



OPEN

# Optimized SVR with nature-inspired algorithms for environmental modelling of mycotoxins in food virtual-water samples

Abdullahi G. Usman<sup>1,2</sup>✉, Sagiru Mati<sup>2,3</sup>, Hanita Daud<sup>4</sup>, Ahmad Abubakar Suleiman<sup>4</sup>, Sani I. Abba<sup>5</sup>, Hijaz Ahmad<sup>2,6</sup> & Taha Radwan<sup>7</sup>✉

The accurate determination of mycotoxins in food samples is crucial to guarantee food safety and minimize their toxic effects on human and animal health. This study proposed the use of a support vector regression (SVR) predictive model improved by two metaheuristic algorithms used for optimization namely, Harris Hawks Optimization (HHO) and Particle Swarm Optimization (PSO) to predict chromatographic retention time of various food mycotoxin groups. The dataset was collected from secondary sources and used to train and validate the SVR-HHO and SVR-PSO models. The performance of the models was assessed via mean square error, correlation coefficient, and Nash–Sutcliffe efficiency. The SVR-HHO model outperformed existing methods by 4–7% in both the two learning (training and testing) phases respectively. By using metaheuristic optimization, parameter adjustment became more effective, avoiding trapping in local minima and improving model generalization. These results demonstrate how machine learning and metaheuristics may be combined to accurately forecast mycotoxin levels, providing a useful tool for regulatory compliance and food safety monitoring. The SVR-HHO framework is perfect for commercial quality assurance, regulatory testing, and extensive food safety programs because it provides exceptional accuracy and resilience in predicting mycotoxin retention times. In contrast to conventional models, SVR-HHO effectively manages intricate nonlinear interactions, guaranteeing accurate mycotoxin identification and improving food safety while lowering hazards to human and animal health.

**Keywords** Mycotoxins, Food virtual-water samples, Machine learning, Harris-Hawks Optimization, Retention time

Many crops and cereals are prone to fungal attacks in the field as well as during storage. These fungi are capable of producing various groups of secondary chemical metabolites popularly known as mycotoxins<sup>1</sup>. Various factors influence the levels of mycotoxins in certain food or food materials, which increased the favorable conditions for the fungal growth and invasion<sup>2</sup>. Most of the mycotoxins are identified after they have instigated different subacute health issues in humans, as well as other livestock, negatively affecting various systems and organs. Based on this, most of the mycotoxins have been reported to be potent agents for human and animal carcinogens. Despite climatic and geographical differences in the occurrence and production of mycotoxins, the exposure to these chemical substances is global, with most of the world's food supply being contaminated to some extent<sup>3</sup>. Therefore, monitoring to check the presence of mycotoxins is of paramount significance.

<sup>1</sup>Department of Analytical Chemistry, Faculty of Pharmacy, Near East University, 99138 Nicosia, Turkish Republic of Northern Cyprus. <sup>2</sup>Operational Research Centre in Healthcare, Near East University, Nicosia, Turkish Republic of Northern Cyprus. <sup>3</sup>Department of Economics, Northwest University, Kano, Nigeria. <sup>4</sup>Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Malaysia. <sup>5</sup>Department of Civil Engineering, Prince Mohammad Bin Fahd University, 31952 Al Khobar, Saudi Arabia. <sup>6</sup>Department of Mathematics, College of Science, Korea University, 145 Anam-ro, Seongbuk-gu, Seoul 02841, South Korea. <sup>7</sup>Department of Management Information Systems, College of Business and Economics, Qassim University, 51452 Buraydah, Saudi Arabia. ✉email: abdullahigusman@gmail.com; t.radwan@qu.edu.sa

Different investigative tools can be used in checking the presence of these chemical substances as well as to check their amounts in order to design a protocol towards their prevention and control<sup>4</sup>. Knowledge of chromatographic techniques such as gas chromatography (GC)<sup>5</sup>, thin layer chromatography (TLC)<sup>6</sup>, high performance thin layer chromatography (HPTLC), ultra-high performance chromatography (UPLC) and high performance liquid chromatography for the detection of mycotoxins using different detection approach such as mass spectrometry (MS) and diode array detector (DAD) etc<sup>7–12</sup>.

For example, Zhang et al. also indicated the instantaneous estimation of diverse aflatoxin G1, G2, B1, B2 T-2 toxin, Zearalanone and Ochratoxin A using the A locally manufactured column was used to hyphenate tandem mass spectrometry (HPLC) data. The suggested approach satisfies every prerequisite taken into account for rapid sample preparation and increased sensitivity for multiple discovery<sup>13,14</sup>, also demonstrated the examination of several beer-brewing samples to detect aflatoxin and ochratoxin. Mycotoxins known as ochratoxin and aflatoxin are produced from samples of fungi, opaque beer, sorghum malt, and maize. The HPLC technique was used to perform the analysis. The outcome demonstrated the necessity of employing the chromatographic approach to assess the mycotoxins present in beer samples. Pakshir et al. used an HPLC method developed with a fluorescence detector (FD) to equally assess different classes of aflatoxin and ochratoxin from tea samples. The authors suggested that, given the significant level of mycotoxin contamination in the tea samples, frequent and routine evaluation during the tea's processing could be utilized to enhance the tea's quality<sup>15</sup>.

Even though, various studies have been depicted in the literature regarding the applications of chromatography in the determination and elucidation of mycotoxins, but the major setbacks consist the usage of large amounts of toxic solvents such as chloroform, high cost of analysis and long analysis time<sup>16</sup>. This in turn made chromatographic technique complex and tough to present a green technique that can cover all the essential variables needed to give result within short analysis time, and the need for manual parameter tuning, making them inefficient for large-scale screening programs<sup>17</sup>. Recently, chromatographers have been employing the application of chemometrics in various fields of chromatography to simulate chromatographic systems' behavior. This approach aims to obtain reasonable results with low analysis time, minimize the use of toxic solvents, be cost-effective, and, more importantly, use green methods<sup>18</sup>. However, the traditional chromatographic approach serves as the primary source of the data as well the foundation for this kind of hyphenation<sup>19–22</sup>. The use of artificial intelligence (AI) techniques has proved to be reliable and promising chemometrics-based approach for the determination of mycotoxins using HPLC technique. Among AI-driven techniques, SVR is particularly effective in capturing nonlinear relationships in complex datasets. However, its performance depends on optimal hyperparameter selection, which can be challenging using conventional methods. Metaheuristic algorithms, such as Harris Hawks Optimization (HHO) and Particle Swarm Optimization (PSO), enhance SVR by automating hyperparameter tuning, improving predictive accuracy and robustness. The integration of AI-driven approaches in food safety has gained significant attention, with advancements in predictive modeling contributing to more efficient contaminant detection, quality control, and regulatory enforcement. By leveraging SVR-HHO for chromatographic modeling, this study aligns with global food safety standards, such as those established by the Codex Alimentarius and the European Food Safety Authority (EFSA), offering a scalable, high-precision tool for mycotoxin screening in food industries and regulatory agencies<sup>23</sup>. Numerous studies have emphasized the drawbacks of conventional chromatographic methods, such as their dependence on hazardous solvents, protracted processing periods, and human parameter adjustment<sup>24–28</sup>.

Recently, chromatographers have increasingly turned to chemometrics to simulate chromatographic systems' behavior, aiming to obtain reliable results with reduced analysis time, minimized use of toxic solvents, cost-effectiveness, and environmentally friendly methods<sup>29</sup>. SVR has emerged as a powerful machine learning technique for these purposes due to its robustness in handling nonlinear relationships and its ability to provide high prediction accuracy. SVR works by finding the hyperplane that best fits the data, minimizing the error within a specified margin. Despite its advantages, SVR has some limitations, such as sensitivity to the choice of hyperparameters and potential overfitting, particularly with small or noisy datasets<sup>30</sup>.

To address these limitations, optimizing SVR's hyperparameters is crucial. Traditional methods like grid search can be computationally expensive and may not always find the global optimum. This is where metaheuristic optimization methods, such as Harris Hawks Optimization (HHO) and Particle Swarm Optimization (PSO), come into play. These algorithms can efficiently navigate complex, multidimensional search spaces, enhancing SVR's performance by finding optimal hyperparameters that traditional methods might miss. Recent developments in meta-heuristic optimization have improved food safety and environmental science prediction modeling. Algorithms that have recently been proposed perform better at solving complex problems. While the Fata Morgana Algorithm (FATA)<sup>31</sup> uses refraction mechanisms inspired by mirages for improved convergence, the Educational Competition Optimizer (ECO)<sup>32</sup> mimics student competition for optimal learning. The Rime Optimization Algorithm (RIME)<sup>33</sup> simulates frost formation for efficient worldwide search, while the Polar Lights Optimization (PLO)<sup>34</sup> simulates auroras to strike a balance between exploration and exploitation. In optimization challenges, these new algorithms have demonstrated great promise. By incorporating such cutting-edge methods, predictive models can be further enhanced, especially in the areas of environmental monitoring and food safety, making them useful resources for researchers studying nature-inspired computing and data-driven decision-making. Furthermore, apart from conventional techniques inspired by nature, a number of innovative algorithms have surfaced. To improve convergence time, the gradient-based optimizer (GBO)<sup>35</sup> combines gradient information with metaheuristic techniques. Inspired by slime molds' clever foraging strategies, the Slime Mold Algorithm (SMA)<sup>36</sup> effectively strikes a balance between exploration and exploitation. To increase the searchability of solutions, the Heap-Based Optimizer (HBO)<sup>37</sup> simulates the hierarchical behavior of heap structures in data organization. In order to steer clear of local optima, the Escape Algorithm (ESC)<sup>38</sup> mimics escape processes in prey-predator interactions. Also, JASMA-SVM model was proposed by Shi et al.<sup>39</sup> and achieved 92.998% accuracy in predicting recurrent spontaneous abortion. PM-

SMEKLM was presented by Fei et al.<sup>40</sup> as a transfer learning method for diagnosing brain diseases. MaOPEO, an optimization algorithm created by Chen et al.<sup>41</sup> outperforms current techniques in many-objective problems by increasing convergence and variety. More information on machine learning, optimization, deep learning and data balancing techniques can be found in<sup>41–45</sup>. Recent studies have explored various optimization algorithms to enhance machine learning models in complex problem domains. For instance, Hassan et al.<sup>46</sup> has demonstrated superior convergence speed and accuracy in solving high-dimensional optimization problems. Integrating such techniques into SVR parameter tuning can potentially improve prediction accuracy and model stability. Similarly, Ashraf et al.<sup>47</sup> introduces an enhanced metaheuristic approach for function optimization, offering an alternative strategy for fine-tuning machine learning models.

In recent years, significant findings from previous studies in AI-based methods have laid the groundwork for the development and adoption of metaheuristic methods in various fields, including chromatography and the prediction of mycotoxins. For instance, early AI models like SVM and neural networks demonstrated the ability to accurately predict complex chemical behaviors, including mycotoxin levels in food samples<sup>30</sup>. These models set a high standard for prediction accuracy but also revealed limitations in parameter optimization and overfitting<sup>30</sup>. Also, initial methods for optimizing model parameters, such as grid search and manual tuning, were time-consuming and often failed to find the global optimum, leading to suboptimal performance<sup>48</sup>.

