CIVIL ENGINEERING JOURNAL

Article no. 25

THE CIVIL ENGINEERING JOURNAL 3-2025

APPLICATION OF THE OPTIMIZED REGRESSION TO VOLUME EXPANSION EVALUATION OF CEMENT PASTE WITH FLY ASH AND MGO EXPANSIVE ADDITIVE

Yishuo Wang, Ziyu Liu and Cheng Wang

Xi'an University of Science and Technology, Xi'an710600, Shaanxi, China; 22403050108@stu.xust.edu.cn

Received: 21.10.2024
Received in revised form: 13.02.2025
Accepted: 01.09.2025

ABSTRACT

The synergistic effects of fly ash (FA) and magnesium oxide (MgO) expansive addition (MEA) on the expansion behavior and mechanical qualities of cement paste are validated by studies that manipulate dose and curing temperature as factors. Few studies have been conducted on the use of machine learning (ML) techniques to predict the volume expansion (V_e) of cement paste when FA and MEA are present. Utilizing a set of data that contained 170 experimental results that were obtained from previous research, the purpose of this study was to develop and evaluate ML algorithms for the assessment of V_e. To achieve this, the Extra tree regression (ETR) was created. The technique uses FA, Portland cement (PC), MEA, and sample age (Age) as input parameters. The Red-Tailed Hawk Algorithm (RTHA) and the Electric Eel Foraging Algorithm (EEFA) establish the hyperparameters of ETR, which greatly affect its effectiveness. The variation percentages of the two models developed for these measures are a minimum of 13%; in some cases, this variation decreases by 53%, illustrating the ETR (R)'s predictive reliability and efficacy. For instance, for the RMSE index, ETR (R) achieved values of 0.0053 and 0.01 during the training and testing phases, which are about 28% and 14% lower than the corresponding values of ETR (E) at 0.0068 and 0.0114, correspondingly. The impact of this research allows practical applications: cement formulation optimization, structural integrity, and enhanced quality control in the building and construction industry. An example could be an exact prediction of volume expansion V_e by the model, which gives very valuable cost reductions by preventing cracks and material performance consistency in cement paste with FA and MgO additives MEA.

KEYWORDS

Cement Paste, Volume Expansion, Fly Ash; Machine Learning, Extra Tree Regression, Optimal models

INTRODUCTION

For large-scale concrete constructions, thermal stresses that can lead to fracturing are a considerable issue, necessitating the performance of temperature control strategies throughout the creation phase. In the case of hydraulic mass cement-based composite constructions, the primary cause of gaps is temperature-induced warping occurring shortly after the completion of creation [1, 2]. The initial fracturing phenomenon observed in cement-based composite primarily arises from the balance between resistance and stress throughout the hardening phase. As cement-based composite matures, its attributes evolve, leading to a gradual growth in resistance, which in turn results in the distribution of high temperatures [3, 4]. The simultaneous effects of hydration heat and

the cooling of the surroundings create a temperature gradient within the cement-based composite construction.

Various temperature control techniques, including the application of ice for $\mathrm{pre-cooling}$ aggregates, the installation of cooling water pipes within the construction, and the performance of surface heat retention techniques, have been employed to mitigate the occurrence of temperature-related gaps [5]. Nevertheless, they are not cost-effective, and certain approaches, like pipe cooling, are quite intricate and could significantly impact the construction actions timeline [6].

In both investigation and functional applications, the addition of a suitable amount of magnesium oxide, which has been calcined at elevated temperatures, can be incorporated into the dam's cement-based composite preparation procedure. This inclusion facilitates the expansion of the cement-based composite, effectively counteracting contraction distortion that may occur during a diminish in temperature [7, 8]. The incorporation of magnesium oxide as an enhancer factor in mass concrete, alongside routine auxiliary measures such as surface heat protection, can effectively manage self-induced volume distortion, thereby decreasing and stopping gaps in dam concrete. This approach has the potential to avert gaps in dams and minimize the reliance on conventional cement-based composite temperature management techniques, such as aggregate $\mathrm{pre-cooling}$, ice blending, and the use of embedded refrigerating water conduits [9].

The distribution strategy of magnesium oxide concrete is notably difficult [10]. The hydration procedures of both the cement and the magnesium oxide contribute to endogenous distortion. Additionally, the presence of FA within the cementing matter significantly affects the features of magnesium oxide cement-based composite [11].

The labor-intensive and expensive characteristics of laboratory studies, coupled with the possible negative impacts on the environment, have led to a growing interest in alternative approaches. These include non-linear and advanced regression techniques, tuned models [12], optimized models, and hybrid models [13–16]. It is generally crucial to mathematically model the endogenous volume distortion (distribution) and the corresponding distribution stress during the construction phase.

Numerous engineering actions in China have resulted in the collection of remarkable distortion data [17]. B. Zhu (2002, 2003) [18, 19] proposed the measured quantities for the endogenous volumetric distortion of micro-distribution concrete and founded a computation model of the endogenous volume distortion in light of these variations. According to the engineering test data and pertinent laboratory analyses, G. Yang et al. (2004) [20] proposed that the endogenous volumetric distortion features of magnesium oxide cement-based composite at a fixed temperature align with a $\mathrm{hyperbolic}$ model. G. Zhang et al. (2002, 2005) [21–23] conducted a thorough analysis utilizing imitation methods to examine the effects of the growth of cement-based composite's elastic modulus, the degree of creep over time, and temperature on the distribution of MgO. Liu et al. (2006) [24] developed a numerical model to describe the correlations among the endogenous volume distortion of magnesium oxide cement-based composite, the content of magnesium oxide, the volume of FA, and the $\mathrm{temperature}$. P. Xu and colleagues introduced the equivalent age method to assess the un-planned volume distortion of magnesium oxide cement-based composite in their 2008 study [25]. Moreover, $\mathrm{sub-stitute}$ models have been suggested, including the arctangent curve model developed by C. Chen and associates in 2008 [26].

According to the references [27–29], it is evident that model ETR was developed in 2014, making it a relatively newer model compared to older models like ANN and SVR. In contrast, the optimization algorithms $RTHA$ and $EEFA$ are very recent developments and have not yet been applied in civil engineering and concrete technology. A comprehensive review of the technical literature shows that the optimized models of $ETR + RTHA$ and $ETR + EEFA$ are less used in various fields of civil engineering. These models will be predicted and optimized in future research.

The estimation of cement paste volume expansion (V_e) FA and MEA using ML techniques have not received much attention. The objective of this study was to develop and evaluate ML techniques for assessing V_e via a dataset of 170 test results derived from previously published research. ETR algorithm has been devised to achieve this objective. The approach utilizes input

parameters like PC, FA, MEA, and sample age (Age) to produce the layouts. The effectiveness of ETR is considerably influenced by its hyperparameters, which may be enhanced through optimization techniques. The present work used advanced optimizers, namely the EEFA and the RTHA, to achieve this particular target. The efficacy of the established regression analysis is assessed by several metrics, taking into account the values of observed and anticipated V_e. Engineers may reduce the danger of cracking and other expansion-related problems by fine-tuning mixes to obtain desired qualities by forecasting the volume expansion behavior of cement paste. It makes use of sophisticated optimizers such as RTHA and EEFA, which could save time and money in experimental research.

Overall, it was reviewed that recent studies demonstrate the effectiveness of incorporating magnesium oxide in cement-based composites to reduce temperature-induced warping and crack reduction in large-scale concrete structures. Advanced modeling techniques, including the ETR algorithm and optimization methods such as EEFA and RTHA, are used to predict volume expansion in cement pastes and increase the understanding of endogenous volume distortion. The goal of integrating these models is to improve construction practices by minimizing the risk of cracking and optimizing material properties, thereby reducing reliance on traditional temperature control methods.

