namely updating rule, vector combining, and local search, to adjust the positions of these vectors. The initialization process of the INFO algorithm involves randomly generating a population of vectors according to the definition provided below.

$$X_{l,j}^{g} = \{x_{l,1}^{g}, x_{l,2}^{g}, ..., x_{l,D}^{g}\} \quad (1)$$

With $l = 1,2, ..., Np$; Np denoting the population of vectors, and D the search region's dimension. The weighted mean factor (δ) and scaling factor (σ), which vary dynamically throughout generations, are also employed in the initialization phase and are provided in Equation (2) and Equation (3).

$$\delta = 2\beta \times \text{rand} - \beta \quad (2)$$

$$\sigma = 2a \times \text{rand} - \sigma \quad (3)$$

With $\beta = 2e^{-4(g/Maxg)}$ and $\alpha = ce^{-d(\frac{g}{Maxg})}$. Here, g signifies the current generation, and $Maxg$ denotes the maximum number of generations. Additionally, c is equal to 2 and d is equal to 4. The increase of population diversity is accomplished through the utilization of the updating rule operator, which employs the weighted mean of vectors to generate novel ones. Within the INFO algorithm, the determination of the weighted mean involves the utilization of a number of differentially selected vectors, as opposed to directing the vector towards an improved solution. To enhance population diversity, a *MeanRule*, as shown in equation (4), is incorporated:

$$MeanRule = r \times WM1_{l}^{g} + (1 - r) \times WM2_{l}^{g} \quad (4)$$

$$WM1_{l}^{g} = \delta \times \frac{w_1(x_{a1} - x_{a2}) + w_2(x_{a1} - x_{a3}) + w_3(x_{a2} - x_{a3})}{w_1 + w_2 + w_3 + \varepsilon} + \varepsilon \times \text{rand} \quad (5)$$

$$WM2_l^{\mathrm{g}} = \delta \times \frac{w_1(x_{bs}-x_{bt})+w_2(x_{\mathrm{bs}}-x_{ws})+w_3(x_{bt}-x_{ws})}{w_1+w_2+w_3+\varepsilon} + \varepsilon \times \text{rand} \quad (6)$$

w_1, w_1 and w_1 denote wavelet functions [15]. $l = 1,2,\ldots,Np$; r is a number generated randomly from the [0, 0.5]; and $a1; a2; a3$ are distinct integers chosen randomly from the range $[1, Np]$. Additionally, rand signifies a normally distributed arbitrary value while ε represents a small constant. On the other hand of $x_{ws,}$ x_{bt} and x_{bs}, the worst, better and best solutions. A convergence acceleration (CA) component is incorporated in the updating rule operator as shown in equation (7). Here randn is a randomly generated number having a normal distribution.

$$CA = \text{randn} \times \frac{(x_{bs}-x_{a1})}{(f(x_{bs})-f(x_{a1})+\varepsilon)} \quad (7)$$

The computation of the new vector, as outlined in equation (8), is achieved by employing the previous equations:

$$Z_l^g = x_l^g + \sigma \times \mathit{MeanRule} + CA \quad (8)$$

The updating rule, formulated depending on $x_{\mathrm{bs},}$ x_{bt}, x_L^Z and x_{a1}^3, is specified to generate new vectors denoted as $z1_1^s$ and $z2z_l^s$ in the g^{th} generation. To determine these vectors in the current generation, a conditional mechanism is introduced based on a randomly generated number. If this number is less than 0.5, equations (9) and (10) come into play.

$$z1_l^g = x_l^g + \sigma \times \text{MeanRule} + \text{randn} \times \frac{(x_{bs}-x_{a1}^g)}{(f(x_{bs})-f(x_{a1}^g)+1)} \quad (9)$$

$$z2_l^g = x_{bs} + \sigma \times \text{MeanRule} + \text{randn} \times \frac{(x_{a1}^g-x_b^g)}{(f(x_{a1}^g)-f(x_{a2}^g)+1)} \quad (10)$$

Equations (11) and (12) present an alternative updating mechanism when the randomly generated number exceeds 0.5. In this case, the algorithm employs Equation (11) to compute $z1_1^s$ and Equation (12) to compute $z2z_l^s$. This demonstrates the INFO algorithm's ability to adjust its methods by picking different strategies randomly. This adaptability lets INFO handle various problem-solving situations more effectively.

$$z1_l^g = x_a^g + \sigma \times \text{MeanRule} + \text{randn} \times \frac{(x_{a2}^g-x_{a3}^g)}{(f(x_{a2}^g)-f(x_{a3}^g)+1)} \quad (11)$$

$$z2_l^g = x_{bt} + \sigma \times \text{MeanRule} + \text{randn} \times \frac{(x_{a1}^g-x_{a2}^g)}{(f(x_{a1}^g)-f(x_{a2}^g)+1)} \quad (12)$$

During the vector combining phase, vector x_1^s is merged with the acquired vectors of $\mathrm{z1}_l^g$ and $\mathrm{z1}_i^g$ to produce the new vector u_1^g.

The introduction of the parameter $\mu = 0.05 \times \text{randn}$, adds a stochastic element to the merging process. The conditions within these equations further contribute to the algorithm's versatility. If a randomly generated number is less than 0.5, the algorithm selects between two merging strategies depending on the value of another random generated as outlined in equations (13) and (14) , each involving a weighted combination of the acquired vectors and their absolute differences. Alternatively, if the random number exceeds 0.5, the merging process simply involves retaining the current vector x_i^{g} (equation 15).

if rand < 0.5:

if rand < 0.5.

$$u_1^g = z1_i^g + \mu \cdot |z|_i^g - \mathrm{z2}_i^g| \quad (13)$$

else

$$u_1^g = z2_i^g + \mu \cdot |z1_i^g - \mathrm{z2}_i^g| \quad (14)$$

end .

else

$$u_1^g = x_i^{\mathrm{g}} \quad (15)$$

The utilization of the vector combining stage is primarily aimed at executing exploitation. The INFO algorithm also incorporates the local search phase to enhance its efficiency, thereby preventing local optimum solutions. In this subsequent stage, the local search agent takes into account both the global optimum $\left(x_{\mathrm{best}}^{\mathrm{g}}\right)$ and adheres to the mean-based rule outlined in equation (16).

$$WM = \frac{x_1 \times w_1 + x_2 \times w_2}{w_1 + w_2} \quad (16)$$

The following can be used to create a new vector close to $x_{\mathrm{best}}^{\mathrm{g}}$ best:

if rand < 0.5

$$u_i^g = x_{ix} + \text{randn} \times (\text{MeanRule} + \text{randn} \times \left(x_{bu}^g - x_{a1}^g\right)) \quad (17)$$

else

$$u_l^g = x_{md} + \text{randn} \times (\text{MeanRule} + \text{randn} \times (v_1 \times x_{bs} - v_2 \times x_{md})) \quad (18)$$

Here, x_{md} denotes a novel solution that further amplifies the algorithm's unpredictability. It can be written as follows: $x_{md} = \varphi \times x_{avg} + (1-\varphi) \times (\varphi \times x_{br} + (1-\varphi) \times x_{bs})$; with $x_{avg} = (x_a + x_b + x_3)/3$. φ is a randomly selected

from (0,1) while v_1 and v_2 In contrast, v1 and v2, which have the following definitions and are two arbitrary numbers, enhancing the influence of the optimal position on the vector:

$$v_1 = \begin{Bmatrix} 2 \times \text{rand}, p > 0.5 \\ 1, \text{otherwise} \end{Bmatrix} \quad (19)$$

$$v_2 = \begin{Bmatrix} \text{rand}, p < 0.5 \\ 1, \text{otherwise} \end{Bmatrix} \quad (20)$$

Here p is a random number from (0, 1).

B. Classical ELM

The Extreme Learning Machine (ELM) has proven to be a successful alternative to training neural networks with the backpropagation (BP) algorithm. ELM is distinguished by its capacity to address the limitations associated with training Single Layer Feedforward Networks (SLFNs) using BP, including local minima and time limitations [32]. There are two essential phases in the ELM learning stage: (1) randomly assigning weights for the relations between the input layer and hidden layer, along with biases, then by the generation of the hidden layer output matrix H. (2) Determining the outcome weights utilizing the least square algorithm. Remarkably, ELM reduces computational effort by converting the learning procedure into a linear system's solution. We seek for the least squares solution β of the linear system Hβ = T in order to train Single Layer Feedforward Networks (SLFN) as effectively as possible. Assuming that N unique samples (x_i, t_i), K neurons in the hidden layer, and an activation function g(x) are to be used in the training of SLFNs. $x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T$ is the n-dimensional input vector of the ith sample. The resultant vector is denoted by $t_i = [t_{i1}, t_2, ..., t_{i1}]^T$. This type of set will contain the following components: the hidden layer bias ($b_{K\times1}$); the input weights; and output weights ($\beta_{1\times xK}$.). The form of the ELM's output function is provided by Equation (21).

$$f_K(x) = \sum_{j=1}^{K} \beta_j h_j(x) = h(x)\beta \quad (21)$$

Here β_j denotes the weight vector connecting the output neurons (≥1) with the hidden neuron j. The consolidated weight vector is denoted as $\beta = [\beta_1, \beta_2, ..., \beta_K]$, connecting the hidden layer to the output layer with a minimum of one neuron, and the hidden layer's output is represented by $h(x) = [h_1(x), h_2(x), ..., h_K(x)]$. $h(x)$ in a specific application can be written as:

$$h_j(x) = G(w_j, b_j, x), w_i, x \in R^d, b_i \in R \quad (22)$$

Here G is a piecewise continuous, nonlinear function. There are numerous activation functions that can be employed in the hidden layer's hidden neurons. The sine function, hardlimit function, and sigmoid function are some of the most often utilized ones. A single instance of the training samples is represented by x , and (w_j, b_j) are the configuration variables of the j_k hidden neuron.
Equation 21 can also be expressed as $H\beta = T$, with $H_{N\times K}$ representing the hidden layer's output matrix.

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} h(x_1) \cdots h_K(x_1) \\ \vdots \\ h(x_N) \cdots h_K(x_N) \end{bmatrix} = \begin{bmatrix} G(w_1, b_1, x_1) & \cdots & G(w_K, b_K, x_1) \\ \vdots & & \vdots \\ G(w_1, b_1, x_N) & \cdots & G(w_K, b_K, x_N) \end{bmatrix} \begin{matrix} x_1 + b_M) \\ \vdots \\ x_N + b_M) \end{matrix}_{NxM} \quad (23)$$

Where the weight vector linking the input neurons to the ith hidden neuron is represented by $W_i = [W_{i1}, W_{12}, ..., W_{\text{it}}]^T$ .A sample at ith position in the training set is represented by $x_i = [x_{i1}, ..., x_{in}]$, T stands for the desired result, β for the output weight matrix, and b_i for the bias value of the ith hidden neuron. In order to improve the feedforward neural network's generalization efficiency, ELM seeks to obtain the lowest output weight norm in addition to the lowest training error.
Minimize:

$$\| H\beta - T\|^2, \| \beta \| \quad (24)$$

The ELM-prepared SLFN in Figure 1 has arbitrarily initialized weights and biases connecting the input neurons and hidden neurons. The weights between the hidden neurons and the output layer are calculated through analysis using the system's least squares solution, $\widehat{\beta} = H^{\dagger}T$, with H† standing for the MP generalized inverse of the matrix obtained in equation (23).