Despite the fact that various scientific articles have reported about determination of mycotoxins using the chromatographic method coupled with the AI-based models<sup>49–54</sup>, however, only few focused on the use of diverse learning methods and optimization of network architecture<sup>55</sup>, such as HHO<sup>47</sup>, which enhance the accuracy of predictions by fine-tuning the parameters of the AI models, leading to more reliable and precise identification of mycotoxins, through reducing the need for extensive data pre-processing and large datasets, making the method more practical and cost-effective, by ultimately improving the detection and analysis of mycotoxins in food samples, and more importantly by improving the efficiency and effectiveness of the predictive models, optimization contributes to reducing the usage of toxic solvents and overall environmental impact, aligning with green chemistry principle. Also, more information on other techniques that can be used in improving the single model performance can be found<sup>22,56–67</sup>.

However, to the best of our knowledge, there is no published paper in the literature showing implementations of the recently evolved Harris Hawks Optimization (HHO) algorithm for the prediction of mycotoxins using chromatographic techniques. The primary goal of the current study is to optimize support vector regression (SVR) using nature-inspired algorithms to improve the predicted accuracy and efficiency of mycotoxin contamination modeling in food. By combining these optimization strategies, we want to enhance SVR's ability to analyze virtual water samples, which will ultimately lead to more accurate environmental risk assessments and food safety monitoring. Furthermore, this paper focuses on (1) the use of hybrid SVR with HHO and PSO (i.e., SVR-HHO and SVR-PSO) to model the qualitative properties of different classes of mycotoxins from food samples using the HPLC technique. (2) Compare the feasibility of SVR with PSO and HHO (i.e., SVR-PSO) and SVR-HHO for modelling the retention behaviour of mycotoxins using different input combinations. The environmental prediction of mycotoxins in food-virtual water samples is a critical area of research, given the significant health risks associated with mycotoxin contamination. Accurate prediction models are essential for effective monitoring and mitigation strategies. The current manuscript was structured into four different sections, whereby; section one demonstrates the introduction part of the study, while the second section presents the methodology part, which discussed in detail regarding the selected algorithms used in the current study, the third section depicts the results obtained based on the modelling process using various algorithm techniques as well as discuss the findings, whereas section four presents the conclusion of the study.

## Materials and method

### Proposed computational algorithms

#### Support vector regression (SVR)

Support vector machine (SVM) was introduced by Vapnik et al.<sup>68</sup>. The SVM is an approach to problem solving that combines pattern recognition, regression analysis, prediction, and classification. It is a concept of a machine learning. Hence, SVR is a form of SVM designed for regression task. There are two types of SVR; non-linear and linear<sup>69</sup>. The layer-based SVM known as SVR combines the function-weighted sum of the kernel outputs and the kernel function weighting on the input parameter. SVR aims to find a function that approximates the relationship between input features and continuous target values while minimizing prediction errors. Additionally, it offers a test error rate bound approximation<sup>70</sup>. SVR can be denoted using

$$f(x) = w \times \phi(x) + b \quad (1)$$

In other words,  $w$  represents the weight of the vector,  $\phi$  represents the transfer function, and  $b$  represents the bias. To visualize the SVR function  $f(x)$ , the regression problem can be represented as follows:

$$\text{Minimise } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (2)$$

$$\text{Subject to the condition: } \begin{cases} y_i - f(x) \leq \varepsilon + \xi_i \\ f(x) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad i = 1, 2, 3, \dots, N \end{cases} \quad (3)$$

where  $\varepsilon$  indicates the tube's size and shows the optimization performance,  $C$  indicates the variable, and  $\xi_i$  and  $\xi_i^*$  are the slack variables. Applying the Lagrangian functions in the following manner will yield the non-linear regression function's solution:

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (4)$$

In this case,  $C$ ,  $\alpha_i$ , and  $\alpha_i^* > 0$ , and  $K(x, x_i)$ ,  $\alpha_i$ , and  $\alpha_i^*$  are the kernel function and dual variables, correspondingly.

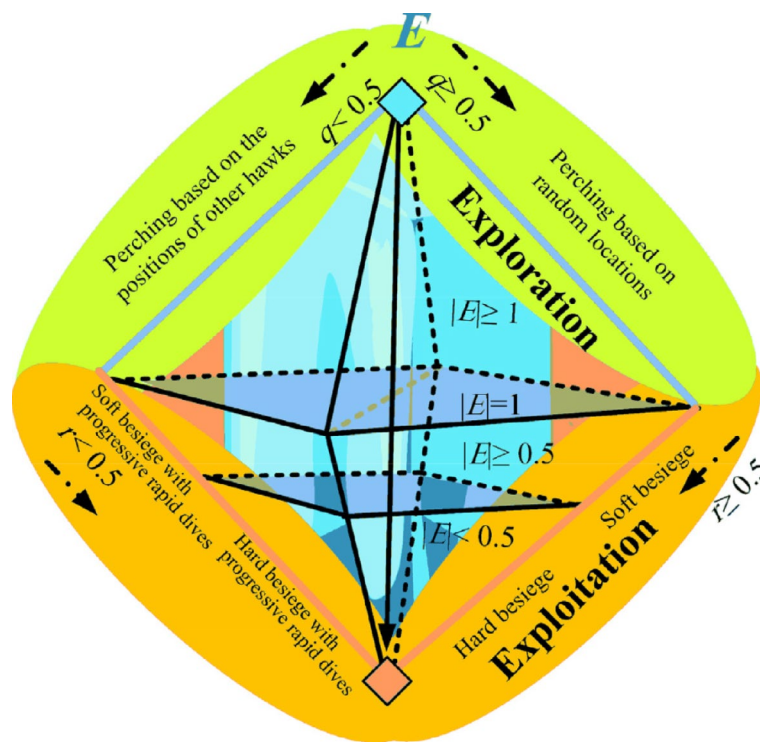
#### Harris Hawks optimization (HHO)

The Harris Hawks Optimization (HHO) algorithm is regarded as new since it was created by imitating the hawk's hunting technique. In recent years, the method has been effectively applied to resolve a variety of intricate problems in science and engineering<sup>71,72</sup>. Hawks used to hunt and pursue primarily on their own, whereas Harris hawks hunt in pairs. Thus, the HHO method is similar to the natural Harris hawks' cooperative mechanism and hunting style. Tracing, surrounding, approaching, and attacking are the mechanisms used in HHO hunting, which consist of three primary steps: investigation, exploitation, and a transition from exploitation to exploration (Fig. 1).

Exploration is the phase in which the algorithm searches the solution space broadly to identify potential locations. This phase of HHO replicates the activity of Harris hawks assessing their surroundings and scouting for prey. During exploration, the algorithm prioritizes diversifying the search to avoid early convergence to a local optimum. The hawks select how aggressively to explore based on the escape energy of the victim (rabbit). The escape energy ( $E$ ) is a factor that diminishes each repetition, balancing exploration with exploitation. Various tactics, such as Levy flight or random walks, are used to ensure that the search space is thoroughly covered.

The exploration phase is the initial stage and is illustrated as follows:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 X(t)| & \text{if } q \geq 0.5 \\ X_{rabbit}(t) - X_a(t) - r_3 (LB + r_4 (UB - LB)) & \text{if } q < 0.5 \end{cases} \quad (5)$$



**Fig. 1.** Phases of the Harris Hawks Optimization process<sup>73</sup>. Where  $r$  is a random number between 0 and 1, which is used to introduce randomness in the position updates of the hawks. It helps in diversifying the search process and is crucial for both exploration and exploitation phases.  $q$  is another random number between 0 and 1, which is used to determine the type of behavior (strategy) the hawks will adopt. It helps in deciding whether the hawks will perform a hard or soft besiege strategy.  $|E|$  is considered as the absolute value of the escape energy of the prey. This parameter decreases over iterations, starting from a positive value and gradually becoming negative. The value of  $|E|$  determines whether the algorithm is in the exploration phase or the exploitation phase. Hence, for  $|E| \geq 0.5$ : The hawks are in the exploration phase, meaning they are searching broadly across the solution space. While for  $|E| < 0.5$ : The hawks are in the exploitation phase, meaning they are focusing on fine-tuning the solutions around the best-found solutions.

$$X_a(t) = \frac{1}{N} \sum_1^N X_i(t) \quad (6)$$

The number of Harris eagles is represented by  $N$ ,  $X_a(t)$  is the average of eagles, and  $X(t+1)$  is the position of the eagle in subsequent iterations. LB and UB are the lower and upper equations, and  $q$ ,  $r_1$ ,  $r_2$ ,  $r_3$  and  $r_4$  are the differences between 0 and 1.  $t+1$ ,  $X_i(t)$  is the current position of the Harris Hawk in the iteration order  $t$ ,  $X_{\text{rand}}(t)$  is the randomly selected hawk.

If the hawks' energy is consumed as little as possible throughout the hunt, the second phase is regarded as a shift from investigation to exploitation. The following diagram illustrates the energy expended in evading the hunt:

$$E = 2E_0 \left(1 - \frac{t}{T}\right) \quad (7)$$

$T$  is the maximum number of iterations, and  $E_0$  is the initial energy used for each advancement ( $E_0 \in (-1, 1)$ ).

If we consider the third phase, the development phase focuses on the development of traditional solutions into new solutions. Based on the principle of animals escaping and eagles hunting, eagles quickly attack the species detected in the previous stage at this stage. Choose the type of enclosure according to the  $E$  and  $r$  values to catch the rabbit; Select hard when  $E < 0.5$  and select soft when  $E \leq 0.5$ .

Also, exploitation is the phase in which the algorithm intensifies its search for the best solutions discovered during exploration. This phase in HHO is similar to how Harris hawks encircle and dive into their prey. The goal here is to fine-tune the solutions to find the global optimum. Hawks descend on the prey (optimal solution) using a combination of soft and harsh besiege methods. In HHO, the balance between exploration and exploitation is dynamically regulated dependent on the prey's escape energy ( $E$ ). High escape energy encourages hawks to roam more broadly while low escape energy encourages exploitation by bringing hawks together and fine-tuning the search for promising answers.

The HHO algorithm can use four strategies: soft flanking, hard flanking, fast escalation, and rapid escalation to simulate the attack phase (Fig. 1).

#### Particle Swarm Optimization (PSO) algorithm

Kennedy and Eberhart<sup>74</sup> first developed this optimization method. It is based on population search algorithm, which is stimulated through the social movement and behavior dynamics of some animals<sup>74</sup>. The original idea behind the PSO concept was to provide a graphic representation of the social behavior of various animals, such as birds. The purpose of this is to identify the processes that allow the birds to fly in unison and quickly alter course by optimally regrouping<sup>75</sup>. The idea developed into an effective and straightforward optimization technique based on this notion. PSO refers to the population as a "swarm" and the individuals as "particles". To maintain track of its position inside the group, each member moves through the search space at a specific speed. One member of the swarm informs the others of this message<sup>76</sup>. The position of a particle in the search space is updated based on its current velocity and position<sup>77</sup>. The velocity is influenced by the particle's own experience and the experience of its neighbors. Generally, there are three steps in PSO optimization; (1) Initialization: which evaluate the fitness of each particle and initialize the personal best position and global best position. (2) The iteration stage: which evaluate the new fitness of each particle and (3) Termination: Repeat the iteration step until a stopping criterion is met (e.g., a maximum number of iterations or a satisfactory fitness level)<sup>78</sup>.

