Artificial liver classifier: a new alternative to conventional machine learning models

Mahmood A. Jumaah[1]*, Yossra H. Ali[1] and Tarik A. Rashid[2]*

[1]Department of Computer Science, University of Technology, Baghdad, Iraq, [2]Department of Computer Science and Engineering, Artificial Intelligence and Innovation Centre (AIIC), University of Kurdistan Hewlêr, Erbil, Iraq

Introduction: Supervised machine learning classifiers sometimes face challenges related to the performance, accuracy, or overfitting.

Methods: This paper introduces the Artificial Liver Classifier (ALC), a novel supervised learning model inspired by the human liver's detoxification function. The ALC is characterized by its simplicity, speed, capability to reduce overfitting, and effectiveness in addressing multi-class classification problems through straightforward mathematical operations. To optimize the ALC's parameters, an improved FOX optimization algorithm (IFOX) is employed during training.

Results: We evaluate the proposed ALC on five benchmark datasets: Iris Flower, Breast Cancer Wisconsin, Wine, Voice Gender, and MNIST. The results demonstrate competitive performance, with ALC achieving up to 100% accuracy on the Iris dataset—surpassing logistic regression, multilayer perceptron, and support vector machine—and 99.12% accuracy on the Breast Cancer dataset, outperforming XGBoost and logistic regression. Across all datasets, ALC consistently shows smaller generalization gaps and lower loss values compared to conventional classifiers.

Discussion: These findings highlight the potential of biologically inspired models to develop efficient machine learning classifiers and open new avenues for innovation in the field.

1 Introduction

Artificial intelligence (AI) has many branches according to the tasks to be performed, with machine learning (ML) being one of the most well-known branches that has gained prominence alongside the development of computer science. It focuses on developing systems and algorithms that automatically learn from data without explicit programming (Kolides et al., 2023; Dwivedi et al., 2021; Khudhair et al., 2024c). However, two main types of ML are categorized according to the problem to be solved: supervised learning and unsupervised learning. Supervised learning relies on having pre-labeled input data (denoted *X*) and the desired output (denoted *y*). This type of learning aims to understand the hidden relationship between inputs and outputs to predict new outcomes based on unseen (new) input data (Jiang et al., 2020; Beam and Zupancic, 2022; Zhao et al., 2024). On the other hand, unsupervised learning uses input data that is not pre-labeled (does not contain output y). Instead, an unsupervised learning model is applied to discover patterns and hidden relationships in the data autonomously based on the input data only (Watson, 2023; Molnar et al., 2020). Furthermore, there are other types of ML such as reinforcement learning (RL), which interact directly with the problem's environment to build policies that guide decision-making based on rewards and penalties

obtained through trial and error (Gao and Schweidtmann, 2024; Mutar and Jawad, 2023; Kumar et al., 2023; Khudhair et al., 2024b).

In the early stages of AI, researchers focused on building systems (with minimal intelligence) capable of performing specific tasks using fixed rules (conditional and logical operations). As the field evolved, scientists realized that intelligent systems needed methods to learn from data, rather than relying on rigid rule-based methods with minimal capabilities (Grzybowski et al., 2024; Jabber et al., 2023). As a result, supervised learning algorithms, specifically classifiers, emerged as tools for learning systems to make predictions or decisions based on the available experiences. However, one of the most preeminent algorithms in supervised learning is artificial neural network (ANN), inspired by the fundamental concept of neurons in the human brain and how they are interconnected (Palanivinayagam et al., 2023). These networks are based on the concept of neurons, which are basic units in the brain that communicate with each other to perform processes such as thinking and learning (Schmidgall et al., 2024). The algorithm simulates the functions of brain cells by proposing multiple layers of artificial neurons (an input layer and an output layer). These neurons interact with each other using weights assigned to each connection, and the role of the algorithm is to optimize these weights to minimize the error resulting from interactions with the input data, thereby producing accurate outputs (Jumaah et al., 2024). Moreover, an older algorithm inspired by mathematics is the logistic regression (LR), which aims to find a perfect line that best fits the data points, minimizing the error between actual and predicted labels. These methods were used in statistical analyses before being adopted in ML (Jumin et al., 2020). The complexity of linear operations increased, leading to more sophisticated methods, such as support vector machine (SVM), where the main idea is to create clear boundaries between different data classes by maximizing the margin between them (Quan and Pu, 2022). Comprehensively, most of ML classifiers have drawn their inspirations from mathematical operations or nature (e.g., simulating the functioning of human brain cells) to create robust systems (classifiers) for solving complex problems. Current ML classifiers face multiple challenges related to performance, accuracy or loss, overfitting, and handling data with complex and non-linear patterns (Tufail et al., 2023; Khudhair et al., 2024a).