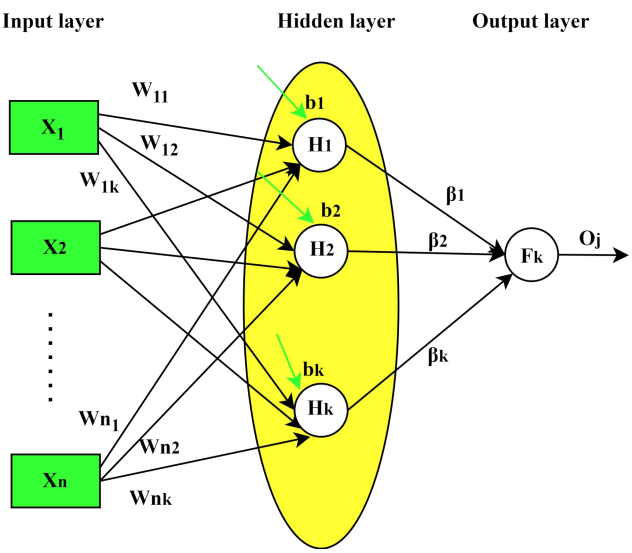

FIGURE 1. Extreme Learning Machine.

III. Performance Evaluation of Proposed INFO Algorithm

A. Classic Functions

A detailed comparison with well-known algorithms, such as the Grey Wolf Optimizer [33], Salp Swarm Algorithm [34], and Sine Cosine Algorithm [35], has been done in order to properly understand the efficacy of INFO. This evaluation confirms that the INFO algorithm is competent for CO2 emission prediction. Four metrics (average, standard deviation, Friedman rank, and Wilcoxon test) have been applied in order to systematically assess the output generated by each method. To put it another way, better performance is shown by a lower mean, and more stability is shown by a smaller standard deviation. A good measure for analysing how the newly suggested algorithm performs overall is the Friedman Rank, where a smaller Friedman Value indicates better algorithm performance overall. Furthermore, the approach performs noticeably better than the comparison algorithm if the Wilcoxon P-value is less than 0.05.

To uphold experiment fairness, uniform parameters are maintained across all comparison algorithms. The number of populations is fixed at 30, and the maximum iteration is limited to 200. Specific parameters for each algorithm can be found in Table I. Broadly, the search capabilities of algorithms can be categorized into exploration ability, exploitation ability, and the skill to balance these two capacities is essential. Exploration ability signifies an algorithm's capability to focus on searching around a specific region for an optimal solution. Exploitation ability, on the other hand, characterizes an algorithm's proficiency in traversing the entire problem space to discover regions with optimal solutions.

TABLE I
ALGORITHM PARAMETERS

Optimizer	Settings
GWO	$a_0 = 2$
INFO	$c = 2, d = 4$
SSA	c1 $=[2/e, 2]$
SCA	$a = 2$

Through a comprehensive comparison and analysis of the evaluation outcomes for INFO and other algorithms on the classical optimization test functions [36] set from Table II, we can gain insights into the outstanding performance of the INFO algorithm. In Table II, F1-F3 correspond to unimodal functions, serving as tests for an algorithm's exploitation ability since they possess a single optimal solution.

TABLE II
OPTIMIZATION TEST FUNCTIONS

Function	Range	Dim	Fmin
$f_1(x) = \sum_{i=1}^{n} x_i^2$	[-100,100]	30	0
$f_2(x) = \sum_{f_{\min}}^{n} \lvert x_i \rvert + \prod_{i=1}^{n} \lvert x_i \rvert$	[-10,10]	30	0
$f_3(x) = \sum_{i=1}^{n} \left(\sum_{j-1}^{i} x_j\right)^2$	[-100,100]	30	0
$f_4(x) = \sum_{i=1}^{n} - x_i \sin\left(\sqrt{\lvert x_i \rvert}\right)$	[-500,500]	30	-418.9892xdim
$f_5(x) = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left((1/n)\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + e$	[-32,32]	30	0
$f_6(x) = \pi/n\{\sum_{i=1}^{n-1}(y_i - 1)^2[1 + 10\sin^2(\pi y_{i+1})] + (y_n - 1)^2\} + \sum_{i=1}^{n} u(x_i, 10,100,4) + \pi/n 10\sin(\pi y_1)$ $y_i = 1 + x_i + (1/4)u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < -a \end{cases}$	[-50,50]	30	0

As illustrated in Figure 2, these functions exhibit less complex landscapes, with the darkest shade of purple indicating the optimal solution. On the other hand, F4-F6 represents multi-modal functions, assessing the exploration ability of algorithms. These functions are characterized by multiple local optimal solutions and a single global optimum solution; as depicted in Figure 2, they have a more complex landscape.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3390408

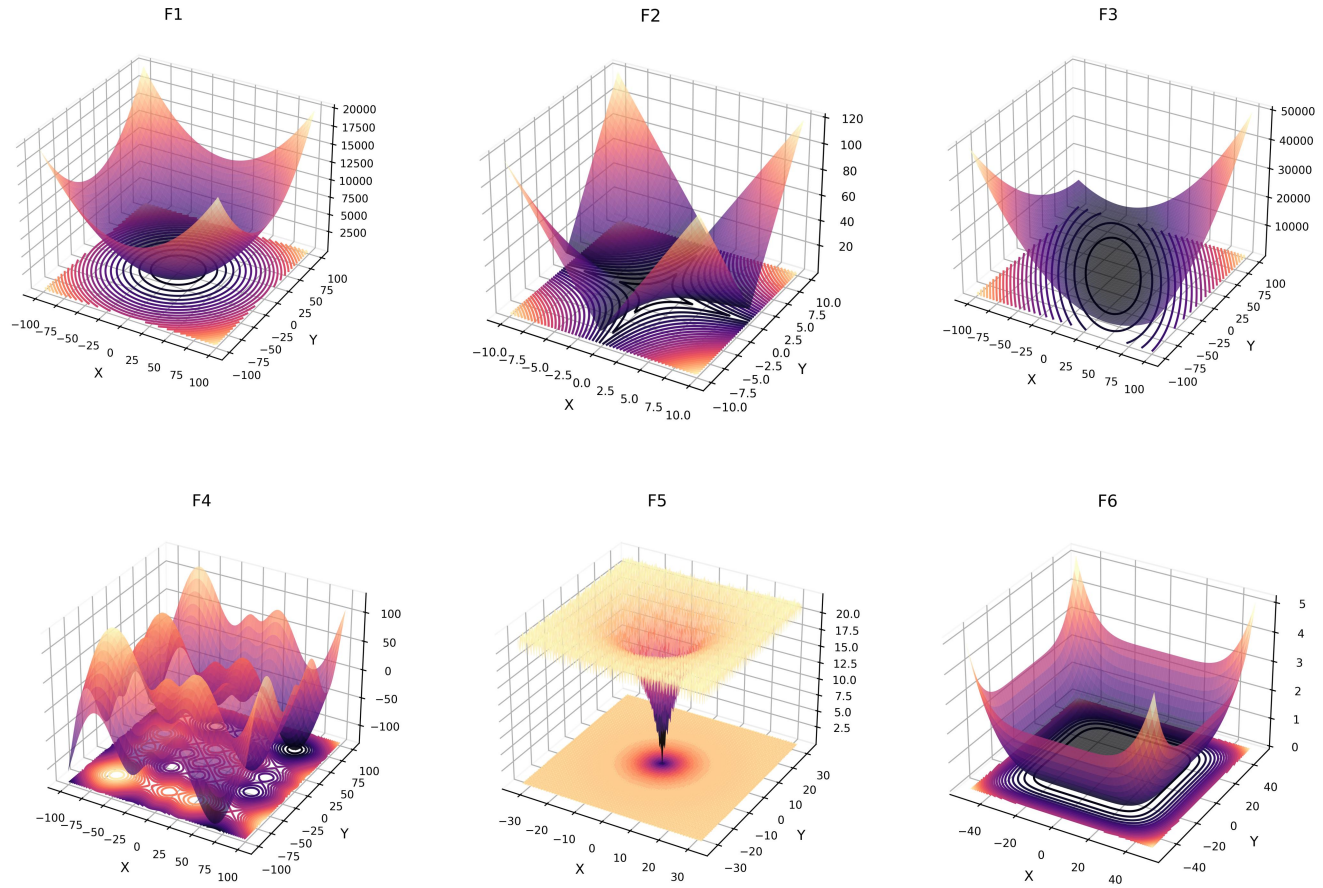

FIGURE 2. **Landscape Plot of Optimization Test Functions**

The test function outcomes for INFO, along with other comparative algorithms, are presented in Table III for unimodal functions with 30 dimensions. The results in Table III show that INFO surpasses the performance of SCA, SSA, and GWO on F1-F3. Specifically, INFO consistently identifies solutions closest to the theoretical value of 0 for F1-F3, outperforming all other algorithms in terms of average results and standard deviation. GWO performed second best to INFO, followed by SSA, while the worst-performing optimizer is SSA. Taking into account the outcomes of F1-F3 functions, these experimental findings conclusively demonstrate that INFO possesses a stronger exploitation ability. The test results for INFO and other comparison algorithms for multi-modal functions are also listed in Table III. Table III makes it clear that out of all the tested multi-modal functions, INFO achieves the most competitive performance compared to SSA and SCA. Both the average and the standard deviation of results generated by INFO are significantly better than those of other comparison algorithms. The testing findings clearly demonstrate INFO's superior exploration capability over other algorithms in these functions, F4–F6.

TABLE III
OPTIMIZATION TEST FUNCTIONS RESULTS OF GWO, INFO, SCA AND SSA

		GWO	INFO	SCA	SSA
F1	AVG	3.4600E-10	**3.4600E-39**	1.0202E+3	2.6783
	STD	3.1200E-10	**5.0800E-40**	9.6153E+2	1.7393
F2	AVG	8.8400E-7	**7.0000E-21**	2.5454	4.0166
	STD	4.9500E-7	**4.5900E-21**	2.3207	1.7268
F3	AVG	7.7681	**2.6900E-30**	2.3497E+4	2.4879E+3
	STD	7.6407	**1.6500E-30**	6.8372E+3	9.5660E+2
F4	AVG	-1.1980E+3	**-1.5883E+3**	-9.1186E+2	-1.4241E+3
	STD	1.0009E+2	9.5471E+1	**8.8935E+1**	1.1048E+2
F5	AVG	3.3900E-6	**6.2200E-16**	1.6311E+1	3.7573
	STD	1.5800E-6	**5.6500E-16**	6.2346	1.0234
F6	AVG	8.5442E-2	**1.5594E-2**	2.9754E+7	8.1217
	STD	8.0246E-2	**7.5300E-3**	2.7454E+7	3.1321

Convergence graphs of INFO, GWO, SCA, and SSA for the test functions are shown in Figure 3. Of all the methods, INFO has the fastest convergence for the unimodal functions (F1–F3). On the other hand, INFO exhibits the best results considering the convergence speed and accuracy for the multi-modal functions (F4-F6). This discovery bolsters the accepted conclusion that, in comparison to the other algorithms, INFO has greater convergence performance [37]. Additionally, the plots' convergence trajectory indicates a faster convergence to the ideal outcome, which is crucial for improving the ELM's weight and bias. This faster convergence improves the optimization algorithm's effectiveness while also improving the ELM model's overall performance by allowing it to quickly adjust to and capture information from the training set. Faster optimisation reduces the amount of time and computational work needed for training, which is especially useful when working with large-scale datasets or in situations where computational resources are scarce.