#### Data normalization

Equation (8) is used to normalize the data used in the current data<sup>79</sup>. One of the primary purposes of data normalization before AI modelling is to avoid the use of higher numerical errors and data redundancy<sup>80</sup>.

$$y = 0.05 + \left(0.95 \times \left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right)\right) \quad (8)$$

#### Model evaluation criteria

Any data-driven strategy compares the experimental outcome with the projected results to verify the performance criteria using various indices<sup>81–89</sup>. According to Legates and McCabe<sup>90</sup>, and Elkiran et al.<sup>91</sup>, the efficiency performance of any data intelligence model should comprise at least one goodness-of-fit (e.g., NSE) and at least one absolute error measure (e.g., RMSE). The use of five performance criteria in this study was due to the fact that multi-criteria indicators for measuring model performance are commonly used in contemporary studies. Another key reason for employing several criteria is that data qualities such as normalcy, size, and linearity influence the performance accuracy of any model that may be evaluated using these criteria<sup>90,91</sup>.

Three metrics were used in determining the performance of the AI based model established during the modelling phases. The correlation co-efficient (CC), mean square error (MSE), and Nash–Sutcliffe efficiency (NSE).

Mean square error (MSE) measures the average of the squares of the errors, which are the differences between observed and predicted values. It is a common measure of the accuracy of a predictive model.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (Y_{\text{obsi}} - Y_{\text{comi}})^2 \quad (10)$$



NSE measures how well the model predictions match the observed data compared to the mean of the observed data, with values closer to 1 indicating better performance.

$$NSE = \frac{\sum_{i=1}^N (Y_{obs} - \bar{Y}_{obs}) (Y_{com} - \bar{Y}_{com})}{\sqrt{\sum_{i=1}^N (Y_{obs} - \bar{Y}_{obs})^2 \sum_{i=1}^N (Y_{com} - \bar{Y}_{com})^2}} \quad (11)$$

The correlation coefficient (CC), measures the strength and direction of the linear relationship between two variables. For two datasets, observed values (obs,*j*) and computed values (com,*j*).

$$CC = 1 - \frac{\sum_{j=1}^N [(Y)_{obs,j} - (Y)_{com,j}]^2}{\sum_{j=1}^N [(Y)_{obs,j} - (\bar{Y})_{obs,j}]^2} \quad (12)$$

### Model validation

Generally, in deep learning and machine learning approaches, one of the primary objectives is to ensure that the models fit with good performance based on the indicators employed, in order to obtain a solid and trustworthy simulation of the test dataset<sup>29,89,92–100</sup>. However, problems such as local minima and overfitting necessitate the validation of the dataset<sup>101–105</sup>. Because of this, the training data's performance might not be enough, particularly if a smaller dataset is being used for the analysis. A variety of validation approach, including holdout, leave-one-out, and cross-validation (also known as k-fold), can be used. The k-fold approach was used in this work to prevent issues like overfitting. There are two phases to the data used in this study: training (70%) and testing (30%). More information about model validation can be found in<sup>49,85,98,106,107</sup>.

### Proposed modelling techniques

It is important to precisely define the model's parameters in order to improve SVR's performance accuracy. Selecting different suitable conditions to reach the optimal state determines how accurate it is. Finding the worldwide ideal solution is crucial since it can help you achieve the finest results. As a result, the SVR-PSO and SVR-HHO hybrid methods were created by combining the SVR approach with nature-inspired algorithms (PSO and HHO). The three main SVR model parameters (C,  $\gamma$ , and  $\epsilon$ ) were found using techniques inspired by nature. The flowchart depicts the hybrid SVR-PSO and SVR-HHO models utilized in this work based on Fig. 3.

Furthermore, the complexity of SVM algorithms depends on several factors, including the type of SVM (linear or kernelized), the optimization algorithm used to train the SVM, the number of training samples (N), training complexity, space complexity and the number of features (D) in the dataset.

The data involved in this study was derived from a research conducted by Nielsen and Smedsgaard<sup>108</sup>. Prior to dwelling into the modelling step various pre-processing analysis were conducted including; handling missing data, data cleaning, feature engineering, outlier detection and handling. Nowadays, the novel AI-based methods (both hybrid and simple models) are now becoming interestingly popular in the field of mycotoxins in classifications as well as regression for determining their chemical structure as well as grouping them into various respective classes based on their features and properties demonstrated<sup>109</sup>. These models are equally gaining popularity in the areas of chromatography in determining the qualitative and quantitative properties of various chemical compounds using their respective retention properties and peak areas as well as in other field of spectroscopy and spectrometry such as mass spectrometry in predicting and classifying various compounds based on their peak areas and mass-to-charge ratio<sup>110</sup>. Using the retention index, mono isotopic mass, relative sensitivity factor, and peak symmetry as the independent variables for each associated mycotoxin, the retention time as the dependent variable was predicted. To simulate the retention time of the associated mycotoxins, 150 data points from seven different mycotoxin groups were used.

More also, the overall methodology flowchart is demonstrated in Fig. 2.

Various researchers such as Rajaei et al.<sup>111</sup> depicts that 70–30% usually presents better performance accuracy. But nevertheless, the current study checked for other partitioning such as 60–40%, 65–35%, 80–20% and 90–10%. However, 70–30% not only showed higher performance but equally solve the major modelling difficulties such as overfitting and under fitting issues. Hence, 70% was utilized as the training data and 30% as the testing data. The performance of the models was checked using the testing data set to provide an unbiased determination of the models' accuracy. Besides, various part of the datasets was employed for validation and training of the chromatographic behaviour of the mycotoxins.

### Results and discussions

The main objective of this study is to develop an AI-based model (SVR) and hybridize it with two different metaheuristic algorithms, namely HHO and PSO, in order to find the optimum values of the SVR parameters, which are used to accurately model the chromatographic behaviour of various classes of mycotoxins in food samples. By fine-tuning these parameters, the AI-based model can provide more precise simulations, leading to better predictions and analyses when using the HPLC technique as illustrated in Fig. 3. In this section, the outcomes obtained are established both in visualized and quantitative format. Table 1 demonstrates a pre-analysis of the raw data for different classes of mycotoxins using both statistical and correlation analysis. Various methods, including modeling, are used for preliminary data before predictions are made.

Based on Table 1, it can be observed that the retention index has the highest correlation with the dependent variable (retention time (tR)), having an R-value of 0.8826, followed by peak symmetry with an R-value of 0.7210. On the other hand, the mono-isotopic mass has the lowest correlation result with an R-value of –0.0924.

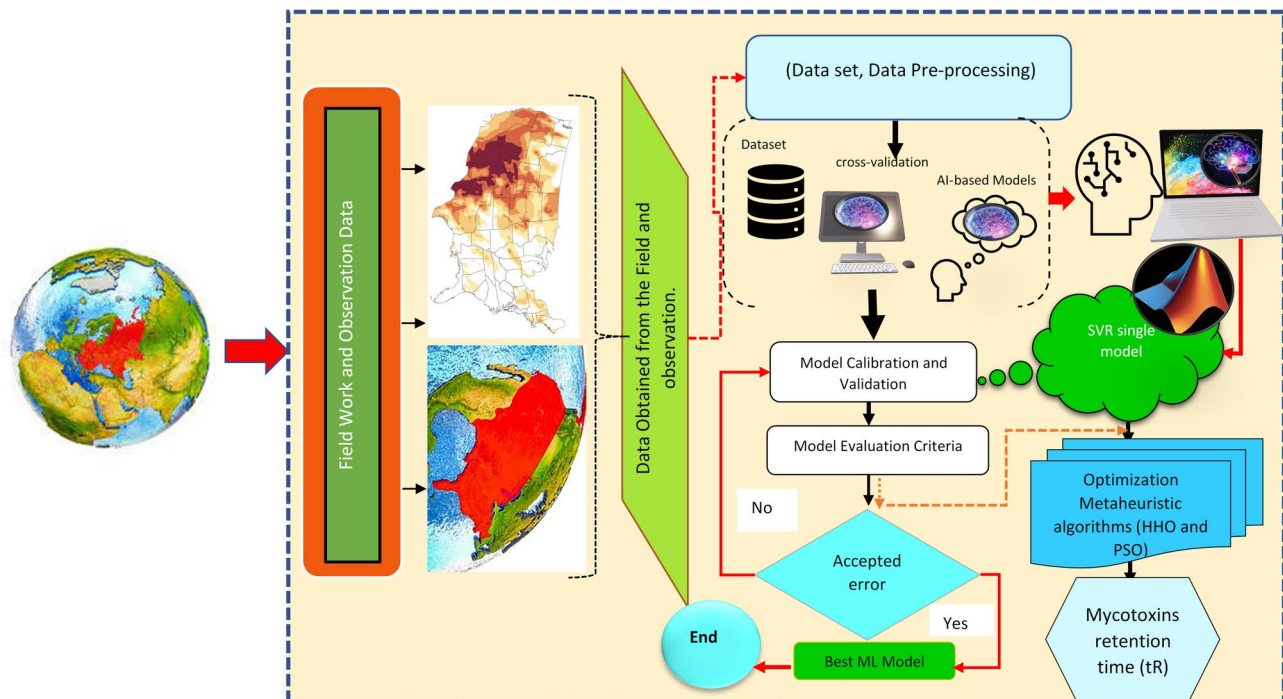


Fig. 2. Overall methodology flowchart.

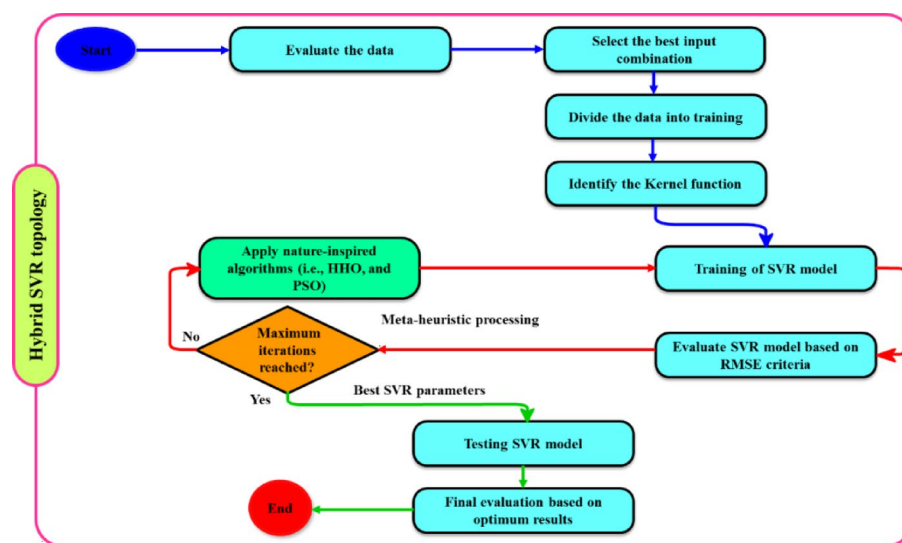


Fig. 3. Hybrid support vector regression models flowchart.

Variables	RI	PS	MIM	RSF	RT
Retention index	1				
Peak symmetry	0.0082	1.0000			
Mono isotopic mass	0.2658	−0.0611	1.0000		
Relative sensitivity factor	−0.6140	−0.1604	−0.1918	1.0000	
Retention time (tR)	0.8826	0.7210	−0.0924	0.2901	1

Table 1. Correlation analysis of the experimental variables.