Previous studies indicated that a limited application of machine learning techniques was employed to forecast the volume expansion parameter of cement paste (V_e). Additionally, the predominant reliance on laboratory experiments proved to be both expensive and time-consuming. Consequently, this research aims to implement advanced machine learning methodologies that offer predictions that are not only more cost-effective but also more precise and efficient.

RESEARCH METHODOLOGY

Dataset Pre-Processing

Different quantitative studies have been conducted to estimate the growth in volume (V_e) of cement paste, and the result is that the precise and correct selection of input parameters when using ML approaches is one of the main agents used to achieve perfect estimation. The initial step for using an ML approach requires the crucial selection of input and outcome parameters and data sets, which directly affects the precision and practicality of soft computing approaches. Therefore, data must be collected from reliable resources to have a positive impact on estimation precision [30]. Following the set criteria for feature selection, a whole of 170 data were collected, which were divided into two parts, training and testing, with quantities of 70% and 30%, respectively. This data-splitting strategy has been adopted using the existing literature [31–34]. The input parameters include four items: FA, PC, extensive addition of MgO (MEA) and example age (age), and the outcome parameter is the volume growth component (V_e) of the cement paste. Regarding the volumetric growth (V_e) of cement paste, it can be said that it is an important agent in understanding the behavior and properties of cement base matters. This parameter refers to the increase in volume of cement paste due to hydration and other chemical reactions that occur when water is added to cement. A higher water/cement proportion can lead to increased volumetric growth, but may negatively affect resistance. Also, it is very essential to predict the performance of concrete and cement-based materials during curing and long-term use. Also, in Table 1, by using commonly used statistical indicators including Minimum, Maximum, Average, Standard Deviation, Skewness and Kurtosis, more features of the related dataset have been divided into two parts. The standard deviation shows the changeability of the data, the skewness shows the disproportion of the data broadcast, for example, the value 0 indicates a symmetrical broadcast, quantities greater than 0 indicate a right-skewed broadcast, and quantities less than 0 indicate a left-skewed broadcast. kurtosis also indicates the dispersion of the broadcast, i.e. the number 0 is normal, the positive value of the broadcast has more outliers, and the negative values of the broadcast have less outliers. For example, in this table, in the training part, the FA ranges from 0 to 40% with an average of 21.68%

and a slight broadcast to the left. Also, in the test part, the parameter V_e declined by an average of 0.046% compared to the value of the same parameter in the training part and has a positive skew, which is similar to the broadcast pattern of the same parameter in the training part. Consequently, this table shows the general insight of the input and outcome parameters for the two mentioned levels and shows the general changes in the amount of data broadcast and their central tendency. This method is very useful for understanding the performance of the model and recognizing its attributes.

In order to understand the data set used deeply, graphic displays are used to understand the attributes of the data. For example, in Figure 1, an example of this graphic display is given to compare the outcome parameter against four input parameters. This chart is a combination of a violin diagram and a box chart that depicts the broadcast of two parameters in front of each other for comparison. For example, part a of Figure 1 shows the two parameters PC and V_e. The left y-axis displays the numbers of the PC parameter, ranging from 55% to 100%. The y-axis on the right side represents V_e, ranging from 0 to 0.18%. The violin chart for the cyan PC variable illustrates the broadcast of quantities in percentage. Its width also shows various density numbers of data dots. The violin chart for the orange parameter V_e also shows the broadcast of the numbers of this parameter. Each violin plot overlaps with the box plot, which forms the line inside the box (middle). This chart effectively shows the broadcast between two parameters and displays their central tendencies, variability, and potential outliers. It is especially useful for comparing the acting of two parameters in relation to each other and also for identifying patterns or trends in the data.

Tab. 1 - Statistics for each variable included in the dataset

Variable	Type	Amount	Unit	Statistics					
				$Min.*$	$Max.$	$Avg.$	$St.D.$	$Skew.$	$Kurt.$
Training phase	Input	PC	%	60	100	78.31933	16.31205	0.155304	-1.49253
	Input	FA	%	0	40	21.68067	16.31205	-0.1553	-1.49253
	Input	MEA	%	0	10	5.042017	3.969199	-0.01496	-1.42207
	Input	Age	$Days$	0.829	91.25	36.93034	30.48136	0.355487	-1.32172
	Target	V_e	%	0	0.166	0.050824	0.044724	1.049979	0.170372
Testing phase	Input	PC	%	60	100	82.7451	15.85207	-0.24905	-1.38556
	Input	FA	%	0	40	17.2549	15.85207	0.249051	-1.38556
	Input	MEA	%	0	10	5	4.315953	0	-1.70516
	Input	Age	$Days$	1.029	91.044	29.69829	29.47453	0.665697	-0.94112
	Target	V_e	%	0	0.166	0.045608	0.049919	1.048993	-0.09609