It is inadequate to evaluate optimisation algorithms only on the basis of mean and standard deviation values [38], [39]. Two well-known non-parametric statistical tests are used in this study to evaluate the improvements provided by INFO in more detail. First and foremost, one uses the Wilcoxon test. The null hypothesis is rejected if the P-value, or Wilcoxon probability value, is equal to or higher than 0.05. This suggests that the compared algorithms do not vary statistically significantly [40].

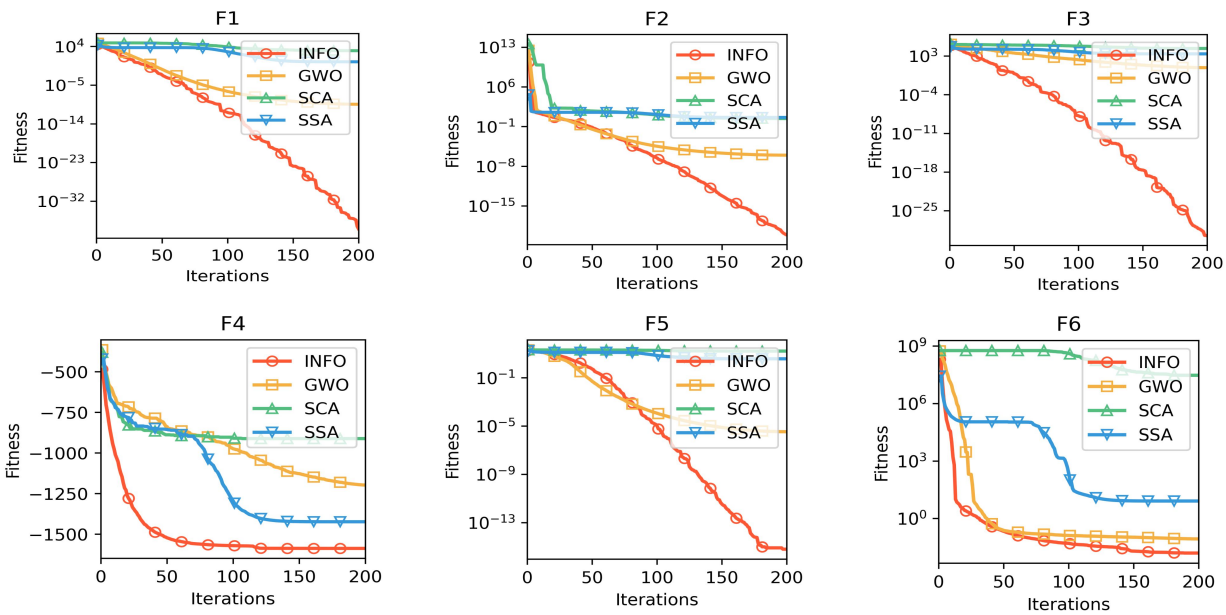

FIGURE 3. Optimization Test Functions Convergence of GWO, INFO, SCA and SSA

In contrast, the null hypothesis is accepted if the P-value is smaller than 0.05, indicating a significant difference between the procedures that were compared. Each algorithm's overall performance over a variety of benchmark functions is measured by the Friedman Value [38], [39]. With a Friedman Value of 1.0416, INFO secured the top spot among the algorithms, as can be observed in Table IV. The superiority of INFO is further shown by the related Friedman Rank of 1. Furthermore, evaluating the statistical significance of the variations between INFO and each of the other methods depends heavily on the Wilcoxon P-Values. With a P-value of 0.035 for GWO, a statistically significant difference is shown. Similarly, INFO shows notable differences for SCA and SSA, with P-Values of 0.0135 and 0.0005, respectively, confirming its superiority over the other optimisation methods in handling challenging optimisation situations.

TABLE IV
WILCOXON AND FRIEDMAN NON-PARAMETRIC TESTS

	GWO	INFO	SCA	SSA
Friedman Value	2.1250	1.0416	3.8916	2.9416
Friedman Rank	2	1	4	3
Wilcoxon P-Value	3.5035E-2	-	1.3489E-2	4.9278E-4

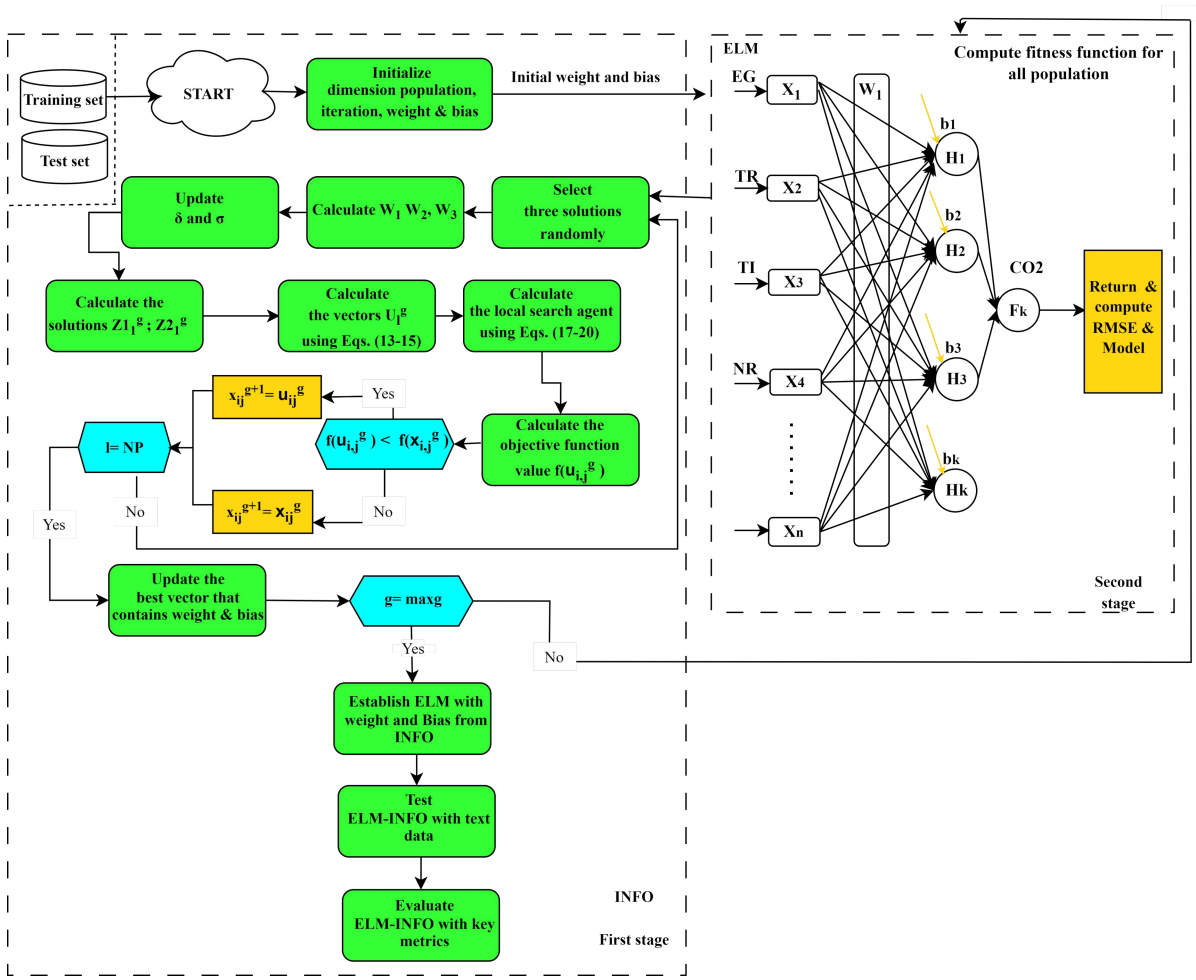

FIGURE 4. **Flow Chart of ELM-INFO Prediction Model**

B. CEC 2019 Functions

The Ten CEC 2019 benchmark functions [41] will allow for a more rigorous comparison of INFO method against existing approaches. The section maintains the same parameters used in the classic functions. The average and standard deviation of 30 runs is indicated in table V. The comparison of the INFO algorithm against other optimization algorithms (GWO, SCA, and SSA) across CEC 2019 optimization functions is expressed in Table V. Notably, INFO demonstrates superior performance on Function C1 by achieving the global optimal solution, surpassing all other algorithms with an average value of 1, indicating its effectiveness in solving a complex optimization problem. Also, INFO performs competitively on several other functions, such as C2, C3, C5, C6, C7, C8, and C9. In some cases, such as Function C4, INFO outperforms certain algorithms but falls slightly behind GWO. However, INFO generally exhibits lower standard deviation values compared to its counterparts, suggesting greater stability in its performance. These findings underscore the INFO algorithm's ability to tackle challenging optimization tasks.

TABLE V: RESULTS OF OPTIMIZERS ON CEC 2019 FUNCTIONS

		GWO	SCA	SSA	INFO
C1	AVG	4.0334E+5	1.1579E+7	2.1551E+6	**1**
	STD	3.5854E+5	1.0185E+7	1.7763E+6	**0**
C2	AVG	1.0324E+3	6.7669E+3	1.6047E+3	**4.5548**
	STD	4.0951E+2	2.7778E+3	9.8096E+2	**2.8660E-1**
C3	AVG	4.32	1.0046E+1	4.7244	**3.3635**
	STD	2.3453	2.6227	2.1159	**1.1842**
C4	AVG	**2.3061E+1**	5.6099E+1	3.1598E+1	2.7362E+1
	STD	1.1510E+1	1.2077E+1	1.3877E+1	**1**
C5	AVG	2.7286	1.4560E+1	1.1864	**1.1522**
	STD	2.4838	6.0108	1.7579E-1	**1.0435E-1**
C6	AVG	**3.2864**	8.8210	4.5492	3.5979
	STD	**1.1432**	1.3580	1.9419	1.2462
C7	AVG	1.1160E+3	1.7625E+3	1.0538E+3	**1.0454E+3**
	STD	5.1030E+2	3.5302E+2	3.5216E+2	**1.6061E+2**
C8	AVG	4.2256	4.7429	4.3458	**4.0981**
	STD	4.4283E-1	3.5147E-1	4.2650E-1	**2.6014E-1**
C9	AVG	1.2545	1.7949	1.3694	**1.2122**
	STD	9.5083E-2	2.5290E-1	1.6255E-1	**7.9582E-2**
C10	AVG	2.1570E+1	2.1569E+1	**2.1045E+1**	2.1336E+1
	STD	1.3436E-1	**9.8020E-2**	1.0692E-1	9.9109E-2