Model type	Model combination
M1	MIM + RSF
M2	MIM + PS + RSF
M3	RI + PS + RSF + MIM

**Table 2.** Input variable selection.

	Training			Testing		
	NSE	CC	MSE	NSE	CC	MSE
SVR-M1	0.48	0.68	0.03	0.42	0.65	0.007
SVR-M2	0.47	0.69	0.03	0.56	0.75	0.006
SVR-M3	<b>0.97</b>	<b>0.98</b>	<b>0.002</b>	<b>0.94</b>	<b>0.97</b>	<b>0.001</b>
SVR-HHO-M1	0.78	0.88	0.01	0.73	0.85	0.004
SVR-HHO-M2	0.73	0.86	0.01	0.78	0.88	0.003
SVR-HHO-M3	<b>0.98</b>	<b>0.99</b>	<b>0.001</b>	<b>0.99</b>	<b>0.995</b>	<b>0.000137</b>
SVR-PSO-M1	0.70	0.83	0.015	0.68	0.82	0.004
SVR-PSO-M2	0.72	0.85	0.014	0.75	0.87	0.003
SVR-PSO-M3	<b>0.91</b>	<b>0.95</b>	<b>0.005</b>	<b>0.90</b>	<b>0.95</b>	<b>0.001</b>

**Table 3.** Results of evaluation criteria for SVR, SVR-HHO and SVR-PSO model. NB: Bold indicates the best performing model combination.

Furthermore, the retention index, as the dominant input variables, has the highest skewness value and, as seen in Table 1, the highest correlation. Furthermore, the skewness values of all the other input variables were greater and better. As indicated in Table 1, this has an impact on the input and output parameter correlation analyses, which in turn influences the modeling.

The understanding between the variables, which support the formation of three distinct model types, were further evaluated using the correlation analysis approach (inform of models M1, M2 and M3). Table 1 presents the assurance of the correlation coefficient conducted in order to improve the models' methodology. The guiding symbols (+ or -) indicate the relationship between the variables. We can identify the highest and lowest variables that are related to one another with the use of correlation analysis. Furthermore, by displaying the input parameter that has the highest correlation with the output variable, it aids in understanding the science and mechanism of the data before delving into the modeling method. This is also beneficial for experimental analysis, particularly for the optimization process. As a result, different models were created based on the correlation analysis in order to actually see how the method affected the simulation process. Table 3 displays these models.

**Results of metaheuristic algorithms and machine learning**

MATLAB 9.3 (R2020a) was used for the modelling, as mentioned in the literature section. Metaheuristic methods such as PSO and HHO were used to optimize the structure of the SVR model. NSE and CC were used to assess the simulation's findings in terms of the fitness among the actual and forecast values, and MSE was used to ascertain the error that the models in both the training and testing phases represented. Based on the model combinations in Table 2, the simulated outcome are quantitatively presented in Table 3. Table 3 exhibits the multi-model information regarding the performance efficiency of the models using various input variable selection in both the training and testing stages. For modelling mycotoxins from food samples using SVR technique the predictive approaches showed different efficiencies in accordance with the evaluation metrics. In general, for the single AI-based model (inform of SVR) SVR-M3 with four input variables depict higher performance based on NSE, CC and MSE. Although it is difficult to rank the model based on exact accuracy, the SVR-M3 method also showed the topmost prediction accuracy for modelling the mycotoxins retention behaviour with more than 90% performance efficiency in both the calibration and verification stages.

Regardless of the nonlinear association among the predictors and their corresponding targets, the overall accuracy of the SVR is inadequate, particularly in M1 and M2. By creating a hybrid technique and using the optimization algorithms (HHO and PSO), this could be improved. Essentially, it is important to remember that the hopeful estimations happened during the calibration phase, which is largely used to successfully calibrate the models using targets and known input variables. However, as testing evaluates the model's accuracy based on unknown goal values, it is essential to evaluating a model's performance. The training set does not includes this benefit. In light of this, the trustworthy model ought to operate steadily and evenly during the calibration and verification stages. Regarding the single model, the section on optimization algorithms typically showed promising abilities. This is not at all surprising, as numerous works of literature demonstrate<sup>71,73,112–115</sup>. Just like in the case of the SVR model, the performance simulation of the metaheuristic's technique was equally evaluated using NSE, CC and MSE. Based on Table 3 SVR-HHO-M3 showed the highest NSE value=0.990096 and the lowest MSE value=0.000137 in the verification stage. Furthermore, the results proved the ability of SVR-HHO over the single SVR model as well as the SVR-PSO metaheuristic approach for the simulation of mycotoxins in



food samples using the chromatographic technique. The predictive ability of HHO over others could be seen in various recent literature for example<sup>71,72,115</sup>.

Moreover, the optimization capacity of HHO on SVR technique in the current study is in line with ability of GSK algorithm in optimizing the structure of various techniques to provide higher accuracy. For instance, Muhammad et al.<sup>82</sup> proposes the application of GSK in solving binary optimization problems through a novel technique called novel binary Gaining Sharing knowledge-based optimization algorithm (NBGSK) that enables exploitation and exploring of the search space effectively and efficiently for solving problems in the binary space. Additionally, in order to keep the solutions from trapping into local optima, which progressively reduce the population size via a linear function, the population size reduction (PR-NBGSK) is implemented. When applied to a set of knapsack examples of both small and big dimensions, the suggested NBGSK and PR-NBGSK demonstrate their superior efficacy and efficiency in terms of convergence, robustness, and accuracy. In general, six issues in 10D, four in 30D, three in 50D, and one in 100D were successfully solved at least once by the GSK performance result. In 51 runs across 10 and 30 dimensions, GSK consistently finds the global optimal solution, demonstrating exceptional performance on unimodal issues (f1–f3). The optimum in 50 dimensions is found in 1 case; the standard deviation and mean error vary from 1.24E+03 to 1.51E+03, and from 1.09E+03 to 3.85E+03. The standard deviation in 100 dimensions is 4.63E+03 to 2.15E+04, whereas the mean error is 5.80E+03 to 1.15E+05. As a result, the inaccuracy that their data show is generally smaller than the error that HHO produced in our investigation. Hence, even though it was applied to different datasets, demonstrated a good potential over the HHO model.

Furthermore, Zhong et al.<sup>116</sup> presented the first hybridization of DE with HHO and GSK (DEGH), which was used as bench mark over 32 other techniques, more also, experimental results equally showed that the proposed DEGH algorithm is significantly superior to the compared algorithms. Hence, the performance results based on the goodness-of-fit demonstrates that FS-BGSK, FS-pBGSK both demonstrated performance accuracy of 1.00 at M6. This indicates that GSK has the ability of improving the performance of HHO with 0.98 and 0.99 goodness-of-fit in the training and testing steps respectively. More information regarding GSK can be found at<sup>117</sup>.

Moreover, Khosrokhavar et al. reported utilizing data from liquid chromatography with UV and MS detector to predict the retention behaviour of various mycotoxins by the application of SVM and MLR approaches. The performance results inform of correlation and predictability measure by  $R^2$  and  $q^2$  are 0.931 and 0.932, respectively, for SVM and 0.923 and 0.915, respectively, for MLR<sup>30</sup>. This indicates that the performance of SVR-M3 (0.97 and 0.94) and SVR-HHO-M3 (0.98 and 0.99) depicted in the training and testing phases of the current study presented higher performance than that<sup>30</sup>.

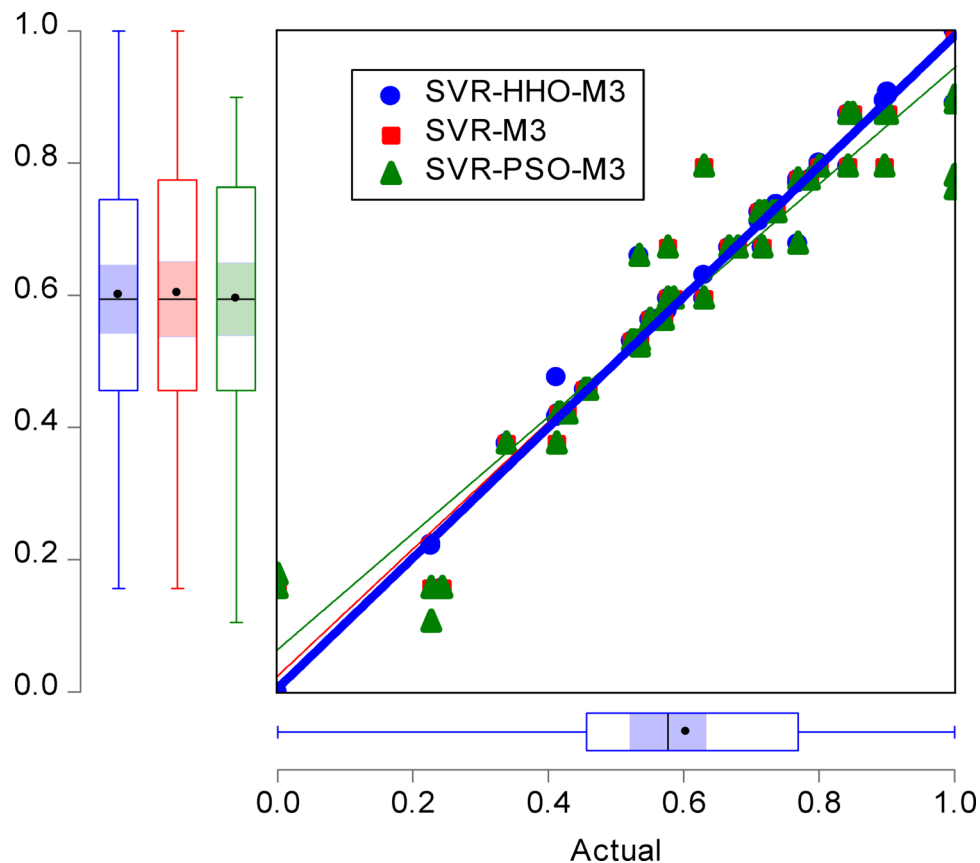
Furthermore, Gilandeh et al. demonstrated the use of image processing to identify and distinguish between different wheat grains that are contaminated with a mycotoxin known as *F. graminearum*. They did this by utilizing SVM and DA algorithms<sup>118</sup>. The obtained results showed that classification using SVM method is better than non-linear SVM, having a 100% performance accuracy. This demonstrates that their performance accuracy is higher than the one obtained in the current study. Even though, the nature of the data and analysis differs, in which their analysis involves image processing techniques that involves the use of images as the datasets, while in our study, we employ the use of regression techniques using numerical dataset.

Additionally, Ge et al. reported on a comparison performance analysis between MLR and ANN models for cheese quality prediction. Using the principle component regression (PCR), partial least squares (PLS), SVM, and principal component analysis (PCA-SVM) techniques, linear and nonlinear regression models are built to connect the absorption spectra and the concentrations of 160 samples. Hence, the techniques depict the following RMSE-values for PLS (0.753, 0.691), PCR (0.587, 0.643), SVM (1.365, 1.674) and PCA-SVM (1.864, 1.953) in the training and testing phases respectively. Whereby, the outcomes of our study depict a range of MSE-values of 0.001 to 0.005 and 0.000137 to 0.001 in the training and testing stages respectively<sup>119</sup>. Hence, this indicates that the performance of our study outperformed the previous study with an outstanding performance.

In general, the single AI-based model in the form of SVR, specifically SVR-M3 with four input variables, shows higher performance based on NSE, CC, and MSE. Despite the difficulty in ranking the models according to the achieved accuracies, the SVR-M3 approach relatively showed the best prediction accuracy for modeling the mycotoxins retention behavior, with more than 90% performance efficiency in both the testing and training stages.