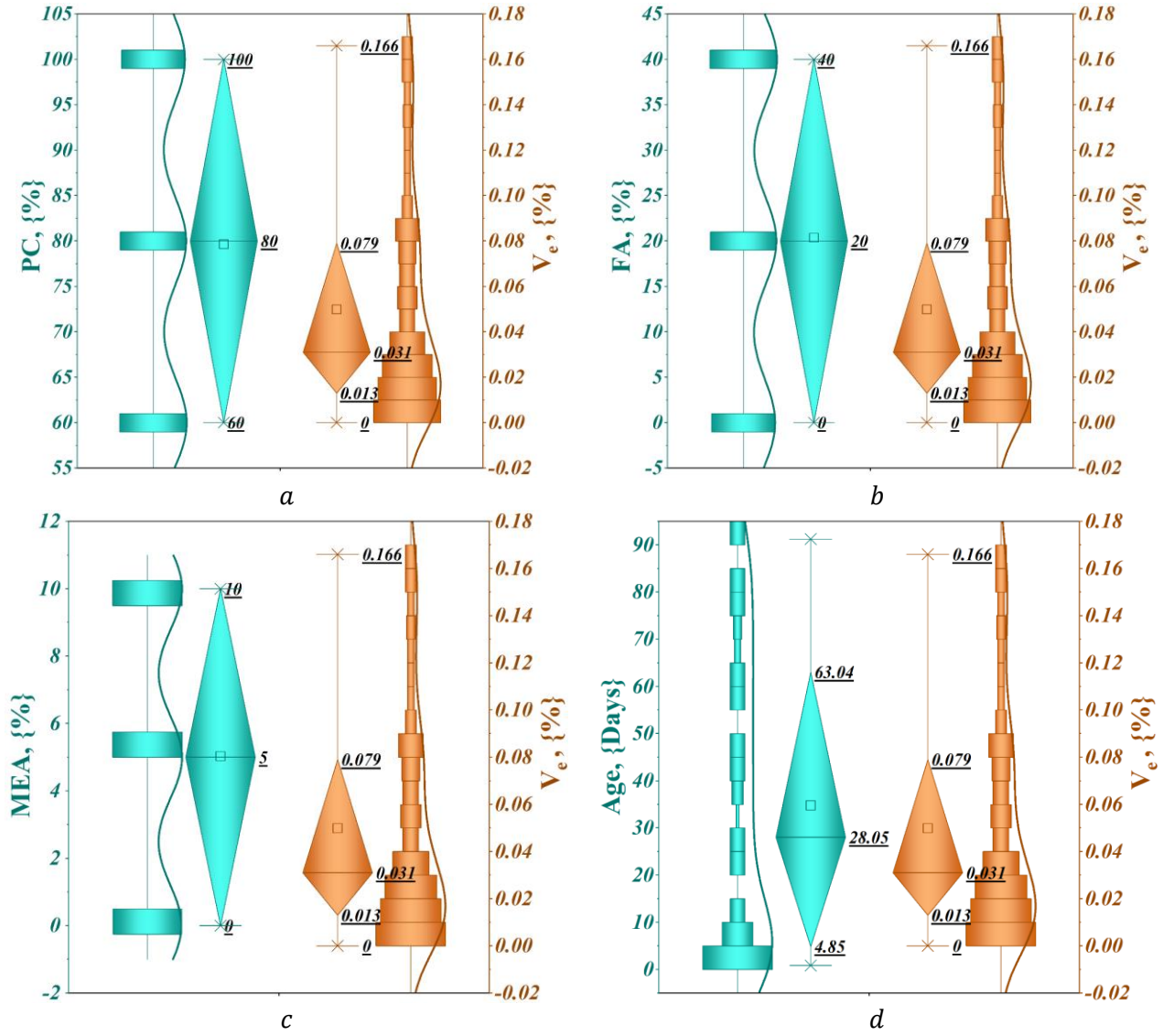


Fig. 1 - Visualization form of the variables versus target

The diagram presented in Figure 2 shows the results of the Morris analysis of the data set, where a method of sensitivity analysis was used to evaluate the importance of the input parameters in the model. Moise analysis examines the effect of changes in input parameters on the outcome parameter. This method actually determines which changes in the input parameter have a great impact on the results of the outcome parameter. For example, part c of Figure 2 shows the input parameters including (PC, FA, MEA, Age) and the corresponding sensibility criterion. The vertical axis of section c shows the sensibility measure (σ), which quantifies the change in the model outcome against changes in each input parameter. PC and FA parameters have the highest sensibility numbers (0.1417), which indicates that they have the most influence on the model result. Bar graphs allow visual comparison of how each parameter contributes to result changeability. This can inform decisions about which parameters to prioritize for further estimation or correction in the model. Morri's analysis, as shown in this chart, is a valuable tool for recognizing important input parameters that significantly affect model outcomes.

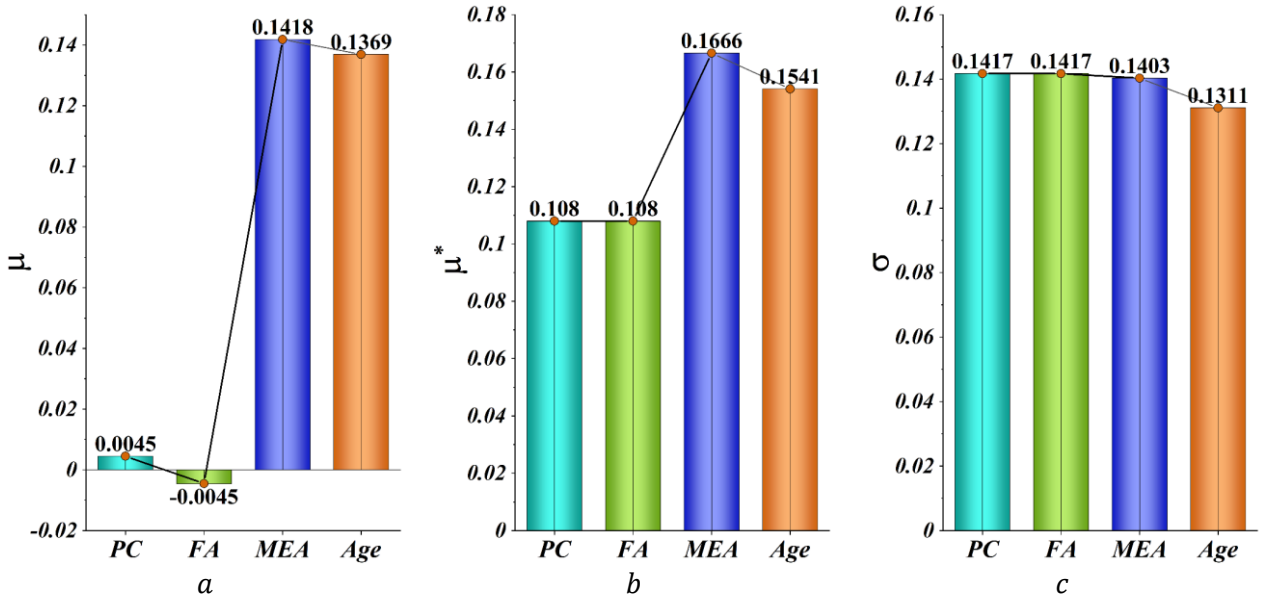


Fig. 2 - Morris analysis for variables' importance evaluation

The chart shown in Figure 3 shows a correlation matrix that shows the Spearman correlation coefficients among different parameters associated with a particular data set. The following explains the main components and concepts of Spearman's correlation: quantities in cells represent Spearman's correlation coefficients, which can range from -1 to 1. A value of 1 indicates a perfect positive correlation (in which a growth in one variable corresponds to a growth in another parameter), while -1 indicates a perfect negative correlation (in which a growth in one parameter leads to a decline in the other). A coefficient of 0 indicates no correlation. Diagonal cells, which reflect the correlation of a variable with itself, show a quantity of 1.0. For instance, the correlation coefficient between the parameters Age and V_e is significantly high at 0.69, which indicates a robust positive correlation. Also, the correlation between MEA parameter and V_e has an acceptable correlation with a quantity of 0.6. Spearman's correlation matrix provides important insights about the connections among various parameters in the data set. It helps identify positive and negative correlations, thereby facilitating further analysis and informed decision-making about matter function. Spearman's correlation is particularly useful in this context, as it evaluates monotonic connections and makes it pliable against the influence of outliers and non-linear associations.

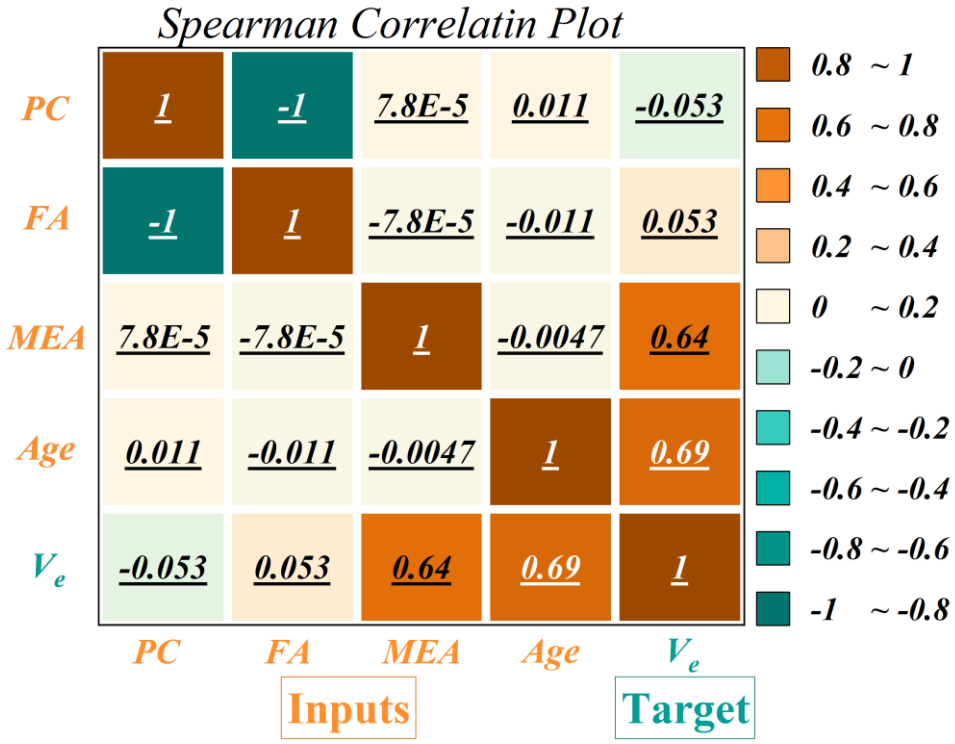


Fig. 3 - Introduced variables correlation matrix values

Electric Eel Foraging Algorithm ($EEFA$)

Electric eels, part of the Gymnotidae family, are deceptive hunters in the animal area, renowned for their terrific capacity to manufacture electric evacuations in freshwater rsurroundings. The phases of utilization and discovery in electric eel foraging optimization are designed based on the social hunt acting observed in electric eels, which encompass their interactions, relaxing tactics, movement strategies, and chasing methods. Statistical models illustrating the searching acting of EEFA optimizations are presented in the following part [27].