In this context, this paper proposes a new classifier called artificial liver classifier (ALC), inspired by the human liver's biological functions. Specifically, it draws on the detoxification function, highlighting its ability to process toxins and convert them into removable forms. Additionally, improvements have been made to FOX optimization algorithm (FOX), a state-of-the-art optimization algorithm, to enhance its performance and ensure compatibility with the proposed ALC. The research aims to bridge the gap in current ML's algorithms by combining the simplicity of mathematical design with solid performance by simulating the detoxification function in the human liver. Furthermore, the proposed classifier aims to improve classification performance by processing data dynamically, simulating the human liver's adaptive ability, enabling its application in fields requiring high-precision solutions and flexibility in dealing with different data patterns. The main challenge lies in transforming the liver's detoxification function into a simplified mathematical model that effectively incorporates properties such as repetition, interaction, and adaptation to the data (Tan et al., 2024). By comparing the proposed classifier with established ML classifiers, the study expects to improve the performance of ML, including increased computation speed, better handling of overfitting problems, and avoidance of excessive computational complexity. Additionally, this paper introduces a new concept for drawing inspiration from biological systems, opening up extensive opportunities for researchers to develop mathematical models based on other biological functions of the liver, such as filtering blood or amino acid regulation (Ishibashi et al., 2009). Moreover, it represents a starting point for interdisciplinary applications combining biology, mathematics, and AI, enhancing our understanding of incorporating natural processes into ML techniques to create efficient, reliable, and intelligent systems.

The proposed ALC has been evaluated using a variety of commonly used ML datasets, including Wine, Breast Cancer Wisconsin, Iris Flower, MNIST, and Voice Gender (Hoffmann et al., 2019), which are explained in detail in Section 4.1. This diversity in the datasets ensures extensive coverage of different data types, including text, images, and audio, and enables handling binary and multi-class classification problems (Seliya et al., 2021; Parimala and Muneeswari , 2023; Sidumo et al., 2022). The purpose of using these datasets is to conduct comprehensive tests to assess the performance of the proposed ALC and compare it with the established classifiers. The originality and contributions that distinguish this research are as follows:

1. Introducing a new classifier inspired by the liver's biological functions, specifically detoxification, highlighting new possibilities in designing effective classification algorithms based on biological behavior.
2. Enhancing the FOX to improve its performance, address existing limitations, and ensure better compatibility with the proposed ALC.
3. Relying on simple mathematical models that simulate the liver's biological interactions, ensuring a balance between design simplicity and high performance.
4. Opening new avenues for researchers to draw inspiration from human organ functions, such as the liver, and simulate them in computational ways to contribute innovative solutions for real-world challenges.
5. Testing the proposed ALC on diverse datasets demonstrates its effectiveness through experimental results and comparisons with established classifiers.

This paper is structured as follows: Section 2 reviews the literature that has attempted to address classification issues across various data types. Section 3 provides an analytical overview of the human liver, focusing on detoxification function and the study's motivation. Section 4 present the used materials and the proposed methodology, including the improvement of classifier design and FOX training algorithm. Sections 5,6 cover the presentation and analysis of results, including comparisons with previous works. Finally, the study concludes with findings, recommendations, limitations, and future research directions in Section 7.

2 Related works

This section reviews the standard algorithms used in ML classification, with their practical applications across various

datasets highlighted (Sarker, 2021). Additionally, recent studies in the field are discussed to identify existing challenges and to shed light on research gaps requiring further attention (Azevedo et al., 2024). Accordingly, the extent to which the proposed classifier can offer practical solutions to these gaps and contribute to the future advancement of the field will be investigated. However, Xiao et al. utilized 12 standard ML classifiers on the MNIST dataset, demonstrating its suitability as a benchmark for evaluating the proposed ALC. Their results identified the Support Vector Classifier (SVC) with a polynomial kernel (C = 100) as the best-performing model, achieving an accuracy of 0.978 (Xiao et al., 2017). This comparable result poses a challenge for the proposed ALC to surpass. Furthermore, the study (Cohen et al., 2017) employed online pseudo-inverse update method (OPIUM) to classify the MNIST dataset, achieving an accuracy of 0.9590. However, the author noted that these results do not represent cutting-edge methods but rather serve as an instructive baseline and a means of validating the dataset. This makes it feasible to compare the performance of the proposed ALC against OPIUM, as surpassing this baseline would demonstrate an improvement over existing methods. On the other hand, in a comparative study by Cortez et al., three classifiers—SVM, multiple regression (MR), and ANN—were evaluated on the Wine dataset. The SVM model demonstrated superior performance, achieving accuracies of 0.8900 for red wine and 0.8600 for white wine, outperforming the other methods with an average accuracy of 0.8790 (Cortez et al., 2009). Hence, the findings of Cortez et al. serve as a foundation for further advancements in ML applications, providing a basis for evaluating the proposed ALC.

Another study utilized a recursive recurrent neural network (RRNN) on Breast Cancer Wisconsin dataset. The results demonstrated that the proposed model achieved an accuracy of 0.9950 (Rajeswari and Sakthi Priya, 2025). Despite its outstanding performance, the computational demands of RRNN require substantial resources, which may limit their applicability in resource-constrained environments. Moreover, the study (Fan et al., 2024) presents a new classification model called CS3W-IFLMC. This model incorporates intuitionistic fuzzy (IF) and cost-sensitive three-way decisions (CS3WD) approaches, contributing to improved classification accuracy and reduced costs associated with incorrect decisions. The proposed model has been evaluated using 12 benchmark datasets, demonstrating superior performance compared to large margin distribution machine (LDM), FSVM, and SVM. However, the study remains limited in scope, as it focuses solely on binary classification tasks and does not extend to multi-class classification problems (Fan et al., 2024). Furthermore, in another study, the researchers examined gender classification (male or female) based on voice data using multi-layer perceptron (MLP). The findings showed that the MLP model outperformed several other methods, including LR, classification and regression tree (CART), random forest (RF), and SVM. The MLP achieved a classification accuracy of 0.9675. This study concluded that the proposed model demonstrates strong discriminative power between genders, which enhances its applicability in auditory data classification tasks (Buyukyilmaz and Cibikdiken, 2016).