However, the overall accuracy of the SVR is unsatisfactory, particularly in M1 and M2, regardless of the nonlinear relationship between predictors and their corresponding targets. This could be improved by developing a hybrid technique through the application of optimization algorithms (HHO and PSO). It should be considered that the promising estimations occurred during the calibration phase, which is primarily employed to effectively calibrate the models based on known input variables and targets. However, the testing step is vital in assessing the performance of a model, as it examines the model's accuracy based on unseen target values. This benefit does not exist in the training set. Therefore, a reliable model should have stable and balanced performance in both the calibration and verification phases. The section on optimization algorithms generally depicted promising ability compared to the single model, which is not surprising as seen in much of the literature<sup>71,73,112–115</sup>. Just like in the case of the SVR model, the performance simulation of the metaheuristic's techniques was equally evaluated using NSE, CC and MSE. Based on Table 3 SVR-HHO-M3 showed the highest NSE value = 0.990096 and the lowest MSE value = 0.000137 in the verification stage. Furthermore, the results proved the ability of SVR-HHO over the single SVR model as well as the SVR-PSO metaheuristic approach for the simulation of mycotoxins in food samples using the chromatographic technique. The predictive ability of HHO over others could be seen in various recent literature for example<sup>71,72,115</sup>.

Nevertheless, the comparative predictive performance of the best model combinations (i.e., SVR-M3, SVR-HHO and SVR-PSO) can visualized and compared using the scatter plots (see Fig. 4). This plot usually indicated the level of arrangement among the actual and simulated values for the overall goodness-of-fit, which is indicated



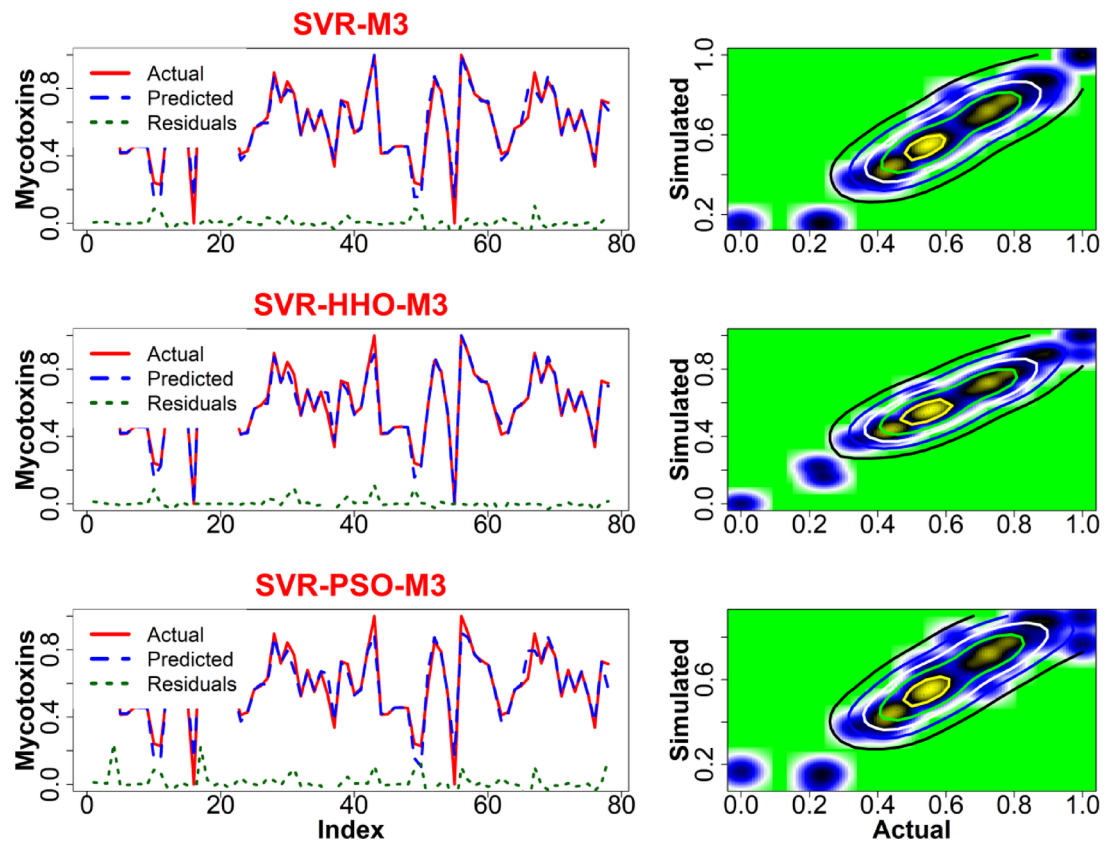
**Fig. 4.** Graphical representation for comparative analysis of the observed and computed values of the best model input combinations.

based on the determination co-efficient value ( $R^2$ ), the higher the  $R^2$  the better the performance of the model and vice-versa. More also, the regression equation  $y = mx + c$ , whereby  $m$  is the slope,  $c$  is the intercept and  $y$  and  $x$  are unknown can be used in determining a line of best fit. It can be used to forecast the  $x$  and  $y$  variables' outcomes in a given data set or sample data and is represented as a scatter plot. Although there are several methods for determining a regression line, the least-squares regression line is typically employed since it yields a uniform line<sup>120</sup>. It is evident from the plot that SVR-HHO-M3 showed higher prediction efficiency as compared with SVR-PSO-M3 and SVR-M3 models.

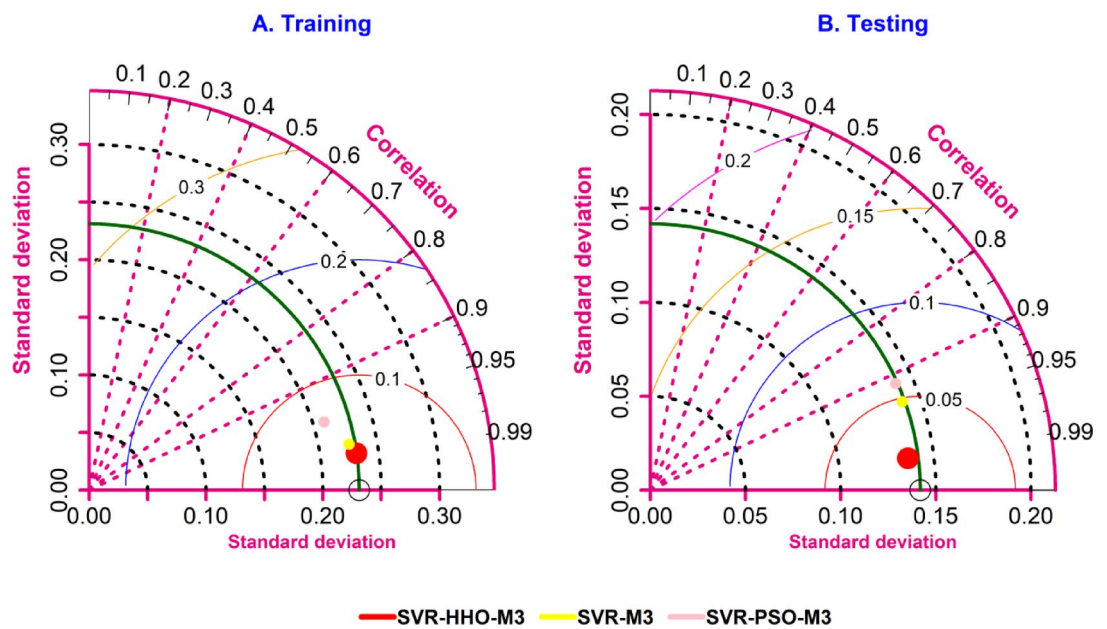
The relative analysis of the models can be better envisaged using the time series plot imbedded with contour plot (see Fig. 5). The time series plot is a robust graphical depiction of data that gives an overview and a numerical summary of a data set.

Furthermore, the overall comparative performance of the topmost AI-based model and the hybrid techniques (SVR-M3, SVR-PSO-M3 and SVR-HHO) is shown through the recently used visualized two-dimensional plot known as the Taylor diagram, as illustrated in Fig. 6. This diagram is generally employed to summarize and highlights various statistical indices such as correlation co-efficient (CC), standard deviation (std.), root mean square error (RMSE), determination co-efficient (DC) and mean absolute error (MAE) between the experimental and the simulated/predicted values. The Taylor diagram has been used in different areas such as engineering, water analysis and rainfall forecasting etc. due to its paramount statistical applications. Remarkably, to best of the author's knowledge this research serves as the first to use this diagram in modelling mycotoxins from food samples. The Taylor diagram is discussed in further detail in<sup>121</sup>. Based on Fig. 6, the SVR-HHO-M3 showed higher goodness-of-fit for the prediction of mycotoxins from food samples using chromatographic method in both the training and testing stages. The obtained results lead to the conclusion that SVR-HHO-M3, SVR-PSO-M3, and SVR-M3 can capture the complex non-linear behaviour of the data involved in this article.

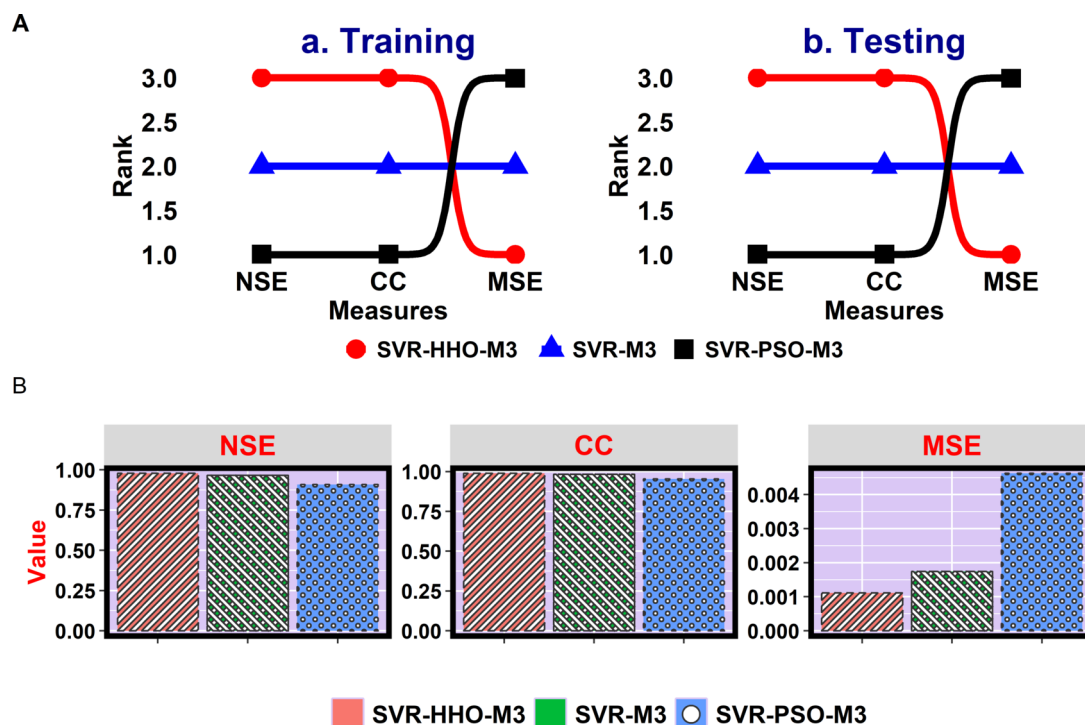
Furthermore, the predictive performance of the metaheuristic algorithms can also be presented using a bump chart coupled with a bar plot (see Fig. 7). A bump chart is a type of chart used to visualize changes in rank over time. It's particularly useful for showing the movement or ranking of different categories or items across multiple periods. Therefore, the ranking given in Fig. 7 for the bump plot includes three different metrics (NSE, MSE, and CC) for the three models (SVR-HHO-M3, SVR-M3, and SVR-POS-M3) in both the training and testing stages respectively. Hence, the illustrations depict the power of SVR-HHO over SVR-POS and conventional SVR techniques in modelling the mycotoxins. Additionally, Fig. 7B illustrates the respective performance metrics of the models using NSE, CC and MSE inform of bar plot. Overall, SVR-HHO presents the highest predictive performance.