Communication Behavior

When eels face a group of fish, they begin to swim together in a coordinated manner, manufacturing a twisting movement. To efficiently capture numerous small fish gathered at the center of this organization, the eels initiate a motion into a large, electrified circle [27]. The interaction indicates that every eel utilizes its situation to communicate with others efficiently. The process of calculating an eel's novel situation involves comparing its location to the center of the crowd and an accidentally selected eel. The distinction between an eel manufactured accidentally within the quest zone and one selected accidentally from the existing crowd is utilized to set the eel's situation. Eels engage with each other through a procedure known as churning, which reflects their erratic motions. This acting is donated by the models shown in the following manner [27].

$$C = n_1 \times B \tag{1}$$

$$n_1 \sim N(0,1) \tag{2}$$

$$B = [b_1, b_2, b_3, \dots, b_k, \dots, b_d] \tag{3}$$

$$b(k) = \begin{cases} 1 & if\ k == g(l) \\ 0 & else \end{cases} \tag{4}$$

$$g = randperm(d) \tag{5}$$

$$l = 1,2, \dots, \left\lceil \frac{T-t}{T} \times r_1 \times (d-2) + 2 \right\rceil \tag{6}$$

T represents the highest value of repetitions allowed.

The concept of inter-active acting was introduced as bellows [27]:

$$\begin{cases} \vartheta_i(t+1) = x_j(t) + C \times \left(\bar{x}(t) - x_i(t)\right), p_1 > 0.5 \\ \vartheta_i(t+1) = x_j(t) + C \times \left(x_r(t) - x_i(t)\right), p_1 \le 0.5 \end{cases}$$
$$\begin{cases} \vartheta_i(t+1) = x_i(t) + C \times \left(\bar{x}(t) - x_j(t)\right), p_2 > 0.5 \\ \vartheta_i(t+1) = x_i(t) + C \times \left(x_r(t) - x_j(t)\right), p_2 \le 0.5 \end{cases} \quad (7)$$

$$\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) \quad (8)$$

$$x_r = Low + r \times (Up - Low) \quad (9)$$

In this context, p_1 and p_2 represent accidental quantities within the range of $(0, 1)$. The parameter x_j denotes the situation of an accidentally chosen eel from the present crowd, where $j \neq i$. The whole crowd size is indicated by n, while $r_1$is another accidental quantity also within the interval $(0, 1)$. The subordinate $fit(i)$ computes the fitness score for the possible crowd of the ith electric eel. Furthermore, r is an accidentally manufactured vector that falls within the range of $(0, 1)$, and Low and Up define the lowest and highest borders of the quest zone in the same order.

It is essential to note that in the primary two instances of Eq. (7), the relation fits as $fit\left(x_j(t)\right) < fit(x_i(t))$. Conversely, in the latter scenarios of this numerical expression, $fit\left(x_j(t)\right) \geq fit(x_i(t))$ is also observed. Additionally, Eq. (9) illustrates that the interactions among electric eels facilitate their motion toward various situations within the discovery region. This acting can be highly advantageous for studying EEFA throughout the quest region.

Relaxing acting

In the quest zone, the zone where any single dimension of an eel's situation vector is moved to the central diagonal is defined as the relaxing region. This resting zone can be described by the following attributes [27]:

$$\left\{X \middle| \left|X - Z(t) \le \alpha_0 \times \left|Z(t) - x_{prey}(t)\right|\right|\right\} \quad (10)$$

$$\alpha_0 = 2.e^{\frac{\mu-1}{T}} \quad (11)$$

$$Z(t) = Low + z(t) \times (Up - Low) \quad (12)$$

$$z(t) = \left(x_{rand(n)}^{rand(d)}(t) - Low^{rand(d)}\right) \quad (13)$$

In this context, α_0 represents the initial scale of the relaxing region while x_{prey} indicates the starting situation vector of the currently identified improved resolution. The parameter r_2 is an accidental quantity manufactured within the interval $(0, 1)$. The expression $\alpha_0 \times \left|Z(t) - x_{prey}(t)\right|$ donates the extent of the relaxing area. Here, z signifies a normalized quantity, and $x_{rand(n)}^{rand(d)}(t)$ refers to a specific situation dimension selected from an accidentally chosen quest factor within the existing crowd.

An eel assumes its relaxing acting within its designated relaxing zone before initiating the relaxing acting described in Equation 14.

$$R_i(t+1) = Z(t) + \alpha \times \left|Z(t) - x_{prey}(t)\right| \quad (14)$$

$$\alpha = \alpha_0 \times sin(2\pi r_3) \quad (15)$$

where r_3 is an accidentally manufactured quantity within the interval $(0, 1)$, and a donates the scale of the relaxing zone.

In its relaxing situation, where this relaxing acting can be replicated, an eel modifies its place as it approaches its relaxing region while maintaining its relaxing place [27]:

$$v_i(t+1) = R_i(t+1) + n_2 \times (R_i(t+1) - round(rand) \times x_i(t)) \quad (16)$$

$$n_2 \sim N(0,1) \quad (17)$$

Application of EEFA

The whole crowd size and the highest value of repetitions are two key control variables that basic EEFA sets up prior to starting the enhancement procedure. An accidentally manufactured set of eels is uniformly dispersed. Each eel employs inter-active acting to manufacture novel applicant resolutions during each repetition when the energy agent $E > 1$. Each eel engages in utilization through relaxing, movement, or chasing, acting with equal probability when the energy agent $E \leq 1$ manufactures fresh applicant resolutions. In every repetition of the loop, the present best resolution is revised following a comparison between the newly found resolution and the available ones. As the procedure continues, the agent E diminishes, compelling each eel to shift from discovery to utilization. The interactive procedure continues until the stopping situation is met. The improved resolution up to that moment is preserved thereafter. A concise overview of EEFA's pseudocode is presented in Algorithm 1, accompanied by its flowchart in Algorithm. (1).

Algorithm 1. A pseudocode highlighting the essential methods of EEFA.
1: Explain the variables n and T 2: Accidentally setting up the crowd $X_i(i = 1,2, \dots, n)$ and evaluate their fitness score fit_i, while X_{prey} represents the improved resolution discovered up to this point. 3: While (The expiration agent has not been met) do 4: for every eel X_i do 5: Calculate E 6: if $E > 1$, then 7: Execute the inter-active acting using Equation 7 8: Calculate the fitness fit_i 9: else 10: if $rand > 1/3$, then 11: Determine the relaxing zone using Equation 14 12: Execute the relaxing procedure 13: Calculate the fitness fit_i 14: else if $rand > 2/3$, then 15: Execute the movement procedure 16: else 17: Determine the chasing zone 18: Execute the chasing 19: end if 20: end if 21: Adjust the situation of each eel 22: end for 23: Revise the resolution achieved to date, X_{prey} 24: end while 25: Revise the resolution X_{prey}

Red-Tailed Hawk Algorithm ($RTHA$)

The procedure of identifying nasty activity utilizes RTHA for attribute selection. This tactic mimics the chasing acting of this creature [28]. The acting exhibited throughout the chasing procedure is examined and exhibited. The three distinct stages of RTHA include low rising, high rising, and bending and stooping.