The reviewed literature, highlights significant advancements in classification models, primarily focusing on improving performance and addressing computational challenges. However,

several limitations and research gaps remain. One major issue is the reliance on computationally intensive methods, which can hinder applicability in resource-constrained environments. The absence of practical hyperparameter tuning or reduction mechanisms may also contribute to overfitting and computational inefficiencies. These limitations underscore the need for a new classifier to address such challenges. Hence, the proposed ALC should emphasize simplicity in design to ensure faster training time with lower cost.

3 Detoxification in liver and motivation

The liver, as illustrated in Figure 1, is the largest internal organ in the human body and is vital in numerous complex physiological processes. It is located in the right upper quadrant of the abdominal cavity and consists of two primary lobes, the right and left, surrounded by a thin membrane known as the hepatic capsule (Moradi et al., 2020). Internally, the liver is composed of microscopic units called hepatic lobules. These hexagonal structures contain hepatic cells organized around a central vein. The lobules are permeated by a network of hepatic sinusoids, which are small channels through which blood flows, facilitating the exchange of oxygen and nutrients between the blood and hepatic cells (Kennedy et al., 2021). Furthermore, the liver receives blood from two sources, each contributing different functions. The oxygenated blood enters via the hepatic artery from the aorta, meeting the liver's energy demands. While, the portal vein delivers nutrient-rich and toxin-rich blood from the gastrointestinal tract and spleen (Schlegel et al., 2023). The blood from both sources mixes in the hepatic sinusoids, allowing the hepatic cells to perform metabolic and regulatory functions efficiently (Gibert-Ramos et al., 2021).

However, detoxification is one of the most important liver's functions, which removes toxins from the bloodstream (Donati et al., 2020). Detoxification occurs in two phases. In the phase I, hepatic enzymes known as cytochrome P450 chemically modify toxins through oxidation and reduction reactions, altering their structures to make them more reactive (Guengerich, 2020). In the phase II, the modified compounds are conjugated with water-soluble molecules such as sulfates or glucuronic acid, making them easier to excrete (Sun and Schanze, 2022). Finally, the toxins are either excreted via bile into the digestive tract or removed from the bloodstream by the kidneys (Zhang et al., 2020).

The complex biochemical system of the liver has inspired us to develop a new ML classifier known as ALC, modeled after the liver's detoxification mechanisms. The design of the proposed ALC was guided by an in-depth understanding of the liver's two primary detoxification phases—Cytochrome P450 enzymes and Conjugation pathways—where toxins are transformed into excretable compounds. The proposed ALC classify feature vectors effectively with minimum training time by simulating these phases using simple ML and optimization methods. This innovation marks a significant step forward, demonstrating how biological systems can inspire advanced computational models. It particularly encourages researchers in computer
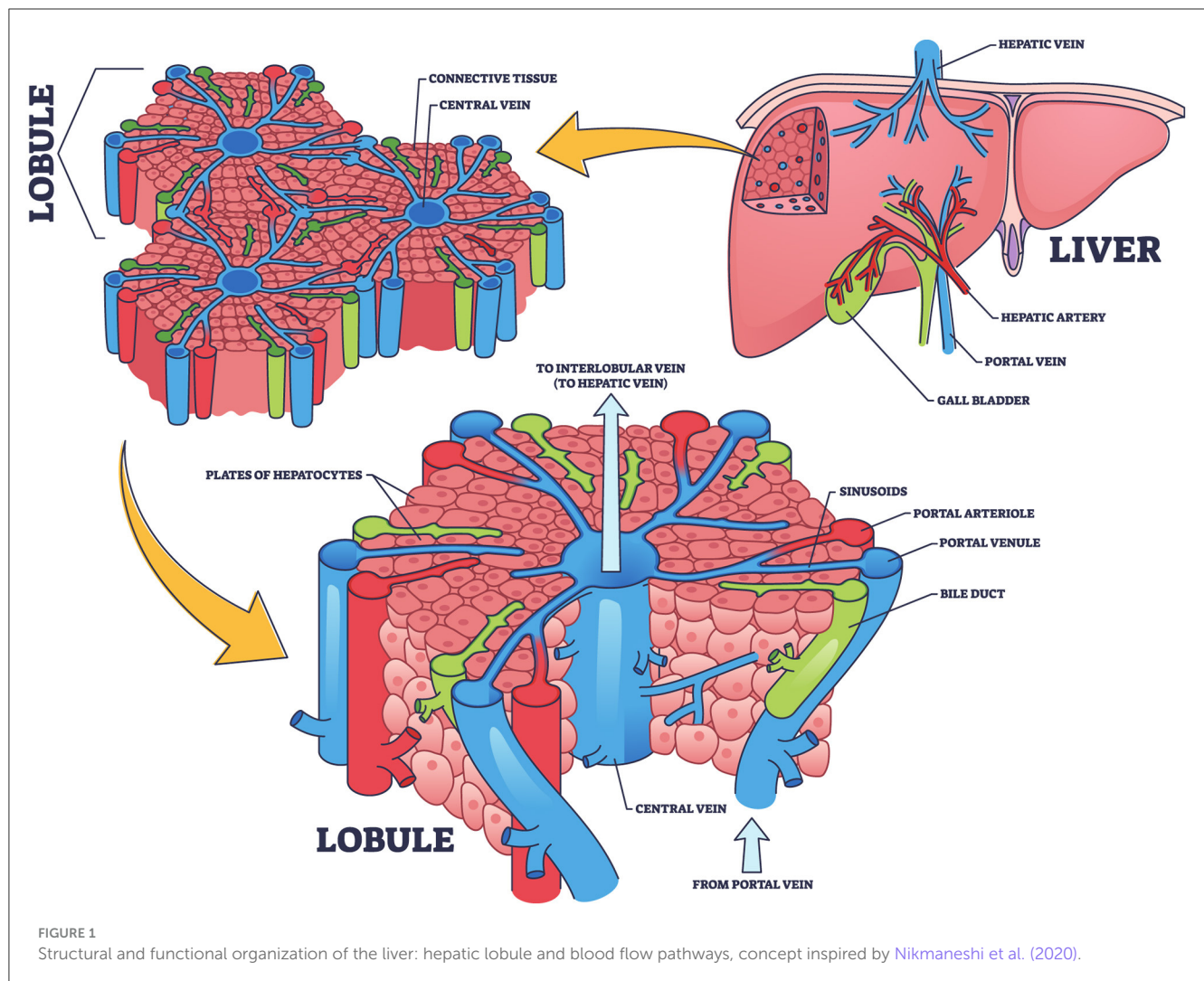
science to explore biological processes for developing intelligent ML models.