**Fig. 5.** Time series for both the actual and computed values for the mycotoxin's determination using chromatographic technique.



**Fig. 6.** Two-dimensional representation showing the models' calibration and verification in terms of the Taylor diagram.



**Fig. 7.** (A) Bump chart and (B) bar plot for the performance metrics.

In general, the findings show that the SVR-M3 model, when hybridized with optimization algorithms (HHO and PSO), achieves over 90% performance efficiency in both training and testing stages. This high level of accuracy is meaningful as it indicates that the model can reliably predict the retention behavior of mycotoxins, which is crucial for ensuring food safety.

The suggested SVR-HHO framework is a good choice for chromatographic retention time modeling since it strikes a compromise between computational cost and prediction accuracy. However, memory and processor cost must be taken into account for large-scale deployment in analytical chemistry labs. Although SVR is still comparatively light in comparison to deep learning models, which require a lot of processing power, the addition of HHO for hyperparameter tuning introduces an optimization layer that raises the computational cost. For moderately small datasets, this is still feasible, but in order to preserve efficiency when scaling to high-dimensional chromatographic data with thousands of samples, parallel processing or GPU acceleration may be needed. One interesting approach to real-time chromatographic analysis is the deployment of SVR-HHO on cloud-based or edge computing platforms. Large-scale data processing and on-demand model retraining are made possible by cloud computing, guaranteeing flexibility in response to changing analytical circumstances. On the other side, edge computing minimizes latency and makes it possible to estimate retention durations almost instantly by enabling localized inference at chromatographic workstations. Optimizing model compression methods and utilizing hardware accelerators to shorten inference times High-throughput sample analysis in industrial and regulatory labs may also be made easier by combining the SVR-HHO model with automated liquid handling devices. Labs might improve operational efficiency, decrease manual involvement, and optimize chromatographic procedures by automating retention time predictions. Future studies might concentrate on creating an end-to-end pipeline for chromatographic quality control systems that combines automated decision-making, AI-driven predictions, and real-time data collecting. Without sacrificing accuracy would be necessary for such implementations to be feasible.

The SVR-HHO model can be successfully applied to food safety regulatory systems, such as FDA guidelines or hazard analysis and critical control points (HACCP), by improving the predictive capabilities for contamination risks and supporting decision-making processes through supply chain risk management, risk-based monitoring, automated alerts, optimized mitigation strategies, improved surveillance programs, etc. Also, using domain-specific datasets and altering input features, the SVR-HHO model can be expanded to detect pesticides, heavy metals, and other pollutants. Furthermore, by altering input characteristics and training the model with a variety of contamination datasets, the SVR-HHO model can generalize across different food matrices, including grains, dairy, and drinks. Custom feature selection and matrix-specific contamination data training are necessary for the model to adapt to various food matrices. SVR can enhance performance in these many categories by using Harris Hawks Optimization (HHO) to optimize the hyperparameters. This integration ensures effective, scalable, and internationally harmonized regulatory compliance by providing a next-generation AI-driven solution to food safety monitoring.



## Conclusion

A multi-model data-driven method using an AI-based model (SVR) with two metaheuristic methods, comprising PSO and HHO (SVR-PSO and SVR-HHO), was used to predict the chromatographic behavior of various classes of mycotoxins in food samples. This approach aims to qualitatively elucidate this type of fungi and develop a mechanism for its prevention. The M3 model combination showed the best predictions in both SVR, SVR-HHO and SVR-PSO, with SVR-HHO having the highest prediction ability. Also, the results indicate that the main objective of the study is the implementation of SVR-HHO (NSE = 0.98, CC = 0.99 and MSE = 0.001 in the training and NSE = 0.99, CC = 0.995 and MSE = 0.000137 in the testing phase), which can enhance the performance of previous methods by 4–7% in the training and testing phases, respectively, in predicting the qualitative characteristics of mycotoxins in food samples. The integration of SVR with metaheuristic optimization enhances predictive accuracy and efficiency, making it suitable for deployment in real-world chromatographic laboratories. By automating hyperparameter tuning, the proposed framework can be integrated into regulatory and commercial food safety programs, improving contaminant detection and compliance monitoring.

This limited scope means that other potentially effective models and hybrid combinations were not explored, model generalization across different food matrices, computational demands, and integration with existing chromatographic software which could have yielded different results. More also, developing and running hybrid AI models, especially with optimization algorithms, requires significant computational resources. This limitation might affect the feasibility of applying these models in environments with limited computational capabilities.

The findings also imply that additional hybrid models, including, ensemble machine learning strategies, and other metaheuristic algorithms, might be used to enhance the capacity to predict the chromatographic characteristics of the mycotoxins. Investigate other AI and machine learning models such as random forest, gradient boosting, neural networks, AutoML and other ensemble methods. Comparing a wider range of models can help identify the most effective approaches for predicting mycotoxins' retention behaviour.

## Data availability

The supported data associated with this research are available upon request from the corresponding author.

Received: 17 January 2025; Accepted: 23 April 2025

Published online: 13 May 2025

## References

- Magan, N. & Olsen, M. *Mycotoxins in Food: Detection and Control* (Woodhead Publishing, 2004). <https://doi.org/10.1533/9781855739086>.
- Vega, A. et al. Comparative study of dehydrated papaya (*Vasconcellea pubescens*) by different drying methods: Quality attributes and effects on cells viability. *J. Food Meas. Charact.* <https://doi.org/10.1007/s11694-021-00845-6> (2021).
- Arroyo-Manzanares, N., Gámiz-Gracia, L. & García-Campaña, A. M. Determination of ochratoxin A in wines by capillary liquid chromatography with laser induced fluorescence detection using dispersive liquid–liquid microextraction. *Food Chem.* **135**, 368–372 (2012).
- Mansilha, C. et al. Quantification of endocrine disruptors and pesticides in water by gas chromatography–tandem mass spectrometry. Method validation using weighted linear regression schemes. *J. Chromatogr. A* **1217**, 6681–6691 (2010).
- Amini, I. & Pal, K. Prediction of two-dimensional gas chromatography time-of-flight mass spectrometry retention times of 160 pesticides and 25 environmental organic pollutants in grape by multivariate chemometrics methods. *Adv. J. Chem. A* **1**, 12–31 (2018).
- Seymour Shephard, G. Determination of mycotoxins in human foods. *Chem. Soc. Rev.* **37**, 2468–2477 (2008).
- Marengo, E., Gianotti, V., Angioi, S. & Gennaro, M. C. Optimization by experimental design and artificial neural networks of the ion-interaction reversed-phase liquid chromatographic separation of twenty cosmetic preservatives. *J. Chromatogr. A* **1029**, 57–65 (2004).
- Zeng, J., Chai, Q., Peng, X. & Li, S. Geographical origin identification for *Tetrastigma hemsleyanum* based on high performance liquid chromatographic fingerprint. In *Proceedings of the 2019 Chinese Automation Congress CAC 2019*, 1816–1820 (2019). <https://doi.org/10.1109/CAC48633.2019.8996240>.
- Xian-hui, L. et al. Determination of 10 kinds of caine-type prohibited ingredients in cosmetics by ultra-performance liquid chromatography–differential mobility spectrometry–mass spectrometry. *Chin. J. Anal. Chem.* **47**, 756–764 (2019).
- Khader, M. et al. Meningkatkan aktivitas fagositosis makrofag tikus betina galur SD (Sprague Dawley) Yang Diinduksi DMBA (7, 12dimetilbenz (α) Antrasen) Secara in vitro the effects of hexane extracts from black cumin seeds to the phagocytic activity of macrophages S. *Molecules* **19**, 1–19 (2017).
- Silva Gracia, M., Köppl, A., Unholzer, S. & Haen, E. Development and validation of an HPLC–UV method for the simultaneous determination of the antipsychotics clozapine, olanzapine and quetiapine, several beta-blockers and their metabolites. *Biomed. Chromatogr.* **31**, 1–11 (2017).
- Markiewicz, B., Komorowicz, I., Sajnog, A., Belter, M. & Baralkiewicz, D. Chromium and its speciation in water samples by HPLC/ICP–MS—Technique establishing metrological traceability: A review since 2000. *Talanta* **132**, 814–828 (2015).
- Zhang, Z., Hu, X., Zhang, Q. & Li, P. Determination for multiple mycotoxins in agricultural products using HPLC–MS/MS via a multiple antibody immunoaffinity column. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **1021**, 145–152 (2016).
- Marume, P., Ndlovu, N., Manhokwe, S. & Ndlovu, C. Detection of mycotoxins in opaque beer production. *Afr. J. Biol. Sci.* **2**, 84–91 (2020).
- Pakshir, K. et al. Mycotoxins detection and fungal contamination in black and green tea by HPLC-based method. *J. Toxicol.* **2020**, 2456210 (2020).
- Agriopoulou, S., Stamatelopoulou, E. & Varzakas, T. Advances in analysis and detection of major mycotoxins in foods. *Foods* **9**, 1–23 (2020).
- Abba, S. I. & Usman, A. G. Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach. *Chemom. Intell. Lab. Syst.* **201**, 104007 (2020).
- Selin, A. G. U. & Abba, I. S. I. A novel multi-model data-driven ensemble technique for the prediction of retention factor in HPLC method development. *Chromatographia* <https://doi.org/10.1007/s10337-020-03912-0> (2020).
- Usman, A. G. et al. Evidential neural network model for groundwater salinization simulation: A first application in hydro-environmental engineering. *Water* **16**, 2873 (2024).