Members ascend high into the sky, searching for an ideal situation in terms of power supply accessibility, as illustrated by the numerical modeling in Eq. (18):

$$X(t) = X_{best} + \left(X_{mean} - X(t-1)\right).Levy(dim).TF(t) \quad (18)$$

In Eq. (18), the situation of these creatures at the t – th repetition is illustrated as $X(t)$. The best position achieved up to that dot is X_{best}, while X_{mean} indicates the mean of the places. The $Levy$ subordinate refers to the Levy flight expansion, and also, the transition agent subordinate is denoted as $TF(t)$.

$$Levy(dim) = s\frac{\mu \times \sigma}{|\vartheta|^{\beta^{-1}}}$$
$$\sigma = \left(\frac{\Gamma(1+\beta).sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right).\beta.2^{\left(1-\frac{\beta}{2}\right)}}\right) \quad (19)$$

In scenarios where the issue size is small, β is a collection at a fixed value of 1.5, s is fixed at 0.01, and also u and v represent accidental integers within the range of 0 to 1.

$$TF(t) = 1 + sin\left(2.5 + \left(\frac{t}{T_{max}}\right)\right) \quad (20)$$

T_{max} represents the highest value of repetitions. In the low-rising acting, members spiral downwards towards the aim, flying at a lower height. The numerical representation of this is provided in Equations 21:

$$X(t) = X_{best} + \left(x(t) + y(t)\right).StepSize(t)$$
$$StepSize(t) = X(t) - X_{mean} \quad (21)$$

In this context, x and y represent the axes of the coordinate system.

$$\begin{cases} x(t) = R(t).sin(\theta(t)) \\ y(t) = R(t).cos(\theta(t)) \end{cases} \begin{cases} R(t) = R_0.\left(r - \frac{t}{T_{max}}\right).rand \\ \theta(t) = A.\left(1 - \frac{t}{T_{max}}\right).rand \end{cases} \begin{cases} x(t) = \frac{x(t)}{max|x(t)|} \\ y(t) = \frac{y(t)}{max|y(t)|} \end{cases} \quad (22)$$

In this context, r illustrated a control achievement [1], [2], while the primary quantity, represented as R_0, falls within the interval $[0.5 - 3]$. The parameter A signifies the angle achievement $[515]$, and rand refers to an accidental integer within the interval $[0,1]$. This variable facilitates the hawk's helix motion around the aim.

The hawk quickly descends and strikes its aim from the arbitrary situation while gliding at a low altitude, as illustrated in Equation 23:

$$X(t) = \alpha(t).X_{best} + x(t).StepSize1(t) + y(t).StepSize2(t)$$
$$StepSize1(t) = X(t) - TF(t).X_{mean}$$
$$StepSize2(t) = G(t).X(t) - TF(t).X_{best} \quad (23)$$

The agents for acceleration and gravity are donated by α and G, in the same order:

$$\alpha(t) = sin^2\left(2.5 - \frac{t}{T_{max}}\right)$$
$$G(t) = 2.\left(1 - \frac{t}{T_{max}}\right) \quad (24)$$

In Equation (24), the acceleration α of this creature's growth as time t progresses, improving the convergence proportion. Additionally, the gravitational impact G declines the variety of utilization as the hawk approaches the aim.

Selecting a suitable attribute that assists the sorter in recognizing an example class within the data presents a substantial challenge [35].

In the chosen procedure, improving the function of categorization duties is important, along with the automatic deletion of unfitting attributes once the desired ones have been recognized. The RTHA is utilized to identify the best sub-collection of properties and leverages the sorter to assess

the categorization function. A_c represents categorization precision, bs indicates the size of the attribute sub-collection, and D_t refers to the whole quantity of properties in the data collection. The sorter fault can be donated as $1 - A_c$, while the selected sub-collection of properties from the data is expressed as $\frac{d_s}{D_t}$. Thus, the fitness subordinate is defined as below:

$$\downarrow Fitness = \mu.(1 - A_c) + (1 - \mu).\frac{d_s}{D_t} \quad (25)$$

In Equation 25, the weight assigned to the fault categorization is μ, which ranges from 0 to 1.

To gain a clearer insight into the properties of the suggested RTHA and its basics, the concepts related to RTHA are outlined in Algorithm 2.

Algorithm 2. Pseudocode of RTHA.
Setting up: Accidental manufacture within the discovery region While $t < T_{max}$ do Elevated level: for $i = N_{pop}$ do Determine the Levy flight broadcast Determine the movement agent TF Revise situations Finish Decline level: for $i = N_{pop}$ do Determine the directional coordinates Revise situations Finish The phase of bending and gliding: for $i = N_{pop}$ do Determine the acceleration and the gravitational agents Determine the StepSize Revise situations Finish Finish

Extra Tree Regression (ETR)

The ETR introduced by Geurts et al. [36] is a group learning technique that depends on a large quantity of decision trees (DTs). The ETR differs from the random forest method in 2 significant ways: (1) the ETR selects an accidental binary dividing number to divide the nodes of the DT, whereas the random forest method employs an improved dividing factor. (2) The ETR leverages all available data dots to diminish divergence, whereas the random forest method employs the boot-strap tactic, a sampling way that allows for substitution. The ETR offers several benefits, including (1) a direct and impactful node dividing procedure compared to other group learning tactics. (2) The ETR competently addresses the issue of over-fitting due to its considerable unpredictability and leverages the whole data collection to decline divergence.

Workability Indicators

To comprehensively assess ETR models, it is essential to include many performance metrics since each criterion provides distinct insights into the model's precision and dependability.

Performance Index (PI) by Equation 26
Coefficient of Determination (R^2) by Equation 27
Root Mean Square Error ($RMSE$) by Equation 28
Symmetric Mean Absolute Percentage Error ($SMAPE$) by Equation 29
Mean Absolute Scaled Error ($MASE$) by Equation 30
Theil's U Statistic (U) by Equation 31

Willmott's Index of Agreement (WI) by Equation 32
Uncertainty with a 95-coincidence level ($U_{95}\%$) by Equation 33
Scatter index (SI) by Equation 34

$$PI = \frac{1}{\overline{N}}\frac{RMSE}{\sqrt{R^2+1}} \tag{26}$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}(N_i-\overline{N})(O_i-\bar{O})}{\sqrt{\left[\sum_{i=1}^{n}(N_i-N)^2\right]\left[\sum_{i=1}^{n}(O_i-\bar{O})^2\right]}}\right) \tag{27}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i-N_i)^2} \tag{28}$$

$$SMAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|N_i-O_i|}{\frac{(|N_i|+|O_i|)}{2}}\times 100 \tag{29}$$

$$MASE = \frac{\frac{1}{n}\sum_{i=1}^{n}|N_i-O_i|}{\frac{1}{n-1}\sum_{i=1}^{n}|N_i-N_{i-1}|} \tag{30}$$

$$U = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i-N_i)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}N_i^2}+\sqrt{\frac{1}{n}\sum_{i=1}^{n}O_i^2}} \tag{31}$$

$$WI = 1-\frac{\sum_{i=1}^{n}(N_i-N_i)^2}{\sum_{i=1}^{n}(|O_i-\bar{O}|+|N_i-\bar{O}|)^2} \tag{32}$$

$$U_{95} = 1.96\sqrt{(SD^2+RMSE^2)} \tag{33}$$

$$SI = \frac{\sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}((N_i-\overline{N})-(O_i-\bar{O}))^2}}{\left(\frac{1}{n}\right)\sum_{n=1}^{i}O_i} \tag{34}$$

where:
N_i : The measured V_e.
$\overline{N}$: The means of the measured V_e.
O_i : The expected V_e.
$\bar{O}$: The mean of the anticipated V_e.
n : The total number of observations.

FINDINGS AND COMPARISON

Step-by-step Development

An adaptive and consistent ML approach capable of addressing both classification and regression challenges with remarkable endurance and accuracy has been implemented utilizing the ETR model. Enhancing feature relevance, scalability, and flexibility are other objectives of this model. The hyperparameters of the ETR framework that significantly influence effectiveness may be modified via trial and error or optimization techniques. This methodology seeks to improve predictive precision and address missing data issues.