4 Materials and methods

This section presents the standard datasets employed for evaluating the proposed ALC in the conducted experiments. Additionally, the architecture of the proposed ALC is provided, including mathematical equations, algorithms, and flowcharts. Furthermore, the section elaborates on the FOX, which serves as the learning algorithm for the proposed ALC, highlighting its improvements.

4.1 Materials

The following datasets are widely used by ML researchers to evaluate their work, making these benchmark datasets suitable for this paper. The MNIST dataset comprises 70,000 grayscale images of handwritten digits (0–9), each of size 28 × 28 pixels. It is widely used for multi-class classification tasks due to its diversity and large size (Elizabeth Rani et al., 2022). To utilize the MNIST dataset with the proposed ALC, each image was preprocessed by flattening it to a vector of 784 dimensions. Each pixel was normalized, with its value transformed to have zero mean and unit variance to ensure consistent scaling. This was then followed by the use of linear discriminant analysis (LDA) to project or reduce the data into a low-dimensional space. Hence, LDA reduces each image to a nine-dimensional feature vector to effectively capture the most discriminative features while also reducing computational requirements (Lasalvia et al., 2022). Additionally, the Iris dataset, a small-scale collection containing 150 instances across three classes with four features per instance, was included in the proposed ALC evaluation (Goyal et al., 2021; Oladejo et al., 2024; Kumar et al., 2024). The Breast Cancer Wisconsin dataset, a binary dataset containing 569 samples with 30 features each, was employed to assess the proposed ALC's performance on high-dimensional data (Alshayeji et al., 2022; Rajeswari and Sakthi Priya, 2025). Furthermore, the Wine dataset, consisting of 178 samples across three classes with 13 features per instance, was selected for its multi-class nature (Oladejo et al., 2024; Waheed and Humaidi, 2023). Finally, the Voice Gender dataset was employed to ensure feature diversity. This dataset

comprises 3,168 samples, each defined by 21 acoustic features, aimed at distinguishing gender (male or female) by leveraging unique vocal characteristics (Buyukyilmaz and Cibikdiken, 2016). These datasets collectively provided a diverse range of classification challenges, enabling a comprehensive evaluation of the proposed ALC's performance.

4.2 Methods

This section begins with a detailed introduction to the architecture of the proposed ALC. Moreover, it delves into the improvements made to the FOX as a learning algorithm, highlighting its key modifications.

4.2.1 Artificial liver classifier

As explained earlier in Section 3, the detoxification process involves the liver's ability to process toxins. Oxygenated blood enters the liver via the hepatic artery, while nutrient-rich blood flows through the portal vein. These sources mix within the hepatic sinusoids, enabling hepatic cells to perform essential functions, including a detoxification function that comprises two phases.