20. Ghali, U. M. et al. Applications of artificial intelligence-based models and multi-linear regression for the prediction of thyroid stimulating hormone level in the human body. *Int. J. Adv. Sci. Technol.* **29**, 3690–3699 (2020).
21. Usman, J. et al. Stabilized oily-wastewater separation based on superhydrophilic and underwater superoleophobic ceramic membranes: Integrated experimental design and standalone machine learning algorithms. *J. Taiwan Inst. Chem. Eng.* **164**, 105704 (2024).
22. Dogara, M., Usman, A. A. G., Ozsahin, D. U., Umar, A. & Abba, I. S. I. Novel ensemble machine learning paradigms for the prediction of antioxidant activity of *Bryophyllum pinnatum* (Lam.) Oken. *Proc. Natl. Acad. Sci. India Sect. B Biol. Sci.* <https://doi.org/10.1007/s40011-024-01619-y> (2024).
23. Martins, C., Assunção, R., Nunes, C., Torres, D. & Alvito, P. Are data from mycotoxins' urinary biomarkers and food surveys linked? A review underneath risk assessment. *Food Rev. Int.* **00**, 1–26 (2020).
24. El-kenawy, E.-S.M. et al. Greylag goose optimization: Nature-inspired optimization algorithm. *Expert Syst. Appl.* **238**, 122147 (2024).
25. Ahmed, M. et al. An AI-based system for predicting renewable energy power output using advanced optimization algorithms. *J. Artif. Intell. Metaheuristics* **8**, 01–08 (2024).
26. El-Sayed, E., Eid, M. M. & Abualigah, L. Machine learning in public health forecasting and monitoring the Zika virus. *Metaheuristic Optim. Rev.* **1**, 01–11 (2024).
27. Eed, M. et al. Potato consumption forecasting based on a hybrid stacked deep learning model. *Potato Res.* <https://doi.org/10.1007/s11540-024-09764-7> (2024).
28. Radwan, M., Alhussan, A. A., Ibrahim, A. & Tawfeek, S. M. Potato leaf disease classification using optimized machine learning models and feature selection techniques. *Potato Res.* <https://doi.org/10.1007/s11540-024-09763-8> (2024).
29. Asnake Metekia, W., Garba Usman, A., Hatice Ulusoy, B., Isah Abba, S. & Chirkena Bali, K. Artificial intelligence-based approaches for modeling the effects of spirulina growth mediums on total phenolic compounds. *Saudi J. Biol. Sci.* <https://doi.org/10.1016/j.sjbs.2021.09.055> (2021).
30. Khosrokhavar, R., Ghasemi, J. B. & Shiri, F. 2D quantitative structure–property relationship study of mycotoxins by multiple linear regression and support vector machine. *Int. J. Mol. Sci.* **11**, 3052–3068 (2010).
31. Aljaidi, M. et al. Optimizing FACTS device placement using the fata morgana algorithm: A cost and power loss minimization approach in uncertain load scenario-based systems. *Int. J. Comput. Intell. Syst.* **18**, 12 (2025).
32. Lian, J. et al. The educational competition optimizer. *Int. J. Syst. Sci.* **55**, 3185–3222 (2024).
33. Su, H. et al. RIME: A physics-based optimization. *Neurocomputing* **532**, 183–214 (2023).
34. Yuan, C. et al. Polar lights optimizer: Algorithm and applications in image segmentation and feature selection. *Neurocomputing* **607**, 128427 (2024).
35. Ahmadianfar, I., Bozorg-Haddad, O. & Chu, X. Gradient-based optimizer: A new metaheuristic optimization algorithm. *Inf. Sci. (N.Y.)* **540**, 131–159 (2020).
36. Li, S., Chen, H., Wang, M., Heidari, A. A. & Mirjalili, S. Slime mould algorithm: A new method for stochastic optimization. *Futur. Gener. Comput. Syst.* **111**, 300–323 (2020).
37. Askari, Q., Saeed, M. & Younas, I. Heap-based optimizer inspired by corporate rank hierarchy for global optimization. *Expert Syst. Appl.* **161**, 113702 (2020).
38. Zhao, W., Wang, L. & Zhang, Z. *Artificial Ecosystem-Based Optimization: A Novel Nature-Inspired Meta-heuristic Algorithm*. *Neural Computing and Applications* Vol. 32 (Springer, 2020).
39. Shi, B. et al. Prediction of recurrent spontaneous abortion using evolutionary machine learning with joint self-adaptive sime mould algorithm. *Comput. Biol. Med.* **148**, 105885 (2022).
40. Fei, X., Wang, J., Ying, S., Hu, Z. & Shi, J. Projective parameter transfer based sparse multiple empirical kernel learning machine for diagnosis of brain disease. *Neurocomputing* **413**, 271–283 (2020).
41. Xu, J., Zhang, Y. & Miao, D. Three-way confusion matrix for classification: A measure driven view. *Inf. Sci. (N.Y.)* **507**, 772–794 (2020).
42. Alkhamash, E. H. et al. Application of machine learning to predict COVID-19 spread via an optimized BPSO model. *Biomimetics* **8**, 457 (2023).
43. Fouad, Y., Osman, A. M., Hassan, S. A. Z., El-Bakry, H. M. & Elshewey, A. M. Adaptive visual sentiment prediction model based on event concepts and object detection techniques in social media. *Int. J. Adv. Comput. Sci. Appl.* **14**, 252–256 (2023).
44. Alkhamash, E. H., Kamel, A. F., Al-Fattah, S. M. & Elshewey, A. M. Optimized multivariate adaptive regression splines for predicting crude oil demand in Saudi Arabia. *Discrete Dyn. Nat. Soc.* **2022**, 8412895 (2022).
45. Alzakari, S. A., Alhussan, A. A., Qenawy, A.-S.T., Elshewey, A. M. & Eed, M. An enhanced long short-term memory recurrent neural network deep learning model for potato price prediction. *Potato Res.* <https://doi.org/10.1007/s11540-024-09744-x> (2024).
46. Ul Hassan, N. et al. Improved opposition-based particle swarm optimization algorithm for global optimization. *Symmetry (Basel)* **13**, 2280 (2021).
47. Ashraf, A. et al. *An Improved Fire Hawks Optimizer for Function Optimization*, 68–79 (2023). [https://doi.org/10.1007/978-3-031-36622-2\\_6](https://doi.org/10.1007/978-3-031-36622-2_6).
48. Alibrahim, H. & Ludwig, S. A. Hyperparameter optimization: Comparing genetic algorithm against grid search and Bayesian optimization. In *2021 IEEE Congress on Evolutionary Computation (CEC)*, 1551–1559 (2021). <https://doi.org/10.1109/CEC45853.2021.9504761>.
49. Ghali, U. M. et al. Advanced chromatographic technique for performance simulation of anti-Alzheimer agent: An ensemble machine learning approach. *SN Appl. Sci.* **2**, 1–12 (2020).
50. Ma, J. et al. Accurate and non-destructive monitoring of mold contamination in foodstuffs based on whole-cell biosensor array coupling with machine-learning prediction models. *J. Hazard. Mater.* **449**, 131030 (2023).
51. Chen, T., Liang, W., Zhang, X., Wang, Y. & Lu, X. Screening and identification of unknown chemical contaminants in food based on liquid chromatography—High-resolution mass spectrometry and machine learning. *Anal. Chim. Acta* **1287**, 342116 (2024).
52. He, P. et al. Rapid and stable detection of three main mycotoxins in rice using SERS optimized AgNPs @ K30 coupled multivariate calibration. *Food Chem.* **398**, 133883 (2023).
53. Lem, P. & Habauzit, D. Prioritization of mycotoxins based on mutagenicity and carcinogenicity evaluation using combined in silico QSAR methods. *Environ. Pollut.* **323**, 121284 (2023).
54. Kim, Y. et al. Detection of aflatoxins in ground maize using a compact and automated Raman spectroscopy system with machine learning. *Curr. Res. Food Sci.* **7**, 100647 (2023).
55. Abdullahi, H. U., Usman, A. G. & Abba, S. I. Modelling the absorbance of a bioactive compound in HPLC method using artificial neural network and multilinear regression methods. *Dutse J. Pure Appl. Sci.* **6**, 362–371 (2020).
56. Ul Hassan, N. et al. Improved opposition-based particle swarm optimization algorithm for global optimization. *Symmetry (Basel)* **13**, 1–23 (2021).
57. Mati, S. et al. Engineering applications of artificial intelligence predicting consumer price index amidst uncertainty: Gaussian random fuzzy number-based evidential neural network for West African economies with COVID-19 and Russia–Ukraine war dynamics. *Eng. Appl. Artif. Intell.* **136**, 109004 (2024).
58. Saleh, A. & Aljundi, I. H. Machine learning insight into inhibition efficiency modelling based on modified graphene oxide of diaminohehexane (DAH-GO) and diaminoctane. *Carbon Trends* **16**, 100373 (2024).