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3390408

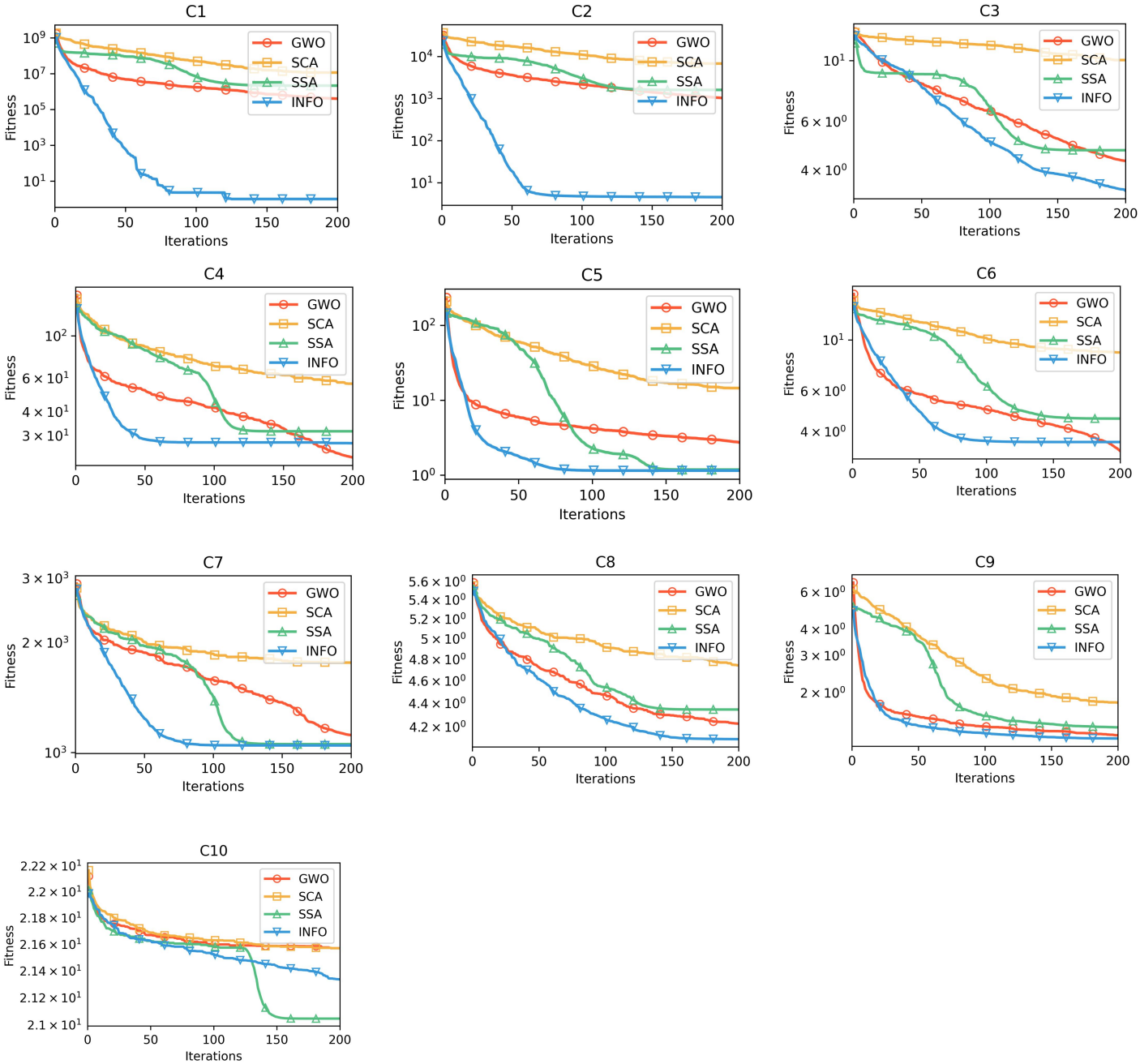

FIGURE 5. **Convergence Curve of GWO, INFO, SCA, and SSA on CEC 2019**

The INFO algorithm demonstrates strong performance on the CEC 2019 test functions when compared to GWO, SCA, and SSA in Figure 5. It exhibits rapid convergence initially, particularly notable in functions C1, C2, C3, and C5, often matching or surpassing the other algorithms' early performance. INFO's stability is a key strength, maintaining a consistent convergence rate across iterations, which is evident in functions C7 and C8. Moreover, the algorithm's ability to find high-quality solutions is highlighted in function C9, where it achieves lower fitness values by the final iterations. This consistent behavior across various functions suggests that INFO is a robust option for solving complex optimization problems, with a balance of speed, stability, and solution accuracy that often outperforms its counterparts in the given test scenarios.

TABLE VI: WILCOXON AND FRIEDMAN NON-PARAMETRIC TESTS ON CEC 2019

	GWO	INFO	SCA	SSA
Friedman Value	2.15	1.35	3.95	2.55
Friedman Rank	2	1	4	3
Wilcoxon P-Value	4.2601E-02	-	5.0620E-03	2.8402E-02

The Friedman test and Wilcoxon signed-rank test were conducted to compare the performance of the optimization algorithms (GWO, INFO, SCA, SSA) across CEC2019 functions in Table VI. According to the Friedman value, INFO achieved the best rank with a value of 1.35, indicating its overall superior performance compared to the other algorithms on CEC 2019. GWO and SSA attained Friedman values of 2.15 and 2.55, respectively, placing them in the second and third positions, while SCA had the worst Friedman value of 3.95, indicating comparatively poorer performance. The Friedman rank further corroborates INFO's dominance, ranking it first among the algorithms. The Wilcoxon P-values provide statistical significance to these findings, with INFO exhibiting significantly better performance compared to GWO ($p = 0.0426$), SCA ($p = 0.005$), and SSA ($p = 0.0284$). These results underscore the effectiveness of the INFO algorithm in optimization tasks and its potential as a preferred choice for solving such problems.

IV. ELM-INFO CO_2 Prediction Model

The INFO algorithm, is a novel optimization technique aimed at improving the performance of the Extreme Learning Machine (ELM) by optimizing its weight and bias parameters. It operates through three main processes:

Updating Rule: INFO adjusts ELM weight and bias parameters iteratively using a weighted mean approach based on the fitness (objective function) value of individual vectors representing weight and bias in the search space. This adaptive updating mechanism enables efficient exploration of the solution space, facilitating rapid convergence towards optimal solutions.

Vector Combining: INFO combines information from multiple candidate solutions by synthesizing vectors based on their fitness values. This strategy enhances diversity in the search process, preventing premature convergence to suboptimal solutions and promoting robustness.

Local Search: INFO incorporates a local search mechanism to refine candidate solutions near promising regions. By perturbing existing solutions and evaluating their fitness in the neighborhood, INFO fine-tunes parameters and improves solution quality, balancing global exploration with local exploitation.

Advantages of the INFO approach include:

1. Efficiency and Scalability: INFO's weighted mean methodology enables rapid convergence, making it suitable for large-scale optimization problems like carbon emission prediction.
2. Robustness and Versatility: INFO's adaptive updating rule and stochastic exploration strategy allow it to handle diverse optimization landscapes effectively, ensuring robust navigation of complex solution spaces.
3. Improved Generalization and Performance: INFO optimizes ELM parameters to dataset characteristics, leading to superior generalization capabilities and performance compared to conventional ELM variants. This adaptability enhances predictive accuracy and reliability in real-world applications like carbon emission prediction.

Algorithm 1 : Pseudocode INFO - ELM

Split data into two parts (training and testing)
Initialize the population size, problem dimension, maximum iteration , weights and bias randomly
Compute the fitness of each individual in the population (ELM Model)
For g < = Max_g
  For i < = N_p
    Select three solutions randomly
    Compute w_1, w_2, w_3
    Update δ and σ
  Updating rule stage
    Compute the solutions $z1_1^g$ and $z2_l^g$ using Eqs. (9-12)
  Vector combining stage
    Compute the vectors U_1^g using Eqs. (13-15)
  Local search stage
    Compute the local search agent using Eqs. (17-20)
    Compute the objective function value $f(U_{i,j}^g)$
      If $f(U_{i,j}^g) < f(x_{i,j}^g)$ then $x_{i,j}^{g+1} = u_{i,j}^g$
      Else $x_{i,j}^{g+1} = x_{i,j}^g$
  End for
  Calculate the objective function value
End for
Return Fitness of best individual
Use Fitness to Train ELM model
Test ELM-INFO with text data
Evaluate ELM-INFO with key metrics

Training and prediction are the two stages of the ELM-INFO model's CO2 emission prediction process. The ELM-INFO model is trained using a dataset of historical CO2 emissions and related input characteristics during the training phase. This is accomplished by minimising the error between the estimated and real CO_2 emissions by optimising the weights and biases of the ELM method. The ELM-INFO method is used to estimate CO_2 emissions using a fresh set of data points that weren't in the training

set during the prediction stage. According to Figure 4, the initial step involves splitting the dataset into training and testing sets. Next, the parameters of the INFO algorithm are initialized, including the population size, problem dimension, maximum iterations, and the random initialization of weight and bias values within the range -1 to 1 for the ELM model. Subsequently, the fitness function of each individual in the population is determined based on the initial weight and bias values. These weight and bias values are utilized to train the ELM model on the training dataset, and the root mean squared error (RMSE) score of the model is returned as the fitness score for that individual. The INFO operators then continuously modify the vectors representing each individual in the population. If the maximum number of iterations is not attained, the process returns to the step of calculating the fitness of each individual and adjusting their vectors to achieve better fitness scores through the INFO operators. Upon reaching the maximum number of iterations, INFO returns the vector of the best individual, which contains the optimized bias and weight values. These values are used to establish the ELM model, which is then tested on the unseen testing dataset. The model's performance is evaluated using key metrics specified in the equations. The ELM-INFO pseudocode is described by Algorithm 1.

A. Data Preparation

In this research, we acquired a quarterly dataset encompassing diverse elements influencing CO_2 emission in Japan (https://data.worldbank.org/country/Japan). This dataset is compiled from real-time measurements of these contributing factors, spanning from the year 1980 to 2021. The dataset comprise of 168 data points, during the experimental phase, all input features underwent normalization using the min–max normalization process specified in Equation (24), ensuring values fall within the [-1, 1] range. Each model employed an 80% training dataset and a 20% test set partition. This division is consistent across all models.

$$x' = \left(\frac{x - \min_a}{\max_a - \min_a}\right) * 2 - 1 \qquad (24)$$

x represents the initial value, x' signifies the scaled value, $\max_a$ denotes the maximum value and $\min_a$ stands for the minimum value of feature a. Table VII displays the initial 5 data points following normalization. In Table VIII, the significance and meaning of each feature is elaborated. The input features for the training models encompass EG, NR, TI, and TR, with CO_2 serving as the output or predicted variable. The correlation score between each feature in the dataset is given in Figure 6, with a heatmap ranging from -1 to 1. Values near to 1 indicate a strong positive correlation. The plot in Figure 7 visually represents the relationship between each input variable and the dependent variable. It is noticeable that EG, TI, and TR exhibit a more linearly separable relationship with CO2.

B. ELM-INFO Complexity

1. Matrix Multiplication and Inversion in ELM (Object Function): $O(n \cdot h^2 + h^3)$ for training and testing the ELM, considering the hidden layer's output calculation and the Moore-Penrose pseudoinverse.

2. Fitness Calculation for Each Individual: Calculating the fitness of each individual via the evaluation of the objective function, the complexity is $O\big(p \cdot (n \cdot h^2 + h^3)\big)$ per iteration.

3. Population Updates and Vector Operations: These operations include INFO vector operations (local search, vector combination, and Update rule) and comparisons, typically $O(p)$ per iteration.