In order to model the detoxification process in the liver, a biologically inspired computational model was chosen to be implemented, where every mathematical operation is merely treated as a simplistic representation of various known mechanisms regarding hepatic detoxification (Donati et al., 2020). The major biochemical steps taken by hepatocytes in the elimination process are considered to include detoxification in two phases: oxidation (Phase I), which is carried out by specialized enzymes and cofactors, and conjugation (Phase II), which is also supported by specific enzymes and cofactors (Sun and Schanze, 2022; Guengerich, 2020). In this formulation, these steps are coordinated by being transformed into operations in matrices. A set of molecular input toxins is first linearly overlaid by a cofactor matrix C to model the oxidation action of cytochrome P450, followed by being passed through a non-linear activation grid to model metabolite selection at a threshold. The second transformation is modeled in the form of conjugation through the interaction of vitamins, and this is subsequently normalized through the use of softmax to represent the classification of detoxified products. Although these operations are not intended to accurately reproduce the actual biochemical kinetics, they are carefully selected so that the structural and functional analogies are preserved—allowing the multi-staged, enzyme-driven, and spatially distributed nature of detoxification to be reproduced within a mathematically consistent and learnable system.

Phase I: toxins are chemically modified to become more reactive. This phase is mathematically simulated by the following equation:

$$A_{ji} = \frac{1}{n} \sum_{k=1}^{n} (X_{jk} \cdot C_{ki}) + \frac{1}{fp} \sum_{k=1}^{f} \sum_{l=1}^{p} C_{kl} \qquad (1)$$

where A_{ji} is the matrix of reactive toxins, X is the input toxins matrix and n is the number of inputs. The C is initialized

```
    Input: toxins, number of features, number of
    lobules p, number of outputs, detoxification
    cycles, and detoxification power.
    Output: predicted classes.
 1: Initialize cofactor matrix C and vitamin matrix
    V randomly.
 2: Initialize the IFOX training algorithm.      ▷ See
    Algorithm 2
 3: Optimize C and V using IFOX over defined cycles.
 4: Compute reactive toxins.        ▷ using Equation 1
 5: Activate reactive toxins.        ▷ Phase I, using
    Equation 2
 6: Perform conjugation to make toxins less harmful.
    ▷ Phase II, using  Equation 3
 7: Normalize outputs to obtain predicted classes. ▷
    Elimination, using  Equation 4
 8: return predicted classes ŷ.
```
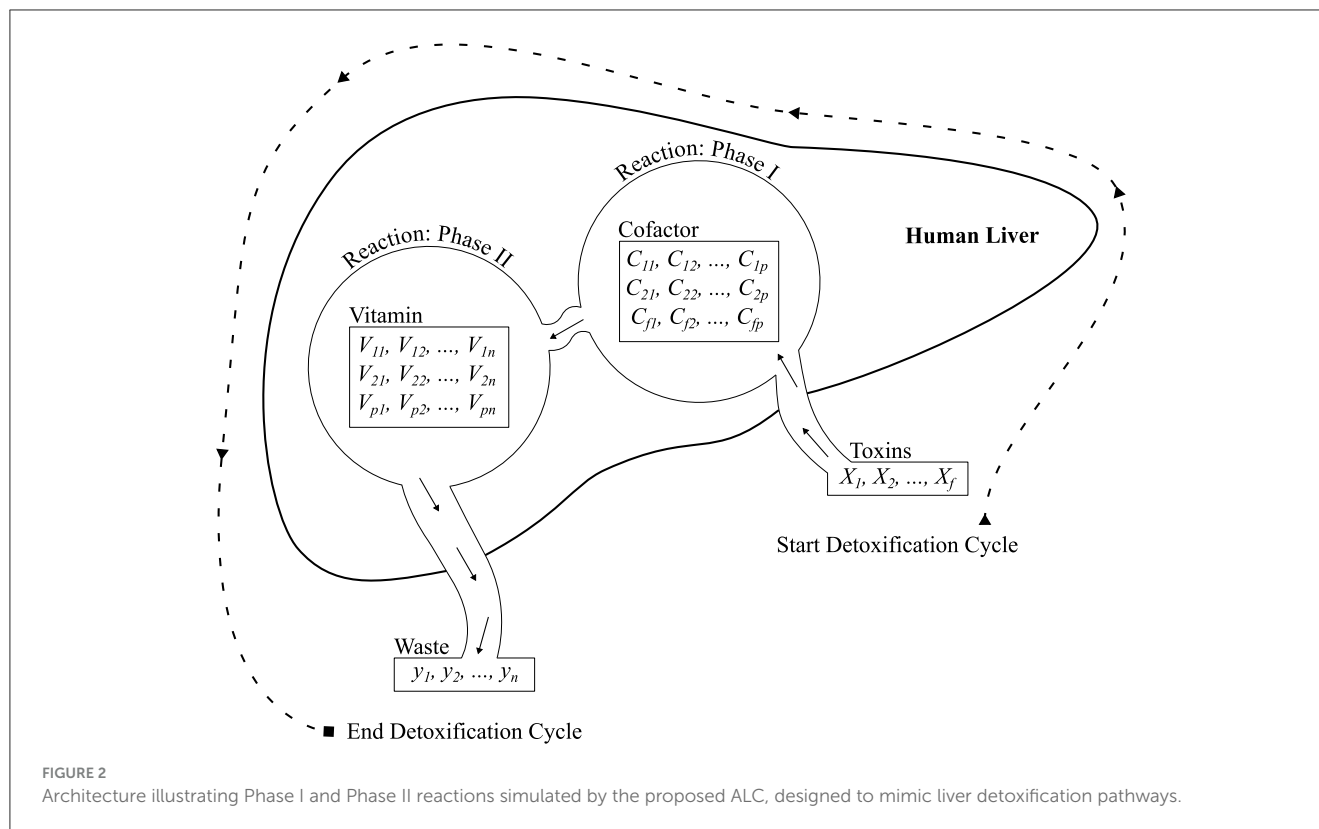
Algorithm 1. Artificial liver classifier (ALC).

randomly within the range $[-1, 1]$ and has dimensions (f, p), where f corresponds to the number of features in the input feature vector, and p is the number of lobules. The term $\frac{1}{fp} \sum_{k=1}^{f} \sum_{l=1}^{p} C_{kl}$ represents the mean of all elements in the cofactor matrix C that used to balance the reaction.