Optimizing an ETR model demands the meticulous adjustment of essential hyperparameters, including $n\_estimators$, max_depth, $min\_samples\_split$, $min\_samples\_leaf$, and $max\_features$, to achieve equilibrium among model intricacy and generalizability, as demonstrated in Table 2. The

procedure is optimized by first partitioning the dataset into 75% for learning and 25% for assessment. Innovative optimization approaches, including the $RTHA$ and the $EEFA$, are used to investigate various hyperparameter mixtures and minimize the $RMSE$ on the validation set. The ideal hyperparameter set is determined after comprehensive training and assessment of many models, taking into account the unique efficacy of each model. Table-based diagram for the research program and the tested models and the adopted variables is supplied in Table 3.

Tab. 2 - The related parameters and hyperparameters of models

Optimization algorithms	Initial Parameters	Value	Coupled models		Named	Hyperparameters	Optimal Value
EEFA	Total of iterations	200	ETR+EEFA	⋙	ETR(E)	$n\_estimators$	105
	Total of tries	5				max_depth	18
	Total of population	30				$min\_samples\_split$	2
	Parameter free	-				$min\_samples\_leaf$	3
						$max\_features$	0.76
RTHA	Total of iterations	200	ETR+ RTHA	⋙	ETR(R)	$n\_estimators$	153
	Total of tries	5				max_depth	28
	Total of population	50				$min\_samples\_split$	3
	A	15				$min\_samples\_leaf$	3
	$R0$	0.5				$max\_features$	0.58
	r	1.5					
	N	30					

Tab. 3 - Table-based diagram for the research program and the tested models and the adopted variables

Aspect	Details
Research Objective	"Predicting Volume Expansion of Cement Paste"
Key Phases	Data Collection and Preprocessing. Model Development and Optimization. Model Evaluation and Analysis.
Models Tested	Extra Tree Regression (ETR) optimized with RTHA (ETR (R)) and EEFA (ETR (E))
Optimization Methods	Red-Tailed Hawk Algorithm (RTHA), Electric Eel Foraging Algorithm (EEFA)
Input Variables	Fly Ash (FA), Portland Cement (PC), MgO Expansive Addition (MEA), Sample Age (Age)
Dataset Size	170 samples
Performance Metrics	PI, R^2, $RMSE$, $SMAPE$, $MASE$, U, WI, U_{95}%, and SI
Best Performing Model	ETR (R) (with RTHA tuning)
Key Findings	ETR (R) achieved RMSE values of 0.0053 (training) and 0.01 (testing), outperforming ETR (E) by 28% and 14% respectively.
Recommendations	Expand dataset, incorporate additional parameters, test hybrid models.

Comparison and Justification

This research demonstrates the results of integrating hybridized ETR approaches with the $EEFA$ and $RTHA$ methodologies. The determination of the V_e of cement paste, consisting of FA and MEA, was performed using these techniques. Figure 4 depicts the projected and actual values of V_e during the training and assessment stages of the $ETR\ (E)$ and $ETR\ (R)$ methodologies, which were developed. Furthermore, the figure illustrates the distribution of the residual %. The accuracy of methods for predicting V_e was assessed using several metrics, including PI, R^2, $RMSE$, $SMAPE$, $MASE$, U, WI, $U_{95\%}$, and SI. Furthermore, the rating-based methodology has been established to allocate a score to each index associated with each strategy, with the greatest score awarded to the best suitable value. Table 4 presents the measurement results for the designs, highlighting the optimal scenario and the corresponding scores attained during both the learning and evaluation phases.

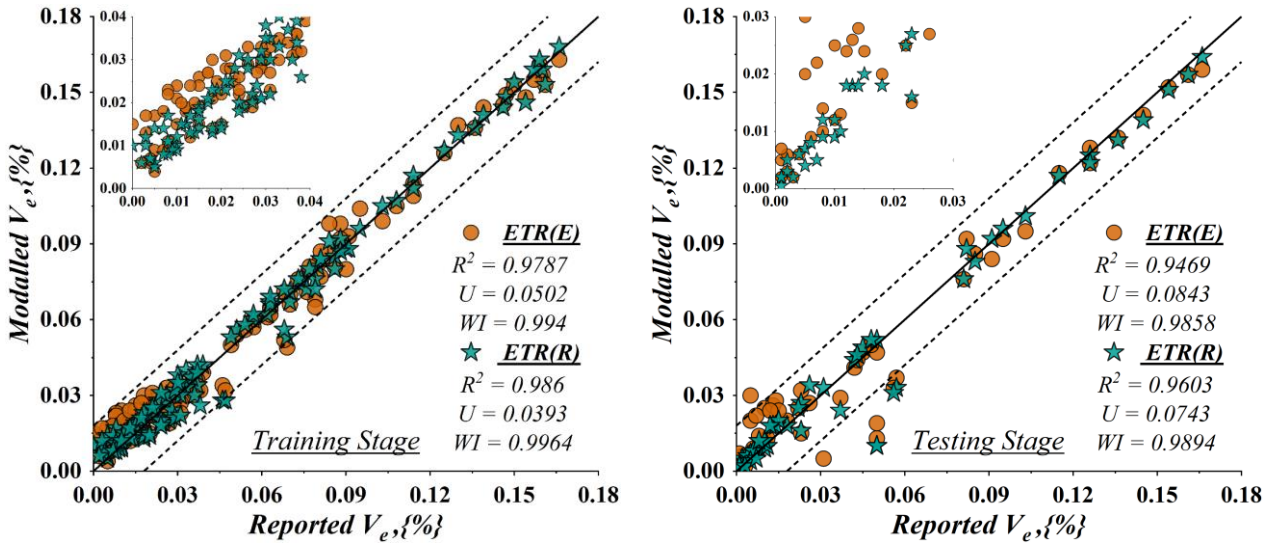
The results indicate that the $ETR\ (E)$ and $ETR\ (R)$ methodologies have significant potential for accurately forecasting the V_e. During the training and assessment phase, the $ETR\ (R)$ approach exhibited R^2 values of 0.986 and 0.9603, signifying exceptional precision. The findings for $ETR\ (E)$ were notably lower than those for $ETR\ (R)$, with R^2 values of 0.9787 and 0.9469, respectively. The optimal model was combined with WI, and the results exhibited a performance trend akin to one of the expected measures. Supplementary error-based indicators are also useful in evaluating the validity of the methodology. In the present study, numerous indicators, including PI, $RMSE$, $SMAPE$, $MASE$, U, $U_{95\%}$, and SI were calculated. It is essential to recognize that decreased values signify superior performance for these metrics. The variation percentages of the two models developed for these measures are a minimum of 13%; in some cases, this variation decreases by 53%, illustrating the $ETR\ (R)$'s predictive reliability and efficacy. For instance, for the $RMSE$ index, $ETR\ (R)$ achieved values of 0.0053 and 0.01 during the training and testing phases, which are about 28% and 14% lower than the corresponding values of $ETR\ (E)$ at 0.0068 and 0.0114, correspondingly.

An uncertainty interval, also known as a confidence interval, is a range of numbers that expresses predicting uncertainty. It signifies the dependability of the estimate. Confidence intervals typically denote 95% or 99% confidence levels. It assesses projected dependability. Numerous confidence intervals are linked to 95% or 99% certainty. This research used a 95% confidence interval (lower numbers imply more reliability). The $U_{95\%}$ index values for the evaluation and learning phases were 0.0147 and 0.0275, respectively, based on the $ETR\ (R)$. The corresponding $U_{95\%}$ values for the $ETR\ (E)$ were 0.0187 and 0.0316. The specs stayed consistent during the project. After evaluating evaluation factors, logical inference, and rating levels, it has been concluded that all models are deemed credible and trustworthy. The $ETR\ (R)$ model is somewhat superior to the alternative model regarding its purpose.