4. Overall Complexity: The loop across iteration multiplies the per-iteration complexity by the number of generations, resulting in $O\big(i \cdot p \cdot (n \cdot h^2 + h^3)\big)$

Thus, the detailed Big O notation for the complexity of integrating INFO with ELM, considering the iteration, population, samples, features, and hidden neurons, is $O(i \cdot p \cdot (n \cdot h^2 + h^3))$. This notation reflects the impact of algorithmic parameters and model architecture on computational complexity. Where h^2 and h^3 relate to operations involving the hidden layer neurons in the Extreme Learning Machine (ELM). i denote the number of iterations in the optimization process. p Denotes the population size of the optimization algorithm. n refers to the number of samples in the dataset. In ELM, it affects the complexity of operations like computing the output of the hidden layer, which involves all samples.

V. ELM-Model Optimization Evaluation

All of the models in this work use Root Square Root Error (RMSE) as their optimization function. To be clear, the comparison model was built and evaluated using the model parameters that correspond to the lowest RMSE values between the predicted data and the actual data across the training period, with respect to the ELM model, the optimization algorithms optimized the weight and bias of the neurons of the ELM model, with upper and lower bounds of -1 and 1 of the weight and bias, respectively. The settings for each of the three optimization techniques are shown in Table I.

TABLE VII
DATASET SAMPLE

S/N	Year	CO2	EG	NR	TI	TR
1	1980Q1	2.090601	0.000000	1.000000	0.000000	0.558465
2	1980Q2	2.086372	0.016164	0.953203	0.163404	0.567669
3	1980Q3	2.081424	0.031345	0.910094	0.285075	0.575037
4	1980Q4	2.075747	0.045571	0.872940	0.376972	0.580606
5	1981Q1	2.069328	0.058866	0.844290	0.446015	0.584406

TABLE VIII
FEATURE DESCRIPTION

Symbol	Variables	Measurement
EG	Economic Growth	Per Capita 2015 US$ Constant
NR	Natural resources	% of GDP
TI	Technological Innovation	Addition of patent resident and nonresident
CO_2	Carbon emissions	Per Capita
TR	Trade Openess	% of GDP

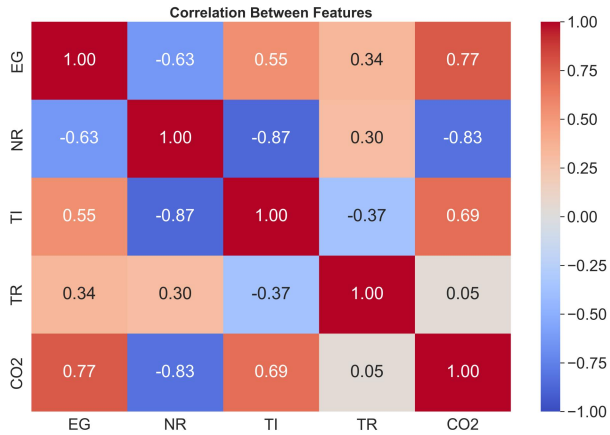

FIGURE 6. **Data Correlation HeatMap**

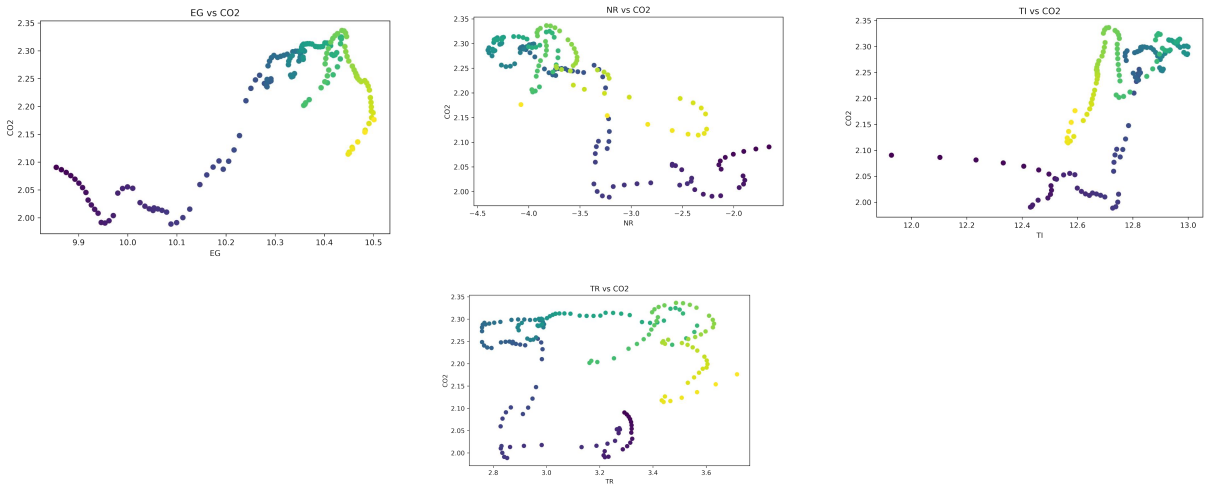

FIGURE 7. **Input Features vs Output Feature**

A. Model Evaluation

To evaluate the efficiency of the five models in this study, we used five statistical metrics: mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), coefficient of determination (R2), and root mean squared error (RMSE). The models' overall performance is measured using R2, RMSE, MAE, MAPE, and MSE.

$$R^2 = \frac{\left[\sum_{i=1}^{n}(Ep_o-\bar{E}p_o)(Ep_s-\bar{E}p_s)\right]^2}{\sum_{i=1}^{n}(Ep_o-Ep_o)^2\sum_{i=1}^{n}(Ep_s-Ep_s)^2} \quad (25)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Ep_o - Ep_s)^2} \quad (26)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Ep_o - Ep_s| \quad (27)$$

$$MAPE = \frac{1}{n}\frac{\sum_{i=1}^{n}|Ep_o-Ep_s|}{\bar{E}p_o} \quad (28)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Ep_o - Ep_s)^2 \quad (29)$$

When it comes to CO_2 prediction, models error is considered to be less when the values of RMSE, MAE, MAPE, and MSE are closer to zero. A R2 value close to 1 indicates better model accuracy, implying less error in real-world use. In equations (25-29), the terms for the observed, simulated, and mean observed CO_2 levels are Ep_o, Ep_s, and $\bar{E}p_o$. The evaluation measures include the RMSE, MAE, and MAPE.

VI. Result and Discussion

In Table IX and Table X, the performance metrics for different models, specifically the R2, RMSE, MSE, MAE, and MAPE are presented for both test and train datasets.

The models under consideration are variants of the Extreme Learning Machine (ELM) enhanced with various optimization algorithms. The integration of metaheuristic algorithms, such GWO, SSA, SCA, and INFO, to optimize the weights and biases of the ELM has yielded promising results in CO_2 prediction. We evaluate the performance of different models in predicting CO_2. Notably, the proposed ELM-INFO consistently demonstrates the most favorable outcomes among the models, showcasing its superiority in both the training and testing phases.

TABLE IX
RESULTS OF EVALUATION METRICS OF MODEL ON TRAINING DATASET

Models	R2	RMSE	MSE	MAE	MAPE
ELM-GWO	0.952339782	0.022051792	0.000486282	0.016819195	0.008582227
ELM-INFO	0.963078653	0.019409066	0.000376712	0.013621682	0.006227307
ELM-SCA	0.933790769	0.025991129	0.000675539	0.019702901	0.010292153
ELM-SSA	0.944950219	0.023699745	0.000561678	0.016889532	0.008007714
ELM	0.905737975	0.031012 32	0.000961764	0.025458567	0.013968315

In the training phase, as seen in Table IX, the ELM-GWO model exhibits a substantial improvement with an R2 value of 0.9523, showcasing the efficacy of the GWO in refining model parameters. The low values in RMSE (0.0221), MSE (0.0005), MAE (0.0168), and MAPE (0.0086) further validate the success of GWO in minimizing prediction errors. Notably, the proposed ELM-INFO model achieves impressive metrics, particularly with the highest R2 (0.9631) and lowest values in RMSE (0.0194), MSE (0.0003), MAE (0.0136), and MAPE (0.0062) among the models, highlighting the superiority of the INFO algorithm in enhancing the learning process of ELM.

In Table X, which demonstrates the results from the testing phase, the models maintain their superior performance, validating their generalization capabilities. ELM-INFO continues to obtain better precision with the highest R2 (0.9742), emphasizing its robustness in predicting CO2 on unseen data. ELM-GWO follows closely with a notable R2 of 0.9588, showcasing the algorithm's success in maintaining predictive accuracy beyond the training dataset ELM-SSA and ELM-SCA also perform admirably in the testing phase, underscoring the adaptability of these algorithms in optimizing the ELM for diverse datasets. While the base ELM model performs competently, particularly in the testing phase with an R2 of 0.9055, the incorporation of metaheuristic algorithms consistently outperforms the baseline. This reinforces the idea that leveraging sophisticated optimization techniques significantly enhances the ELM's predictive capabilities. of each model during the optimization process.

The fitness function, as expressed in equation 26, measures the ability of the model to fit the training data with less error. ELM-INFO exhibits the fastest convergence rate, indicating that it reaches optimal performance more efficiently than the
other models. This rapid convergence is attributed to the INFO algorithm's capacity to efficiently identify the best parameters for ELM for the training data, thereby accelerating the optimization process and leading to better model performance.

TABLE X
RESULTS OF EVALUATION METRICS OF MODEL ON TESTING DATASET

Models	R2	RMSE	MSE	MAE	MAPE
ELM-GWO	0.958780452	0.024512027	0.000600839	0.018633236	0.007544524
ELM-INFO	0.974244166	0.019376057	0.000375432	0.013637 76	0.006097756
ELM-SCA	0.945884136	0.028086001	0.000788823	0.022400228	0.008864959
ELM-SSA	0.958883033	0.024481507	0.000599344	0.017449 19	0.007583483
ELM	0.905530516	0.037108498	0.001377041	0.030311853	0.011542603

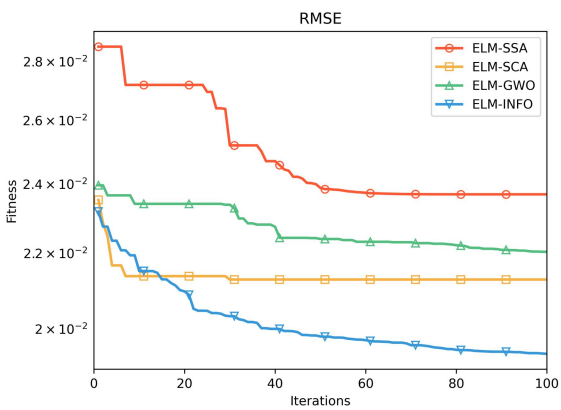

FIGURE 8. Convergence of Curve of Models on Fitness Function

In order to maintain fairness and reduce the potential for bias in our evaluation, we have used 10-fold cross-validation.