A human liver has a very large number of microscopic functional unit called lobules which are estimated to be of around 100,000 (Krebs et al., 2007). In our model the parameter p is an abstraction of a range of choice in lobular diversity, and the columns of the cofactor matrix C implement the various simulated lobular processing units. This structure allows introducing a spatial heterogeneity and an enzymatic variation, which can be witnessed in hepatic tissue, as part of the model. Although the biological premise of p can be that of a number of lobules, direct insertion of a figure like $p = 100,000$ in computation would be inefficient and also inconvenient because it has a high dimensionality and costs more computations. Thus, we introduce a practical range $f \leq p < 100,000$. This will ensure there is diversity, and that every input feature interacts with a number of simulated lobules ensuring the model is computationally straightforward. The parameter p in this range may be chosen by empirical methods of hyperparameter tuning, a compromise between the richness of the representation, and efficiency. It follows that the range is not arbitrary, but rather biological based on biological modeling under constrains. However, the reactive toxins (A) must be activated to enhance their reactivity before progressing to phase II. This activation involves eliminating all negative values, effectively transforming them to zero while retaining only the positive values. This process is mathematically expressed by the following equation:

$$A' = \max(0, A) \qquad (2)$$

where A' is the activated toxins matrix. However, Equation 2 uses ReLU activation to imitate the biological selectivity that occurs in Phase I detoxification in which only reactive (positive) products pass to the next stage. This selection eliminates the non-reactive

FIGURE 2
Architecture illustrating Phase I and Phase II reactions simulated by the proposed ALC, designed to mimic liver detoxification pathways.

outputs keeps the computational efficiency and provides the non-linearity necessary in the downstream processing.

Phase II: involves the conjugation of modified compounds from phase I with water-soluble molecules to make them excretable. This phase reduces the toxicity of compounds and facilitates their elimination from the body. this phase can be mathematically modeled using Equation 1, but with key differences. Instead of toxins, the matrix A' is used as input, representing the modified compounds (activated toxins) generated in phase I. Additionally, a matrix referred to as the vitamin matrix V is employed in place of the cofactor matrix C. This matrix is initialized randomly within the range $[-1, 1]$ and has dimensions (p, n).

$$B_{ji} = \frac{1}{n} \sum_{k=1}^{n} (A_{jk} \cdot V_{ki}) + \frac{1}{pn} \sum_{k=1}^{p} \sum_{l=1}^{n} V_{kl} \qquad (3)$$

where B_{ji} represents the conjugated compounds and $\frac{1}{pn} \sum_{k=1}^{p} \sum_{l=1}^{n} V_{kl}$ represents the mean of all elements in the vitamin matrix V.

Lastly, when the reactions in Phase I and Phase II are finished, detoxification is then complete. The outcome is some less dangerous and water-soluble wastes that can be removed by means of bile, urine, stool, etc. The softmax activation function is used as a model of the elimination process: not only does it allow formulating a probabilistic output for each of the classes, but also captures the selective and competitive characteristic of biological excretion (Bridle, 1990; Arora et al., 2020; Maharjan et al., 2020). Many detoxified compounds simultaneously compete to be eliminated in the liver depending on issues such as solvency,

the availability of transporters as well as priorities at the cellular level. Softmax reflects this behavior by placing higher probabilities on those compounds that are most dominant, or easily excreted, and thereby simulates the preferential clearance mechanism of the body.

$$B'_i = \frac{e^{B_i}}{\sum_{j=1}^{n} e^{B_j}} \qquad (4)$$

where B'_i represents the normalized probability for output class i.

The Algorithm 1 and Figure 2 describes the architecture of the proposed ALC. First, the cofactor and vitamin matrices are initialized randomly, where these matrices are defined based on the dimensions corresponding to the number of features (f), number of lobules (p), and number of output classes (n). Next, the IFOX, as presented in Algorithm 2, is configured, specifying the number of detoxification cycles (maximum number of training epochs) and detoxification power (maximum number of fox agents). The IFOX then optimizes the cofactor and vitamin matrices by minimizing the reaction error (i.e., loss). Finally, the optimized cofactor and vitamin matrices, resulting from this process, are subsequently used together with the toxin inputs (feature vectors) to predict the output classes.

Furthermore, the flowchart visualized the proposed ALC is presented in Figure 3. Additionally, the source code for the implementation of the proposed ALC can be accessed at the following repository: https://github.com/mwdx93/alc, which includes the main ALC implementation, training scripts, and example datasets.