59. Isik, S., Usman, A. G. & Abba, S. I. A chemometrics-based approach for the determination of thymoquinone from *Nigella sativa* L. (Black Cumin) seeds of different geographical regions using the HPLC technique. *Istanb. J. Pharm.* **53**, 320–328 (2023).
60. Shafqat, Z., Iqbal, M., Bangyal, W. H. & Almakhlles, D. Opinion mining of politics and inflation using roman Urdu dataset. In *Proceedings of the 2022 International Conference on Human-Centered Cognitive Systems HCCS 2022*, 1–6 (2022). <https://doi.org/10.1109/HCCS55241.2022.10090276>.
61. Hosney, R., Talaat, F. M., El-Gendy, E. M. & Saafan, M. M. AutYOLO-ATT: An attention-based YOLOv8 algorithm for early autism diagnosis through facial expression recognition. *Neural Comput. Appl.* **7**, 17199–17121 (2024).
62. Magdy, H., Ahmed, B., Shaban, O. & Saafan, M. M. Prostate cancer grading framework based on deep transfer learning and Aquila optimizer. *Neural Comput. Appl.* **36**, 7877–7902 (2024).
63. Magdy, H., Asmaa, B., Sayed, E., Eman, H. & El Gendy, M. An aseptic approach towards skin lesion localization and grading using deep learning and harris hawks optimization. *Multimed. Tools Appl.* **83**, 19787–19815 (2024).
64. Talaat, F. M., Mahmoud, E. M. E. & Samah, M. S. Utilizing social media and machine learning for personality and emotion recognition using PERS. *Neural Comput. Appl.* **35**, 23927–23941 (2023).
65. Magdy, H., Esraa, B., Antar, R., Saafan, M. M. & El, E. M. A comprehensive framework towards segmenting and classifying breast cancer patients using deep learning and Aquila optimizer. *J. Ambient Intell. Humaniz. Comput.* **14**, 7897–7917 (2023).
66. Selim, H., Haikal, A. Y., Labib, L. M. & Saafan, M. M. MCHIAO: A modified coronavirus herd immunity-Aquila optimization algorithm based on chaotic behavior for solving engineering problems. *Neural Comput. Appl.* **36**, 13381–13465 (2024).
67. Akl, D. T., Saafan, M. M. & Haikal, A. Y. IHHO: An improved Harris Hawks optimization algorithm for solving engineering problems. *Neural Comput. Appl.* **36**, 12185–12298 (2024).
68. Vapnik, V., Golowich, S. E., Ave, M. & Hill, M. *Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing*.
69. Jibril, M. M. et al. Implementation of nonlinear computing models and classical regression for predicting compressive strength of high-performance concrete. *Appl. Eng. Sci.* **15**, 100133 (2023).
70. Azimi-Pour, M., Eskandari-Naddaf, H. & Pakzad, A. Linear and non-linear SVM prediction for fresh properties and compressive strength of high volume fly ash self-compacting concrete. *Constr. Build. Mater.* **230**, 117021 (2020).
71. Moayed, H., Gör, M., Lyu, Z. & Bui, D. T. Herding behaviors of grasshopper and Harris hawk for hybridizing the neural network in predicting the soil compression coefficient. *Measurement* **152**, 107389 (2020).
72. Moayed, H. et al. Harris hawks optimization: A novel swarm intelligence technique for spatial assessment of landslide susceptibility. *Sensors (Switzerland)* <https://doi.org/10.3390/s19163590> (2019).
73. Heidari, A. A. et al. Harris hawks optimization: Algorithm and applications. *Futur. Gener. Comput. Syst.* **97**, 849–872 (2019).
74. Kennedy, J. & Eberhart, R. Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks*, vol. IV, 1942–1948 (1995). <https://doi.org/10.1109/ICNN.1995.488968>.
75. Malik, A., Tikhmarine, Y., Sammen, S. S., Abba, S. I. & Shahid, S. Prediction of meteorological drought by using hybrid support vector regression optimized with HHO versus PSO algorithms. *Environ. Sci. Pollut. Res.* **28**, 39139–39158 (2021).
76. Seifi, A., Ehteram, M., Singh, V. P. & Mosavi, A. Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN. *SSRN J.* <https://doi.org/10.2139/ssrn.3636873> (2020).
77. Mirjalili, S., Song Dong, J., Lewis, A. & Sadiq, A. S. Particle swarm optimization: Theory, literature review, and application in airfoil design. *Stud. Comput. Intell.* [https://doi.org/10.1007/978-3-030-12127-3\\_10](https://doi.org/10.1007/978-3-030-12127-3_10) (2020).
78. Musa, B., Yimen, N., Abba, S. I., Adun, H. H. & Dagbasi, M. Multi-state load demand forecasting using hybridized support vector regression integrated with optimal design of off-grid energy systems—A metaheuristic approach. *Processes* **9**, 1166 (2021).
79. Wang, Z., Yashao, C., Qi, B. & Yang, B. Prediction of daytime variations of HO<sub>2</sub> radical concentrations in the marine boundary layer using BP network. *Sci. China Chem.* **53**, 2616–2621 (2010).
80. Pfeifer, N., Leinenbach, A., Huber, C. G. & Kohlbacher, O. Statistical learning of peptide retention behavior in chromatographic separations: A new kernel-based approach for computational proteomics. *BMC Bioinform.* **8**, 1–14 (2007).
81. Saini, R. & Kumar, P. Optimization of chlorpyrifos degradation by fenton oxidation using CCD and ANFIS computing technique. *J. Environ. Chem. Eng.* **4**, 2952–2963 (2016).
82. Mohamed, A. K., Mohamed, A. W., Elfeky, E. Z. & Saleh, M. Enhancing AGDE algorithm using population size reduction for global numerical optimization. *Adv. Intell. Syst. Comput.* **723**, 62–72 (2018).
83. Mustapha, M. T., Uzun Ozsahin, D., Uzun, B. & Ozsahin, I. Evaluation of the thyroid cancer treatment techniques with fuzzy VIKOR. In *2022 Advances in Science and Engineering Technology International Conferences (ASET)* 1–5 (IEEE, 2022). <https://doi.org/10.1109/ASET53988.2022.9735083>.
84. Haruna, S. I. et al. Compressive strength of self-compacting concrete modified with rice husk ash and calcium carbide waste modeling: A feasibility of emerging emotional intelligent model (EANN) versus traditional FFNN. *Arab. J. Sci. Eng.* **46**, 11207–11222 (2021).
85. Ismail, S., Abdulkadir, R. A., Usman, A. G. & Abba, S. I. Development of chemometrics-based neurocomputing paradigm for simulation of manganese extraction using solid—Phase tea waste. *Model. Earth Syst. Environ.* <https://doi.org/10.1007/s40808-022-01369-8> (2022).
86. Abba, S. I. et al. Fluoride and nitrate enrichment in coastal aquifers of the Eastern Province, Saudi Arabia: The influencing factors, toxicity, and human health risks. *Chemosphere* **336**, 139083 (2023).
87. Uzun Ozsahin, D. et al. Clinical modelling of RVHF using pre-operative variables: A direct and inverse feature extraction technique. *Diagnostics* **12**, 3061 (2022).
88. Bala, K., Etikan, I., Usman, A. G. & Abba, S. I. Artificial-intelligence-based models coupled with correspondence analysis visualization on ART—Cases from Gombe State, Nigeria: A comparative study. *Life* **13**, 715 (2023).
89. Mahmoud, K. et al. Prediction of the effects of environmental factors towards COVID-19 outbreak using AI-based models. *IAES Int. J. Artif. Intell.* **10**, 35–42 (2021).
90. Legates, D. R. & McCabe, G. J. Jr. Evaluating the use of 'goodness of fit' measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**, 233–241 (1999).
91. Elkiran, G., Nourani, V. & Abba, S. I. Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *J. Hydrol.* **577**, 123962 (2019).
92. Alotaibi, B. S. et al. Sustainable green building awareness: A case study of Kano green construction. *Buildings* **13**, 2387 (2023).
93. Kazemi, P. et al. Computational intelligence modeling of granule size distribution for oscillating milling. *Powder Technol.* **301**, 1252–1258 (2016).
94. Usman, A. G., Ahmad, M. H., Danraka, R. N. & Abba, S. I. The effect of ethanolic leaves extract of *Hymenodictyon floribundum* on inflammatory biomarkers: A data-driven approach. *Bull. Natl. Res. Cent.* **45**, 1–12 (2021).
95. Khalid, G. M. & Usman, A. G. Application of data-intelligence algorithms for modeling the compaction performance of new pharmaceutical excipients. *Future J. Pharm. Sci.* **7**, 1–11 (2021).
96. Usman, A. G., Isik, S., Abba, S. I. & Meriçli, F. Artificial intelligence-based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method. *Turk. J. Chem.* **44**, 1339–1351 (2020).
97. Usman, A. G., Isik, S., Abba, S. I. & Meriçli, F. Chemometrics-based models hyphenated with ensemble machine learning for retention time simulation of isoquercitrin in *Coriander sativum* L. using high-performance liquid chromatography. *J. Sep. Sci.* **44**, 843–849 (2021).

98. Benaafi, M., Yassin, M. A., Usman, A. G. & Abba, S. I. Neurocomputing modelling of hydrochemical and physical properties of groundwater coupled with spatial clustering, GIS, and statistical techniques. *Sustainability* **14**, 2250 (2022).
99. Jibril, M. M. et al. New random intelligent chemometric techniques for sustainable geopolymer concrete: Low-energy and carbon-footprint initiatives. *Asian J. Civ. Eng.* **25**, 2287–2305 (2024).
100. Jibril, M. M. et al. High strength concrete compressive strength prediction using an evolutionary computational intelligence algorithm. *Asian J. Civ. Eng.* <https://doi.org/10.1007/s42107-023-00746-7> (2023).
101. Gbadamosi, A. et al. New-generation machine learning models as prediction tools for modeling interfacial tension of hydrogen–brine system. *Int. J. Hydrog. Energy* **50**, 1326–1337 (2023).
102. Usman, A. G. et al. Optimization algorithms for modeling conversion and naphtha yield in the catalytic co-cracking of plastic in HVGO. *Process Saf. Environ. Prot.* **196**, 106958 (2025).
103. Abba, S. I. et al. Optimizing sustainable desalination plants with advanced ML-based uncertainty analysis. *Appl. Soft Comput.* **169**, 112624 (2025).
104. Abba, S. I., Benaafi, M., Usman, A. G. & Aljundi, I. H. Inverse groundwater salinization modeling in a sandstone's aquifer using stand-alone models with an improved non-linear ensemble machine learning technique. *J. King Saud Univ. Comput. Inf. Sci.* **34**, 8162–8175 (2022).
105. Abba, S. I. et al. Optimization of extreme learning machine with metaheuristic algorithms for modelling water quality parameters of Tamburawa Water Treatment Plant in Nigeria. *Water Resour. Manag.* **39**, 1377–1401. <https://doi.org/10.1007/s11269-024-04027-z> (2024).
106. Ahmad, M. H., Usman, A. G. & Abba, S. I. Comparative performance of extreme learning machine and Hammerstein–Weiner models for modelling the intestinal hyper-motility and secretory inhibitory effects of methanolic leaf extract of *Combretum hypoleucum* Diels (Combretaceae). *Silico Pharmacol.* **9**, 1–11 (2021).
107. Usman, A. G., Işık, S. & Abba, S. I. Hybrid data-intelligence algorithms for the simulation of thymoquinone in HPLC method development. *J. Iran. Chem. Soc.* **18**, 1537–1549 (2021).
108. Nielsen, K. F. & Smedsgaard, J. Fungal metabolite screening: Database of 474 mycotoxins and fungal metabolites for dereplication by standardised liquid chromatography–UV–mass spectrometry methodology. *J. Chromatogr. A* **1002**, 111–136 (2003).
109. Khajeh, M., Sarafraz-Yazdi, A. & Moghadam, A. F. Modeling of solid-phase tea waste extraction for the removal of manganese and cobalt from water samples by using PSO-artificial neural network and response surface methodology. *Arab. J. Chem.* **10**, S1663–S1673 (2017).
110. Shadrin, D., Pukalchik, M., Kovaleva, E. & Fedorov, M. Artificial intelligence models to predict acute phytotoxicity in petroleum contaminated soils. *Ecotoxicol. Environ. Saf.* **194**, 110410 (2020).
111. Rajae, T., Ebrahimi, H. & Nourani, V. Review papers: A review of the artificial intelligence methods in groundwater level modeling. *J. Hydrol.* **572**, 336–351 (2019).
112. Farajvand, M., Kiarostami, V., Davallo, M., Ghaedi, A. & Fatahi, F. Rapid extraction of copper ions in water, tea, milk and apple juice by solvent-terminated dispersive liquid–liquid microextraction using p-sulfonatocalix (4) arene: Optimization by artificial neural networks coupled bat inspired algorithm and response sur. *J. Food Sci. Technol.* **56**, 4224–4232 (2019).
113. Yaseen, Z. M. et al. Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Adv. Eng. Softw.* **115**, 112–125 (2018).
114. Qin, Q. et al. Deep learning on chromatographic data for segmentation and sensitive analysis. *J. Chromatogr. A* **1634**, 461680 (2020).
115. Moayedi, H., Osooli, A., Nguyen, H. & Rashid, A. S. A. A novel Harris hawks' optimization and k-fold cross-validation predicting slope stability. *Eng. Comput.* <https://doi.org/10.1007/s00366-019-00828-8> (2019).
116. Zhong, X., Duan, M., Zhang, X. & Cheng, P. A hybrid differential evolution based on gaining-sharing knowledge algorithm and Harris hawks optimization. *PLoS ONE* **16**, 1–24. <https://doi.org/10.1371/journal.pone.0250951> (2021).
117. Mohamed, A. W., Hadi, A. A., Mohamed, A. K. & Awad, N. H. Evaluating the performance of adaptive GainingSharing knowledge based algorithm on CEC 2020 benchmark problems. In *2020 IEEE Congress on Evolutionary Computation CEC 2020—Conference Proceedings* (2020). <https://doi.org/10.1109/CEC48606.2020.9185901>.
118. Abbaspour-Gilandeh, Y., Ghadakchi-Bazaz, H. & Davari, M. Discriminating healthy wheat grains from grains infected with *Fusarium graminearum* using texture characteristics of image-processing technique, discriminant analysis, and support vector machine methods. *J. Intell. Syst.* **29**, 1576–1586 (2020).
119. Ge, H., Jiang, Y., Lian, F., Zhang, Y. & Xia, S. Quantitative determination of aflatoxin B1 concentration in acetonitrile by chemometric methods using terahertz spectroscopy. *Food Chem.* **209**, 286–292 (2016).
120. Tao, H. et al. Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing* **489**, 271–308 (2022).
121. Ghorbani, M. A., Deo, R. C., Yaseen, Z. M. H., Kashani, M. & Mohammadi, B. Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. *Theor. Appl. Climatol.* **133**, 1119–1131 (2018).

## Acknowledgments

The Researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QUAPC-2025).

## Author contributions

AGU: conceptualization, writing original draft, visualization, software, validation, reviewing. SM: visualization, software, validation, reviewing. HD: visualization, software, validation, reviewing. AAS: writing original draft, visualization, software. SIA: writing original draft, visualization, software. HA: writing original draft, visualization, software. TR: writing original draft, visualization, software, funding.

## Declarations

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to A.G.U. or T.R.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025