TABLE XI: RESULTS OF EVALUATION METRICS BY PREDICTION MODELS ON 10 FOLD CROSS VALIDATION ON TRAIN DATASET

Models	R2	RMSE	MSE	MAE	MAPE
ELM-GWO	0.9575 13956	0.021764 385	0.000477 113	0.016149 706	0.007293 833
ELM-INFO	0.9672 96012	0.019102 639	0.000365 262	0.014049 223	0.006327 781
ELM-SCA	0.9415 78115	0.025546 874	0.000653 391	0.019058 601	0.008698 642
ELM-SSA	0.9631 09812	0.020310 663	0.000412 736	0.014974 092	0.006764 987
ELM	0.9227 76666	0.029341 083	0.000863 160	0.023096 790	0.011144 280

TABLE XII: RESULTS OF EVALUATION METRICS BY PREDICTION MODELS ON 10 FOLD CROSS VALIDATION TEST DATASET

Tables XI and XII provided compare the performance of different ELM-based models using evaluation metrics from 10-fold cross-validation on train and test datasets. The ELM-INFO model outperforms others in nearly all metrics, indicating its superior predictive accuracy and generalization capability. It achieves the highest R2 values and the lowest error rates across RMSE, MSE, MAE, and MAPE on both train and test datasets, reflecting consistent and reliable performance. ELM-GWO and ELM-SSA also show strong results, significantly better than the standalone ELM. However, ELM-SCA and the standalone ELM lag in performance, as seen by their lower R2 values and higher error measures. The consistent success of ELM-INFO across different metrics and datasets underscores the effectiveness of integrating INFO optimization with ELM for predictive tasks. The ELM-INFO model, with an R2 of 0.967 on the training dataset and 0.966 on the test dataset, showcases its superior fit compared to the other models. Its predictive accuracy is further highlighted by the lowest error metrics, with an RMSE of 0.0191 and 0.0189, MSE of approximately 0.000365 and 0.000363, MAE of 0.0140, and MAPE of 0.0063 on train and test datasets, respectively. ELM-GWO follows closely, while ELM-SCA and ELM-SSA trail behind yet still outperform the standalone ELM, which has the lowest R2 and highest error rates across both datasets, indicating a less accurate model.

Models	R2	RMSE	MSE	MAE	MAPE
ELM-GWO	0.952847151	0.021700895	0.000480003	0.016177996	0.007287746
ELM-INFO	0.966060677	0.018914425	0.000362609	0.014058011	0.006324416
ELM-SCA	0.935078757	0.025408704	0.000672191	0.019223234	0.008616953
ELM-SSA	0.960572212	0.020050024	0.000413190	0.015040702	0.006735040
ELM	0.888964887	0.032153275	0.001118387	0.024519029	0.010480358

Figures 9 and 10 are scatterplots of the predicted and actual CO2 for the training testing dataset. Each point on the plot represents a single data point in the testing dataset, with the x-axis showing the predicted CO2 and the y-axis showing the actual CO2 level. The ideal scenario would be for all of the data points to fall on the diagonal line, which would indicate that the model is perfectly predicting the actual CO2 levels. However, in practice, there will always be some deviation between the predicted and actual values.

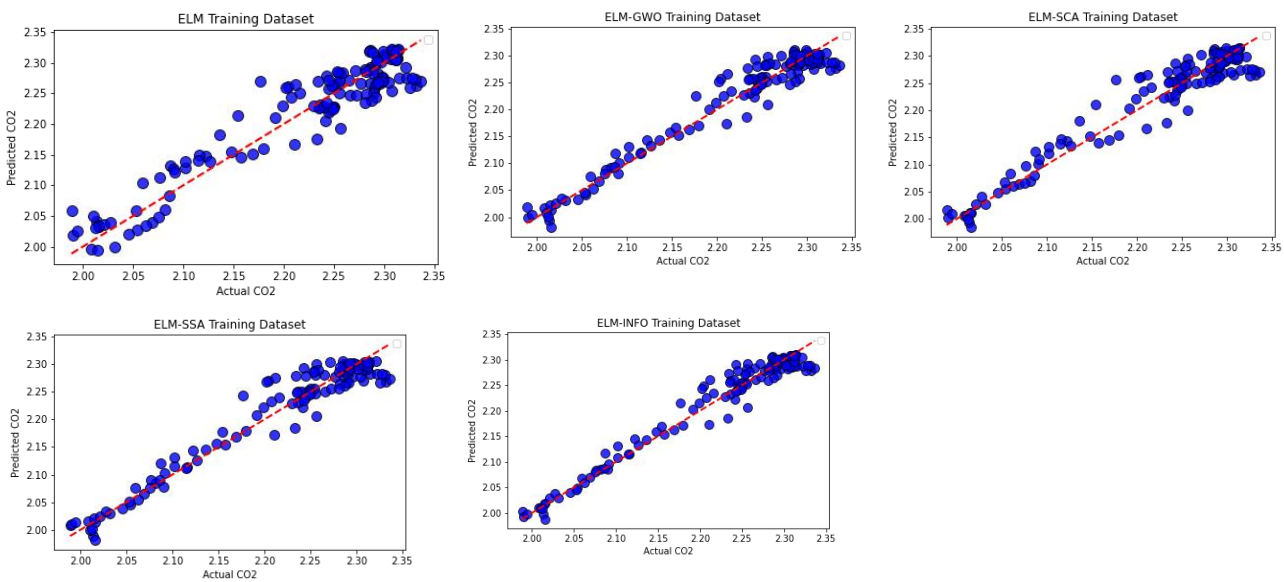

FIGURE 9. Predicted CO2 versus Actual CO2 of Prediction Models on Training Dataset

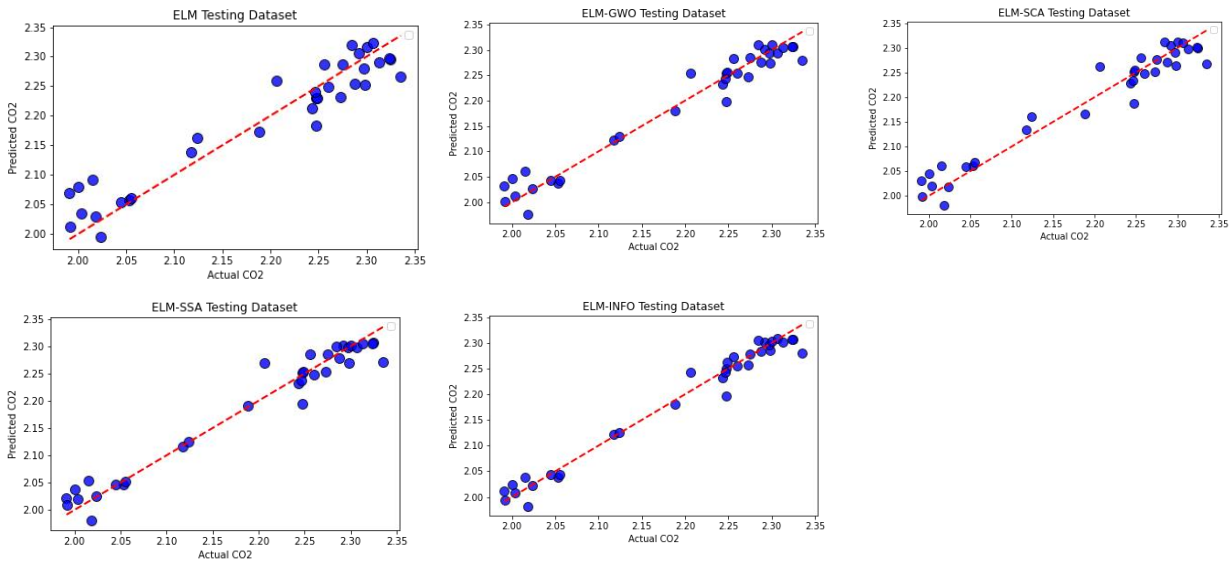

FIGURE 10. **Predicted CO2 versus Actual CO2 of Prediction Models on Testing Dataset**

The scatterplots in Figures 9 and 10 prove that the ELM-INFO model produced more accurate outcomes predicting the actual CO_2. Compared to ELM, ELM-SSA, and ELM-SCA, most of the points are grouped along the diagonal line, with very few exceptions. This suggests that for most of the data points in the training and testing dataset, the model can predict the CO_2 with reasonable accuracy. In comparison to the other models, the scatter plots demonstrate that ELM-INFO can properly learn the connection between the input characteristics and CO_2 and can produce precise estimations for both the training and testing datasets. The training and testing datasets' respective R2 values of 0.9631 and 0.9742 show a very high correlation between the expected and actual CO_2 levels. With a very tiny average variation between the anticipated and actual CO_2 emission, the model is producing highly reliable predictions, as evidenced by the testing dataset's RMSE and MAPE values of 0.0194 and 0.0062, respectively.

C. Significance of Study:

Economic variables, including economic growth (EG), trade openness (TR), and technological innovation (TI), Natural Resources (NR) are carefully selected as features of dataset retrieved from worldbank and integrated into the model as input features. These variables represent key dimensions of economic activity and development, each contributing to the overall dynamics of CO2 emissions. Economic growth (EG) serves as a primary indicator of economic activity and prosperity, capturing the rate of increase in a country's gross domestic product (GDP) over time. The integration of EG into the model allows us to assess the impact of economic expansion on CO2 emissions, reflecting the energy consumption and industrial output associated with economic growth. NR refer to the availability or utilization of natural resources within a country or region. NR also represent factors such as the abundance of raw materials and energy resources (e.g., fossil fuels), The inclusion of NR in the models input data, gives into how the availability or exploitation of natural resources influences CO2 emissions. TI refers to advancements or changes in technology that affect economic productivity, efficiency, and industrial processes. TI also represent factors such as investments in research and development, the adoption of cleaner or more efficient technologies, or shifts in industrial practices. Including TI in the models in input data enables this research to examine how technological innovation influences CO2 emissions. TR involves international trade relationships and patterns of a country or region. TR encompass variables related to imports, exports, trade agreements, and globalization processes. By incorporating TR into the analysis, this research can explore how international trade affects CO2 emissions. For example, countries with strong export-oriented industries may experience economic growth driven by international demand but may also face challenges related to environmental degradation associated with increased production and trade..

Fig. 10 shows the permutation importance of the features for predicting CO_2 using the ELM-INFO. This means that the importance of each feature is measured by the degree to which the model's performance diminishes when that feature is permuted; the higher the permutation importance, the more important the feature is for predicting CO2. It is important to note that the permutation importance of a feature is not the same as its correlation with CO2.

Correlation only measures how strongly two variables are related, while permutation importance measures how significant a feature is for predicting CO2. The permutation importance values in Figure 11 depend on the evaluation metric used to evaluate model performance, with higher values indicating greater importance. We used the MSE metric to evaluate the importance of each feature, and a positive permutation importance score suggests that the feature is important, as permuting its values leads to an increase in mean squared error. The most important feature, Economic Growth (EG), has a permutation importance of 0.199, meaning that the model's error (MSE) increased by 19% when the EG feature is permuted. The other features, such as Technological Innovation (TI) and Trade openness (TR), have a significant impact by increasing the model error by 8.1% and 7.6%, respectively.

The key observations from this study are:

According to the dataset, economic growth is the factor that is most likely to contribute to CO2 emissions, as indicated by its high feature relevance value. This is because rise in economic growth is linked to rise in energy and industrial activity levels, both of which boost CO2 emissions [42]. With a substantial permutation relevance, technological innovation appears to be an important contributor to CO2 emissions in the dataset as well. This is due to the fact that technological advancement can result in the creation of both dirtier and cleaner technology.

For instance, although new methods of extracting fossil fuels might raise CO2 emissions, renewable energy technology can lower CO2 emissions [43], [44].

Additionally, trade openness has a moderate value, indicating that it has a significant role on CO2 emissions. It is crucial to remember that trade openness can affect CO2 emissions in both good and negative ways. Trade openness, for instance, can result in the development of greener technologies that lower CO2 emissions in poorer nations. Open trade, however, may also result in the relocation of companies that produce pollution to emerging nations, raising CO2 emissions [45], [46].