4.2.2 Training algorithm

The FOX, developed by Mohammed and Rashid in 2022, mimics the hunting behavior of red foxes by incorporating physics-based principles. These include prey detection based on sound and distance, agent's jumping during the attack governed by gravity, and direction, as well as additional computations such as timing and walking (Mohammed and Rashid, 2022; Jumaah et al., 2024). These features make FOX a competitive optimization algorithm, outperformed several methods such as particle swarm optimization (PSO) and fitness dependent optimizer (FDO). The FOX is works as follows: Initially, the ground is covered with snow, requiring the fox agent to search randomly for its prey. During this random search, the fox agent uses the Doppler effect to detect and gradually approach the source of the sound. This process takes time and enables the fox agent to estimate the prey's location by calculating the distance. Once the prey's position is determined, the fox agent computes the required jump to catch it. Additionally, the search process is facilitated through controlled random walks, ensuring the fox agent progresses toward the prey while maintaining an element of randomness. The FOX balances exploitation and exploration phases statically, with a 50% probability for each (Aula and Rashid, 2025). Thus, the FOX operates as follows:

1. Computing the distance D_i of sound travel using the best position and random time:

$$D_i = \frac{BestPosition}{T_i} \times T_i \qquad (5)$$

Where T_i is a random time in $[0, 1]$ and i is the fox agent.

2. Determining the distance between the fox agent and its prey:

$$DF_i = 0.5 \times D_i \qquad (6)$$

3. Computing the jump J_i by multiplying half of the gravity acceleration constant with half squared mean of the time:

$$J_i = 0.5 \times 9.81 \times 0.5 \times (\sum_0^n T_i)^2 \qquad (7)$$

4. Updating the fox agent's position based on a directional equation, either northward $c_1 = 0.18$ or in the opposite direction $c_2 = 0.82$ based on the the jump probability p in $[0, 1]$.

$$X_{i+1} = DF_i \times J_i \times \begin{cases} c_1, & \text{if } p > 0.18 \\ c_2, & \text{otherwise} \end{cases} \qquad (8)$$

5. The following equation used for exploration:

$$X_{i+1} = BestPosition \times rand(1, dim) \times Mint \times a \qquad (9)$$

where dim is the problem dimension, $Mint$ is the minimum time iteratively updated based on T_i, a is an adjustment parameter computed as: $2 \times (it - (\frac{1}{Max_{it}}))$, and it is the current iteration.

However, the FOX has some limitations in its design. These limitations were acknowledged by the author of FOX (Mohammed and Rashid, 2022), while others have been identified through further analysis. For instance, one notable drawback is its static

```
   Input: Maximum number of epochs epochs, maximum
          number of fox agents max_fa
   Output: BestX and BestFitness
 1: Initialize the fox agents population X_fa  (fa =
    1, 2, 3, ..., max_fa)
 2: Initialize BestX, BestFitness
 3: while it < epochs do
 4:    for all fa ∈ FAs do
 5:        f ← Fitness(X_fa)
 6:        if f < BestFitness then
 7:           BestFitness ← f
 8:           BestX ← X_fa
 9:        end if
10:    end for
11:    α_min ← 1/(2×epochs)
12:    α ← α_min + (1 − α_min) × (1 − it/epochs)
13:    t ← 0.5 × μ(rand(0, 1, size(BestX)))
14:    Jump ← 4.905 × t²
15:    for all fa ∈ FAs do
16:        β ← rand(−α, α, size(BestX))
17:        if rand(0, 1) < α then
18:           X_fa ← BestX + β × α
19:        else
20:           X_fa ← 0.5 × BestX × (β×α)/Jump
21:        end if
22:    end for
23:    it ← it + 1
24: end while
```

Algorithm 2. IFOX: new variation of FOX optimization algorithm.

approach to balancing exploration and exploitation. This paper aims to address these limitations by proposing a new variation of the FOX called IFOX to make it integrable with the proposed ALC as a training algorithm to optimize the cofactor and vitamin matrices. For reference, the implementation of the FOX can be accessed at https://github.com/hardi-mohammed/fox.

The IFOX, as visualized in Algorithm 2, incorporates several improvements over the FOX. First, it transforms the balance between exploitation and exploration into a dynamic process using the ϵ-greedy method, rather than a static approach (Liu et al., 2020; Abdalrdha et al., 2023). This dynamic adjustment is controlled by the parameter α, which decreases progressively as the optimization process iterate. Second, the computation of distances is eliminated in favor of directly using the best position, facilitated by the parameter β, derived from α. This modification simplifies the FOX by removing Equations 5, 6, and simplifying Equations 8 by eliminating the probability parameter p and the directional variables (c_1 and c_2). Third, in Equations 9, the variables a and $Mint$ are excluded.