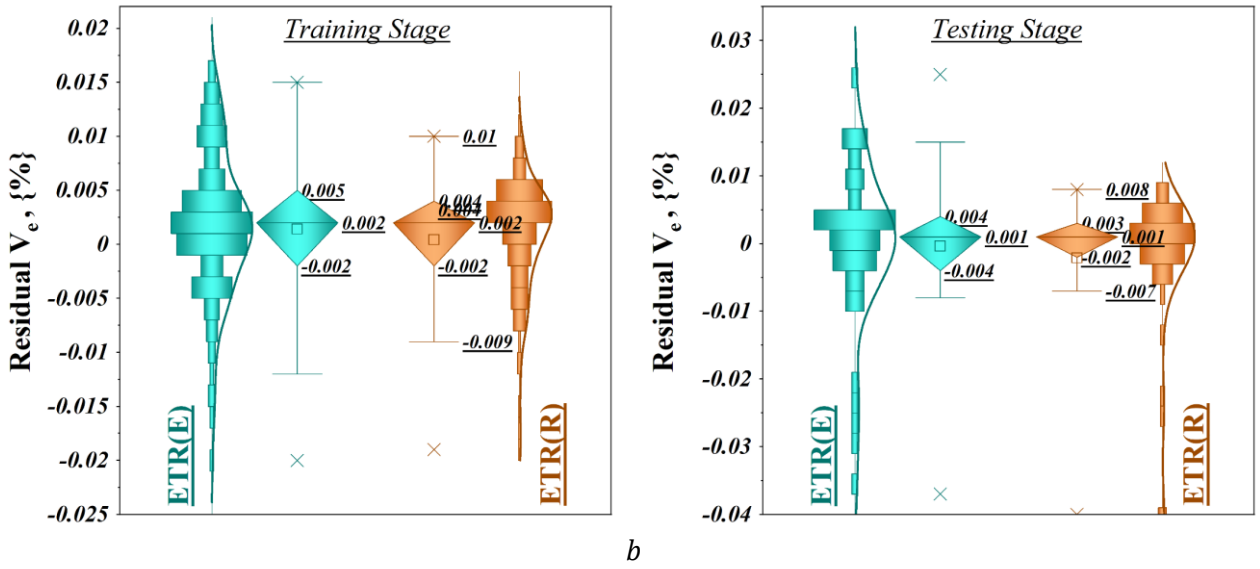
The residual values of the V_e denotes the discrepancy between the model's anticipated value and the actual value derived from experiments. The residual is essential for evaluating the precision of the forecasting model, as it signifies the deviation between the expected and actual volume expansion of cement paste. A positive residual signifies that the model has underestimated the volume expansion of cement paste, whereas a negative residual implies an overestimation. Assessing these residual values enhances model accuracy, augments material evaluation, and guarantees that the volume expansion of cement paste adheres to requisite building criteria. Fig. 4(b) indicates that the mean values of $ETR\ (E)$ are 0.002 in the process of learning and 0.001 in the assessment stage. The $ETR\ (R)$ values are 0.002 in the learning phase and 0.001 in the assessment stage. The residual amount clearly demonstrates that the $ETR\ (R)$ surpasses the $ETR\ (E)$ in both the assessment and procurement phases.

Tab. 4 - The coupled and tunned $ETRs$' results

Trees	Workability indicators								
	PI	R^2	$RMSE$	$SMAPE$	$MASE$	U	WI	$U95\%$	SI
	"Training stage"								
ETR(E)	0.0672	0.9787	0.0068	23.607	0.1111	0.0502	0.994	0.0187	0.1337
Grade	B	B	B	B	B	B	B	B	B
ETR(R)	0.0525	0.986	0.0053	18.934	0.0885	0.0393	0.9964	0.0147	0.1046
Grade	A	A	A	A	A	A	A	A	A
Variance %	-28	0.74	-28.30	-24.67	-25.53	-27.73	0.241	-27.21	-27.82
	"Testing stage"								
ETR(E)	0.121	0.9469	0.0114	41.912	0.1209	0.0843	0.9858	0.0316	0.2387
Grade	B	B	B	B	B	B	B	B	B
ETR(R)	0.1062	0.9603	0.01	27.343	0.0833	0.0743	0.9894	0.0275	0.2102
Grade	A	A	A	A	A	A	A	A	A
Variance %	-13.93	1.395	-14	-53.28	-45.13	-13.45	0.364	-14.90	-13.55



a



b

Fig. 4 - The coupled and tunned ETRs' results

This has very strong practical implications for construction materials, cement formulation, and quality control. By allowing the model to predict the V_e accurately, in a cement paste with FA and MgO expansive additives, MEA, it helps engineers to optimize material selection by reducing costs and improving structural integrity, as it prevents cracks and deformation. The infrastructure, therefore, lasts longer with minimal maintenance needs. It also currently contributes to quality control, ensuring consistent product performance while contributing to sustainability by efficient use of FA, reducing carbon footprint, and the optimization of MEA applications. Furthermore, it enables customized cement formulation according to project requirements in new construction and rehabilitation of existing ones. This model will improve risk assessments and the follow-through of standards on safety, adding more solidity to the general safety and durability of the project, thus fitting well into current construction practice.

Following suggestions can be considered as recommendations for future works. Exploring combination of the ETR with other algorithms (e.g., stacking or ensemble methods) to enhance prediction accuracy. Including other potentially relevant parameters, such as curing temperature, water content, or admixture type, to improve the model's comprehensiveness. And testing the models against noisy or incomplete data to simulate real-world scenarios and evaluate their resilience.

CONCLUSION

This investigation aimed to develop and evaluate ML techniques for assessing the volume expansion of cement paste (V_e) using FA and magnesium oxide (MgO) expansive additives (MEA). An algorithm known as ETR has been devised to achieve this objective. The present work utilized advanced optimizers, namely the EEFA and the RTHA, for optimization purposes.

The results of the Morris analysis of the data set showed which input parameter would have the greatest impact on the result of the model. For instance, in part c, PC and FA parameters have the highest sensibility values (0.1417), which shows that they have the most influence on the model result. Bar graphs allow a visual comparison of how each parameter contributes to the variability of the outcomes.

The presented correlation matrix shows the Spearman correlation coefficients among the different variables associated with a particular data set. For instance, the correlation coefficient among Age and V_e parameters are significant and are high at 0.69, which indicates a robust positive

correlation, as the correlation among MEA variable and V_e has an acceptable correlation with a number of 0.6.

The results indicate that the $ETR\ (E)$ and $ETR\ (R)$ methodologies have significant potential for accurately forecasting the V_e. During the training and assessment, the $ETR\ (R)$ exhibited R^2 of 0.986 and 0.9603, signifying exceptional precision. The findings for $ETR\ (E)$ were notably lower than those for $ETR\ (R)$, with R^2 of 0.9787 and 0.9469. The optimal model was combined with WI, and the results exhibited a performance trend akin to one of the expected measures.

It is essential to recognize that decreased values signify superior performance for workability indicators. The variation percentages of the two models developed for these measures are a minimum of 13%; in some cases, this variation decreases by 53%, illustrating the $ETR\ (R)$'s predictive reliability and efficacy. For instance, for the $RMSE$ index, $ETR\ (R)$ achieved values of 0.0053 and 0.01 during the training and testing phases, which are about 28% and 14% lower than the corresponding values of $ETR\ (E)$ at 0.0068 and 0.0114, correspondingly.