Because natural resources are used less frequently in the data, it obtained the least significant factor influencing CO2 emissions in this dataset, as indicated by their lowest permutation relevance. Despite the fact that most nations possessing natural resources, such coal and oil, are more inclined to use them to produce energy, which increases CO2 emissions [47].

The crucial elements influencing CO2 emissions in the dataset have been revealed by the ELM-INFO's permutation significance analysis, offering a plethora of priceless information to stakeholders and policymakers entrusted with creating effective environmental legislation. These results do more than just indicate the relative significance of each component; they also function as an outline for focused data collecting and analysis. The ELM-INFO model helps decision-makers pinpoint and identify the major causes of CO2 emissions, facilitating more effective and efficient policy actions. In the end, the ELM-INFO is a useful tool for developing comprehensive plans and evidence-based policies targeted at addressing CO2 emissions and reducing their negative consequences.

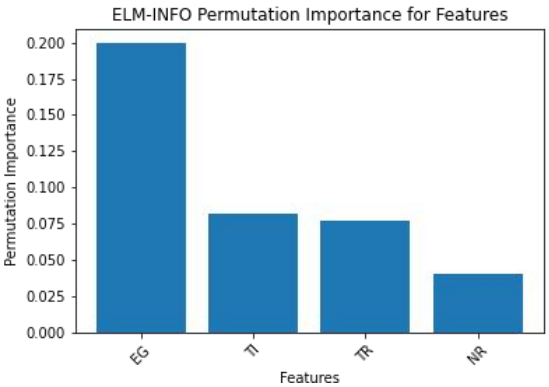

FIGURE 11. Input Features Permutation Importance

The model predictive capacity could be instrumental in informing and shaping environmental policies through the following step

1. Policy Targeting and Prioritization: The model helps identify the main sources and areas of CO2 emissions by analyzing economic growth, trade, and innovation impacts. This allows for focused policy action on the most impactful sectors.

2. Long-Term Planning and Adaptation: Forecasting CO2 emissions assists in planning and adapting for future environmental challenges, enabling the development of strategies to mitigate climate change effects and enhance resilience.

3. Policy Evaluation and Improvement: Continuous assessment and refinement of the model based on actual data enable the iterative enhancement of policy measures, ensuring their effectiveness and efficiency.

4. Stakeholder Engagement and Communication: Utilizing predictive insights as a communication tool aids in engaging stakeholders, increasing awareness of environmental issues, and garnering support for policy initiatives through effective visualization of CO2 impact scenarios.

VII. CONCLUSION

In summary, this study concentrated on the creation and assessment of a brand-new optimisation algorithm called INFO that was combined with the ELM for CO2 emissions in an effort to enhance the model's overall performance. This allowed for quick adaptation and learning from the dataset through the optimisation of the weight and bias. A thorough evaluation of the INFO algorithm's performance was conducted by comparing it to other well-known

algorithms. The thorough examination of INFO's performance on test functions for classical optimisation revealed several noteworthy strengths, especially when it came to obtaining optimum solutions for both unimodal and multi-modal functions. Convergence plots also showed how quickly INFO converged and how effective INFO will be at improving the weight and bias of ELM. Through a two-stage training and prediction procedure, the suggested ELM-INFO model applied to the prediction of CO2 emissions demonstrated its effectiveness. The prediction phase showed the model's capacity to estimate CO2 emissions using unobserved data, whereas the training phase entailed optimising ELM's weights and biases using INFO. Key performance indicators for the model, such as MSE, MAE, MAPE, R2, and RMSE, were used to assess its performance. With an outstanding R2 (0.9631) and lowest values in training for RMSE (0.0194), MSE (0.0003), MAE (0.0136), and MAPE (0.0062), the findings continuously demonstrated the superiority of the ELM-INFO model. Its resilience in forecasting CO2 emissions for training and testing was demonstrated by its strong performance in testing, with an R2 of 0.9742 and the lowest values in RMSE (0.01937), MSE (0.00037), MAE (0.0136), and MAPE (0.0060). Using ELM-INFO, the study also carried out a permutation importance analysis to ascertain the relative importance of important variables influencing CO2 emissions in the dataset. The largest impact was clearly economic growth, which raised the model's MSE by 19%. Trade openness and technological advancement came very close behind, adding 7.6% and 8.1%, respectively, to the model's MSE increases. With only a modest 4.1% contribution, natural resources in the dataset showed the least amount of effect on CO2 emissions. The study provides focused treatments and policy options for minimising the consequences of climate change by identifying and prioritising important variables impacting CO2 emissions. With its improved optimisation features, the ELM-INFO model is a potent instrument for developing evidence-based policies that seek to achieve sustainability and resilience in the face of environmental difficulties.

Even though this study produced encouraging results, there are a few limitations to be aware of. First off, the reliability and the accuracy of the dataset used for testing and training have a significant impact on the model's performance. Thus, getting more and more varied datasets may improve the generalizability and robustness of the model. Furthermore, when the suggested approach is used to various geographic locations or diverse datasets with distinctive features, it may show particular limitations. In order to overcome these constraints, contextual variances must be carefully taken into account, and other variables that could affect CO2 emissions must be included. Subsequent developments of ELM-INFO will concentrate on enhancing the model's precision and suitability through the integration of more and more diverse datasets. The model will undergo additional validations in various temporal and geographical settings to increase its flexibility and dependability in real-world situations. A thorough grasp of the synergistic impacts of various optimisation algorithms and machine learning approaches on enhancing prediction accuracy may also be obtained by investigating the possible integration of these methods. Furthermore, examining the time-varying dynamics of CO2 emissions and their correlation with changing socio-economic variables may provide insightful information for more accurate and prospective forecasts. More work will be put into overcoming these constraints and broadening the study's focus in order to support more potent mitigation techniques for climate change.

REFERENCES

[1] T. Wheeler and J. von Braun, 'Climate Change Impacts on Global Food Security', *Science*, vol. 341, no. 6145, pp. 508–513, Aug. 2013, doi: 10.1126/science.1239402.

[2] G. Kolsuz and A. E. Yeldan, 'Economics of climate change and green employment: A general equilibrium investigation for Turkey', *Renew. Sustain. Energy Rev.*, vol. 70, pp. 1240–1250, Apr. 2017, doi: 10.1016/j.rser.2016.12.025.

[3] M. A. Benevolenza and L. DeRigne, 'The impact of climate change and natural disasters on vulnerable populations: A systematic review of literature', *J. Hum. Behav. Soc. Environ.*, vol. 29, no. 2, pp. 266–281, Feb. 2019, doi: 10.1080/10911359.2018.1527739.

[4] A. Omer, N. A. Elagib, M. Zhuguo, F. Saleem, and A. Mohammed, 'Water scarcity in the Yellow River Basin under future climate change and human activities', *Sci. Total Environ.*, vol. 749, p. 141446, Dec. 2020, doi: 10.1016/j.scitotenv.2020.141446.

[5] T. Arcomano, I. Szunyogh, J. Pathak, A. Wikner, B. R. Hunt, and E. Ott, 'A Machine Learning-Based Global Atmospheric Forecast Model', *Geophys. Res. Lett.*, vol. 47, no. 9, p. e2020GL087776, 2020, doi: 10.1029/2020GL087776.

[6] K. O. Yoro and M. O. Daramola, 'Chapter 1 - CO2 emission sources, greenhouse gases, and the global warming effect', in *Advances in Carbon Capture*, M. R. Rahimpour, M. Farsi, and M. A. Makarem, Eds., Woodhead Publishing, 2020, pp. 3–28. doi: 10.1016/B978-0-12-819657-1.00001-3.

[7] F. Zennaro *et al.*, 'Exploring machine learning potential for climate change risk assessment', *Earth-Sci. Rev.*, vol. 220, p. 103752, Sep. 2021, doi: 10.1016/j.earscirev.2021.103752.

[8] X. Wang, Y. Bouzembrak, A. O. Lansink, and H. J. van der Fels-Klerx, 'Application of machine learning to the monitoring and prediction of food safety: A review', *Compr. Rev. Food Sci. Food Saf.*, vol. 21, no. 1, pp. 416–434, 2022, doi: 10.1111/1541-4337.12868.

[9] H. Mohammad-Rahimi, M. Nadimi, A. Ghalyanchi-Langeroudi, M. Taheri, and S. Ghafouri-Fard, 'Application of Machine Learning in Diagnosis of COVID-19 Through X-Ray and CT Images: A Scoping Review', *Front. Cardiovasc. Med.*, vol. 8, 2021, Accessed: Dec. 11, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcvm.2021.638011

[10] R. Bunker and T. Susnjak, 'The Application of Machine Learning Techniques for Predicting Match Results in Team Sport: A Review', *J. Artif. Intell. Res.*, vol. 73, pp. 1285–1322, Apr. 2022, doi: 10.1613/jair.1.13509.

[11] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, and D. J. Inman, 'A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications', *Mech. Syst. Signal Process.*, vol. 147, p. 107077, Jan. 2021, doi: 10.1016/j.ymssp.2020.107077.

[12] B. E. Odigwe, A. B. Rajeoni, C. I. Odigwe, F. G. Spinale, and H. Valafar, 'Application of machine learning for patient response prediction to cardiac resynchronization therapy', in *Proceedings of the 13th ACM International Conference on Bioinformatics,*

Computational Biology and Health Informatics, in BCB '22. New York, NY, USA: Association for Computing Machinery, Aug. 2022, pp. 1–4. doi: 10.1145/3535508.3545513.

[13] N. Kapilan, R. p. Reddy, and P. Vidhya, 'Recent Applications of Machine Learning in Solar Energy Prediction', in *Artificial Intelligence for Renewable Energy and Climate Change*, John Wiley & Sons, Ltd, 2022, pp. 33–52. doi: 10.1002/9781119771524.ch2.

[14] Y. Amethiya, P. Pipariya, S. Patel, and M. Shah, 'Comparative analysis of breast cancer detection using machine learning and biosensors', *Intell. Med.*, vol. 2, no. 2, pp. 69–81, May 2022, doi: 10.1016/j.imed.2021.08.004.

[15] B. Bochenek and Z. Ustrnul, 'Machine Learning in Weather Prediction and Climate Analyses—Applications and Perspectives', *Atmosphere*, vol. 13, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/atmos13020180.

[16] A. Saini and A. Sharma, 'Predicting the Unpredictable: An Application of Machine Learning Algorithms in Indian Stock Market', *Ann. Data Sci.*, vol. 9, no. 4, pp. 791–799, Aug. 2022, doi: 10.1007/s40745-019-00230-7.

[17] H. Rezapouraghdam, A. Akhshik, and H. Ramkissoon, 'Application of machine learning to predict visitors' green behavior in marine protected areas: evidence from Cyprus', *J. Sustain. Tour.*, vol. 31, no. 11, pp. 2479–2505, Nov. 2023, doi: 10.1080/09669582.2021.1887878.

[18] H. Guo, S. Wu, Y. Tian, J. Zhang, and H. Liu, 'Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review', *Bioresour. Technol.*, vol. 319, p. 124114, Jan. 2021, doi: 10.1016/j.biortech.2020.124114.

[19] R. Zhu, X. Hu, J. Hou, and X. Li, 'Application of machine learning techniques for predicting the consequences of construction accidents in China', *Process Saf. Environ. Prot.*, vol. 145, pp. 293–302, Jan. 2021, doi: 10.1016/j.psep.2020.08.006.