The $U_{95\%}$ values for the evaluation and learning were 0.0147 and 0.0275, based on the $ETR\ (R)$. The corresponding $U_{95\%}$ for the $ETR\ (E)$ were 0.0187 and 0.0316. The specs stayed consistent during the project. After evaluating evaluation factors, logical inference, and rating levels, it has been concluded that all models are deemed credible and trustworthy. The $ETR\ (R)$ model is somewhat superior to the alternative model regarding its purpose.

The impact of this research allows practical applications: cement formulation optimization, structural integrity, and enhanced quality control in the building and construction industry. An example could be an exact prediction of volume expansion V_e by the model, which gives very valuable cost reductions by preventing cracks and material performance consistency in cement paste with FA and MgO additives MEA. In this respect, it also optimizes the use of FA and MEA for sustainability purposes apart from customized cement mixes based on project-specific requirements.

REFERENCES

[1] Bofang Z (2013) Thermal stresses and temperature control of mass concrete. Butterworth-Heinemann

[2] Benemaran RS, Esmaeili-Falak M, Kordlar MS (2024) Improvement of recycled aggregate concrete using glass fiber and silica fume. Multiscale and Multidisciplinary Modeling, Experiments and Design 7:1895–1914. https://doi.org/10.1007/s41939-023-00313-2

[3] Ulm F-J, Coussy O (1995) Modeling of thermochemomechanical couplings of concrete at early ages. J Eng Mech 121:785–794

[4] Esmaeili-Falak M, Sarkhani Benemaran R (2024) Application of optimization-based regression analysis for evaluation of frost durability of recycled aggregate concrete. Structural Concrete 25:716–737. https://doi.org/https://doi.org/10.1002/suco.202300566

[5] Liu X, Zhang C, Chang X, et al (2015) Precise simulation analysis of the thermal field in mass concrete with a pipe water cooling system. Appl Therm Eng 78:449–459

[6] Hassankhani E, Esmaeili-Falak M (2024) Soil–Structure Interaction for Buried Conduits Influenced by the Coupled Effect of the Protective Layer and Trench Installation. J Pipeline Syst Eng Pract 15:. https://doi.org/10.1061/JPSEA2.PSENG-1547

[7] Mo L, Deng M, Wang A (2012) Effects of MgO-based expansive additive on compensating the shrinkage of cement paste under non-wet curing conditions. Cem Concr Compos 34:377–383

[8] Min D, Mingshu T (1994) Formation and expansion of ettringite crystals. Cem Concr Res 24:119–126

[9] Mo L, Deng M, Tang M, Al-Tabbaa A (2014) MgO expansive cement and concrete in China: Past, present and future. Cem Concr Res 57:1–12

[10] Gao P, Xu S, Chen X, et al (2013) Research on autogenous volume deformation of concrete with MgO. Constr Build Mater 40:998–1001

[11] Nobre J, Ahmed H, Bravo M, et al (2020) Magnesia (MgO) production and characterization, and its influence on the performance of cementitious materials: A review. Materials 13:4752

[12] Esmaeili-Falak M, Benemaran RS (2023) Ensemble deep learning-based models to predict the resilient modulus of modified base materials subjected to wet-dry cycles. Geomechanics and Engineering 32:583–600

[13] Liang S, Shen Y, Gao X, et al (2023) Symbolic machine learning improved MCFT model for punching shear resistance of FRP-reinforced concrete slabs. Journal of Building Engineering 69:106257. https://doi.org/https://doi.org/10.1016/j.jobe.2023.106257
[14] Esmaeili-Falak M, Katebi H, Vadiati M, Adamowski J (2019) Predicting triaxial compressive strength and Young's modulus of frozen sand using artificial intelligence methods. Journal of Cold Regions Engineering 33:4019007. https://doi.org/https://doi.org/10.1061/(ASCE)CR.1943-5495.0000188
[15] Esmaeili-Falak M, Benemaran RS (2024) Ensemble Extreme Gradient Boosting based models to predict the bearing capacity of micropile group. Applied Ocean Research 151:104149. https://doi.org/10.1016/j.apor.2024.104149
[16] Zhang K, Zhang Y, Razzaghzadeh B (2024) Application of the optimal fuzzy-based system on bearing capacity of concrete pile. Steel and Composite Structures 51:25
[17] Nobre J, Ahmed H, Bravo M, et al (2020) Magnesia (MgO) production and characterization, and its influence on the performance of cementitious materials: A review. Materials 13:4752
[18] Yang Y, Li C (1999) Study on long term autogenous volume deformation of MgO concrete. J Hydraul Eng 65:54–58
[19] Zhu B (2003) Incremental type of computing model for the volume expansion of concrete with gentle volume expansion. Hydroelec Eng 29:20–23
[20] Zhu B (2002) Computational model and experimental method for bulk expansion of gentle expansive concrete. J Hydraul Eng 33:18–21
[21] Yang G, Yuan M (2004) The hyperbola model for autogenous expansion volume deformation of MgO concrete. J Hydroel Eng 4:38–44
[22] Zhang G, Chen X, Du LH (2005) Analysis of expansion effect of magnesium oxide (MgO) in mass concrete. J Hydraul Eng 2:185–189
[23] Zhang G (2002) MgO micro-expansion thermal integral model considering temperature process effect. Water Power 11:28–32
[24] Feng C, Zhao C, Yu X, et al (2021) A Mathematical Model of the Expansion Evolution of Magnesium Oxide in Mass Concrete Based on Hydration Characteristics. Materials 14:3162
[25] Liu SH, Fang KH (2005) Study on autogenous deformation of concrete incorporating MgO as expansive agent. Key Eng Mater 302:155–161
[26] Xu P, Zhu Y, Ben N (2008) Computing model for autogenous volume deformation of MgO concrete based on degree of hy-dration. Water Resour Hydrop Eng 2:22–25
[27] Zhao W, Wang L, Zhang Z, et al (2024) Electric eel foraging optimization: A new bio-inspired optimizer for engineering applications. Expert Syst Appl 238:122200
[28] Ferahtia S, Houari A, Rezk H, et al (2023) Red-tailed hawk algorithm for numerical optimization and real-world problems. Sci Rep 13:12950. https://doi.org/10.1038/s41598-023-38778-3
[29] Simm J, De Abril IM, Sugiyama M (2014) Tree-based ensemble multi-task learning method for classification and regression. IEICE Trans Inf Syst 97:1677–1681
[30] Zhang J, Lv T, Hou D, Dong B (2023) Synergistic effects of fly ash and MgO expansive additive on cement paste: Microstructure and performance. Constr Build Mater 371:130740
[31] Dawei Y, Bing Z, Bingbing G, et al (2023) Predicting the CPT-based pile set-up parameters using HHO-RF and PSO-RF hybrid models. Structural Engineering and Mechanics 86:673–686
[32] Yaychi BM, Esmaeili-Falak M (2024) Estimating Axial Bearing Capacity of Driven Piles Using Tuned Random Forest Frameworks. Geotechnical and Geological Engineering. https://doi.org/10.1007/s10706-024-02952-9
[33] Li D, Zhang X, Kang Q, Tavakkol E (2023) Estimation of unconfined compressive strength of marine clay modified with recycled tiles using hybridized extreme gradient boosting method. Constr Build Mater 393:131992. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2023.131992
[34] Sun X, Dong X, Teng W, et al (2024) Creation of regression analysis for estimation of carbon fiber reinforced polymer-steel bond strength. Steel and Composite Structures 51:509–527
[35] Akinola OA, Ezugwu AE, Oyelade ON, Agushaka JO (2022) A hybrid binary dwarf mongoose optimization algorithm with simulated annealing for feature selection on high dimensional multi-class datasets. Sci Rep 12:14945. https://doi.org/10.1038/s41598-022-18993-0
[36] Simm J, De Abril IM, Sugiyama M (2014) Tree-based ensemble multi-task learning method for classification and regression. IEICE Trans Inf Syst 97:1677–1681