[20] A. Mahindru and A. L. Sangal, 'MLDroid—framework for Android malware detection using machine learning techniques', *Neural Comput. Appl.*, vol. 33, no. 10, pp. 5183–5240, May 2021, doi: 10.1007/s00521-020-05309-4.

[21] A. Hamrani, A. Akbarzadeh, and C. A. Madramootoo, 'Machine learning for predicting greenhouse gas emissions from agricultural soils', *Sci. Total Environ.*, vol. 741, p. 140338, Nov. 2020, doi: 10.1016/j.scitotenv.2020.140338.

[22] A. Mardani, H. Liao, M. Nilashi, M. Alrasheedi, and F. Cavallaro, 'A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques', *J. Clean. Prod.*, vol. 275, p. 122942, Dec. 2020, doi: 10.1016/j.jclepro.2020.122942.

[23] Q. Nguyen, I. Diaz-Rainey, and D. Kuruppuarachchi, 'Predicting corporate carbon footprints for climate finance risk analyses: A machine learning approach', *Energy Econ.*, vol. 95, p. 105129, Mar. 2021, doi: 10.1016/j.eneco.2021.105129.

[24] M. S. Bakay and Ü. Ağbulut, 'Electricity production based forecasting of greenhouse gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms', *J. Clean. Prod.*, vol. 285, p. 125324, Feb. 2021, doi: 10.1016/j.jclepro.2020.125324.

[25] R. Balakrishnan, V. Geetha, M. R. Kumar, and M.-F. Leung, 'Reduction in Residential Electricity Bill and Carbon Dioxide Emission through Renewable Energy Integration Using an Adaptive Feed-Forward Neural Network System and MPPT Technique', *Sustainability*, vol. 15, no. 19, Art. no. 19, Jan. 2023, doi: 10.3390/su151914088.

[26] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, 'Extreme learning machine: a new learning scheme of feedforward neural networks', in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, Jul. 2004, pp. 985–990 vol.2. doi: 10.1109/IJCNN.2004.1380068.

[27] K. Liao, Y. Wu, F. Miao, L. Li, and Y. Xue, 'Using a kernel extreme learning machine with grey wolf optimization to predict the displacement of step-like landslide', *Bull. Eng. Geol. Environ.*, vol. 79, no. 2, pp. 673–685, Mar. 2020, doi: 10.1007/s10064-019-01598-9.

[28] M. Shariati *et al.*, 'A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement', *Eng. Comput.*, vol. 38, no. 1, pp. 757–779, Feb. 2022, doi: 10.1007/s00366-020-01081-0.

[29] S. Boriratrit, C. Srithapon, P. Fuangfoo, and R. Chatthaworn, 'Metaheuristic Extreme Learning Machine for Improving Performance of Electric Energy Demand Forecasting', *Computers*, vol. 11, no. 5, Art. no. 5, May 2022, doi: 10.3390/computers11050066.

[30] J. Qiu, X. Yin, Y. Pan, X. Wang, and M. Zhang, 'Prediction of Uniaxial Compressive Strength in Rocks Based on Extreme Learning Machine Improved with Metaheuristic Algorithm', *Mathematics*, vol. 10, no. 19, Art. no. 19, Jan. 2022, doi: 10.3390/math10193490.

[31] D. Li, S. Li, S. Zhang, J. Sun, L. Wang, and K. Wang, 'Aging state prediction for supercapacitors based on heuristic kalman filter optimization extreme learning machine', *Energy*, vol. 250, p. 123773, Jul. 2022, doi: 10.1016/j.energy.2022.123773.

[32] J. Wang, S. Lu, S.-H. Wang, and Y.-D. Zhang, 'A review on extreme learning machine', *Multimed. Tools Appl.*, vol. 81, no. 29, pp. 41611–41660, Dec. 2022, doi: 10.1007/s11042-021-11007-7.

[33] S. Mirjalili, S. M. Mirjalili, and A. Lewis, 'Grey Wolf Optimizer', *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.

[34] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, 'Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems', *Adv. Eng. Softw.*, vol. 114, pp. 163–191, Dec. 2017, doi: 10.1016/j.advengsoft.2017.07.002.

[35] S. Mirjalili, 'SCA: A Sine Cosine Algorithm for solving optimization problems', *Knowl.-Based Syst.*, vol. 96, pp. 120–133, Mar. 2016, doi: 10.1016/j.knosys.2015.12.022.

[36] M. Farhat, S. Kamel, A. M. Atallah, M. H. Hassan, and A. M. Agwa, 'ESMA-OPF: Enhanced Slime Mould Algorithm for Solving Optimal Power Flow Problem', *Sustainability*, vol. 14, no. 4, Art. no. 4, Jan. 2022, doi: 10.3390/su14042305.

[37] I. Ahmadianfar, A. A. Heidari, S. Noshadian, H. Chen, and A. H. Gandomi, 'INFO: An efficient optimization algorithm based on weighted mean of vectors', *Expert Syst. Appl.*, vol. 195, p. 116516, Jun. 2022, doi: 10.1016/j.eswa.2022.116516.

[38] S. García, A. Fernández, J. Luengo, and F. Herrera, 'Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power', *Inf. Sci.*, vol. 180, no. 10, pp. 2044–2064, May 2010, doi: 10.1016/j.ins.2009.12.010.

[39] O. R. Adegboye and E. Deniz Ülker, 'Hybrid artificial electric field employing cuckoo search algorithm with refraction learning for engineering optimization problems', *Sci. Rep.*, vol. 13, no. 1, Art. no. 1, Mar. 2023, doi: 10.1038/s41598-023-31081-1.

[40] S. García, D. Molina, M. Lozano, and F. Herrera, 'A study on the use of non-parametric tests for analyzing the evolutionary algorithms' behaviour: a case study on the CEC'2005 Special Session on Real Parameter Optimization', *J. Heuristics*, vol. 15, no. 6, p. 617, May 2008, doi: 10.1007/s10732-008-9080-4.

[41] J. Lu, X. Zhou, Y. Ma, M. Wang, J. Wan, and W. Wang, 'A Novel Artificial Bee Colony Algorithm with Division of Labor for Solving CEC 2019 100-Digit Challenge Benchmark Problems', in *2019 IEEE Congress on Evolutionary Computation (CEC)*, Jun. 2019, pp. 387–394. doi: 10.1109/CEC.2019.8790252.

[42] H. Liu, W.-K. Wong, P. The Cong, A. A. Nassani, M. Haffar, and A. Abu-Rumman, 'Linkage among Urbanization, energy Consumption, economic growth and carbon Emissions. Panel data analysis for China using ARDL model', *Fuel*, vol. 332, p. 126122, Jan. 2023, doi: 10.1016/j.fuel.2022.126122.

[43] A. Mostafaeipour, A. Bidokhti, M.-B. Fakhrzad, A. Sadegheih, and Y. Zare Mehrjerdi, 'A new model for the use of renewable electricity to reduce carbon dioxide emissions', *Energy*, vol. 238, p. 121602, Jan. 2022, doi: 10.1016/j.energy.2021.121602.

[44] T. S. Adebayo, S. Ullah, M. T. Kartal, K. Ali, U. K. Pata, and M. Ağa, 'Endorsing sustainable development in BRICS: The role of technological innovation, renewable energy consumption, and natural resources in limiting carbon emission', *Sci. Total Environ.*, vol. 859, p. 160181, Feb. 2023, doi: 10.1016/j.scitotenv.2022.160181.

[45] T. S. Adebayo, H. Rjoub, G. D. Akinsola, and S. D. Oladipupo, 'The asymmetric effects of renewable energy consumption and trade openness on carbon emissions in Sweden: new evidence from

quantile-on-quantile regression approach', *Environ. Sci. Pollut. Res.*, vol. 29, no. 2, pp. 1875–1886, Jan. 2022, doi: 10.1007/s11356-021-15706-4.

[46] T. S. Adebayo, G. D. Akinsola, D. Kirikkaleli, F. V. Bekun, S. Umarbeyli, and O. S. Osemeahon, 'Economic performance of Indonesia amidst CO2 emissions and agriculture: a time series analysis', *Environ. Sci. Pollut. Res.*, vol. 28, no. 35, pp. 47942–47956, Sep. 2021, doi: 10.1007/s11356-021-13992-6.

[47] Danish, M. A. Baloch, N. Mahmood, and J. W. Zhang, 'Effect of natural resources, renewable energy and economic development on CO2 emissions in BRICS countries', *Sci. Total Environ.*, vol. 678, pp. 632–638, Aug. 2019, doi: 10.1016/j.scitotenv.2019.05.028.

Credits authorship contribution statement:

Afi Kekeli Feda Methodology: Writing, Investigation, Editing. Oluwatayomi Rereloluwa Adegboye: Conceptualization, Formal analysis, Original Draft, Methodology. Ephraim Bonah Agyekum: Methodology, Resources, Review & Editing. Abdurrahman Shuaibu Hassan: Methodology, Writing – Review & Editing. Salah Kamel: Methodology, Writing – Review & Editing.

Data Availability

The data obtained through the experiments are published in the paper.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Afi Kekeli FEDA received Her B.Sc. in Maintenance and computer Network at Université de Lomé, Togo. She concluded Her M.Sc. degree in Management Information System from European University of Lefke. She is currently pursuing her Ph.D in Management Information System at European University of Lefke. Her research interests include data science and search optimization algorithms.

OLUWATAYOMI RERELOLUWA ADEGBOYE received His B.Sc. in Electrical and Electronic Engineering from Osun State University, and concluded His M.Sc. degree in Computer Engineering from European University of Lefke. He also received his Ph. D in Computer Engineering at European University of Lefke. His research interests include machine learning, text mining, data science and search optimization algorithms.

EPHRAIM BONAH AGYEKUM received his BSc. Applied Physics degree from the University for Development Studies in Ghana in 2014, an MSc. Degree in Nuclear Power Operations Installations from the National Research Tomsk Polytechnic University in Russia in 2018, and a Ph.D. in Renewable Energy Engineering from Ural Federal University in 2022. He is currently a Senior Research and Postdoctoral Fellow at the Ural Federal University. He has published over 100 peer-reviewed articles in reputable journals like, Journal of Cleaner Production, Utilities Policy, Energy Reports, Sustainable Energy Technologies and Assessments, Environmental Science and Pollution Research, Heliyon, Energy, Journal of Energy Storage among others. He also serves as a reviewer for various reputable journals. His research interests include energy systems analysis, hydrogen energy, energy policy, multi-criteria decision-making analysis and smart grids.

ABDURRAHMAN SHUAIBU H, PhD is a senior lecturer in the Department of Electrical Engineering at the School of Engineering and Applied Sciences at Kampala International University. He is passionate about academia, the industry, research, and development and has always strived to join and become part of a competitive program which meets long-term educational and research needs, to drive a scientific approach to engineering processes. He holds a PhD. In Electrical Electronics Engineering. His research interest include power system analysis and optimizations, Economics of power system, Smart grid, Energy and Nonconventional energy system.

SALAH KAMEL received the international Ph.D. degree from Jaen University, Spain (Main), and Aalborg University, Denmark (Host), in January 2014. He is currently an Associate Professor with the Department of Electrical Engineering, Aswan University. He is also the Leader of the Advanced Power Systems Research Laboratory (APSR Lab), Power Systems Research Group, Aswan, Egypt. His research interests include power system analysis and optimization, smart grid, and renewable energy systems.