5 Results

This section presents the performance results of the proposed ALC on multiple benchmark datasets, as described in Section 4.1. The experimental parameter settings were configured for

FIGURE 3
Flowchart of the proposed ALC.
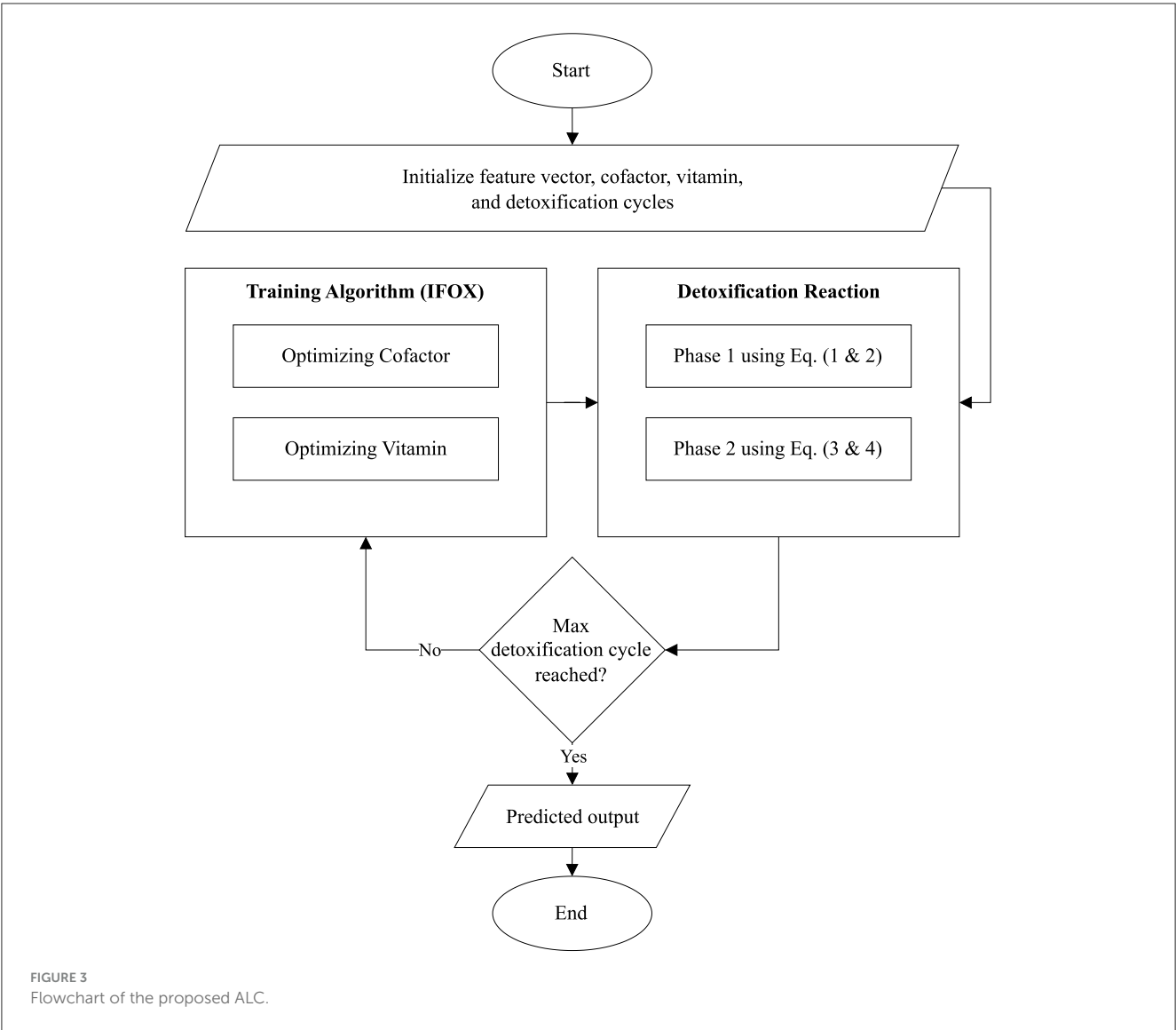
TABLE 1 Comparison of best objective values obtained by the proposed IFOX algorithm across selected CEC2019 benchmark functions over 30 independent runs.

Function/optimizer	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
GOOSE	1.8E12	6.8E3	13.70	1.60E3	6.09	4.79	274.35	5.57	3.81	20.98
ANA	-	4.00	13.70	38.50	1.22	-	116.59	5.47	2.00	2.71
LEO	7.3E9	17.47	12.70	69.86	2.76	3.01	195.56	5.06	3.26	20.01
FDO	4585.27	4.00	13.7	34.08	2.13	12.13	120.40	6.1	2.01	2.71
FOX	1.00	4.72	9.88	147.21	5.13	298.10	1.017	1.38	1.41	21.49
IFOX	1.00	4.61	2.42	35.80	1.88	1.00	1.00	1.24	1.33	20.99

The results are compared with those of FOX and several recent optimization algorithms. Lower objective values indicate better performance.

each dataset as follows: 500 detoxification cycles, a detoxification power of 10, and dataset-specific numbers of lobules. Specifically, the number of lobules was set to 10 for Iris Flower and Breast Cancer Wisconsin, 15 for Wine and Voice Gender, and 50 for MNIST. The choice of these values was done through a systematic empirical search, where a predetermined set of possible values

has been considered, a validation based approach used. Different values of p were searched over each dataset and the value that provided the maximum average classification accuracy on a held-out validation split was selected. This strategy makes sure that the hyperparameters being chosen are tuned in a reproducible and performance-based fashion. In order to have an efficient and fair

FIGURE 4
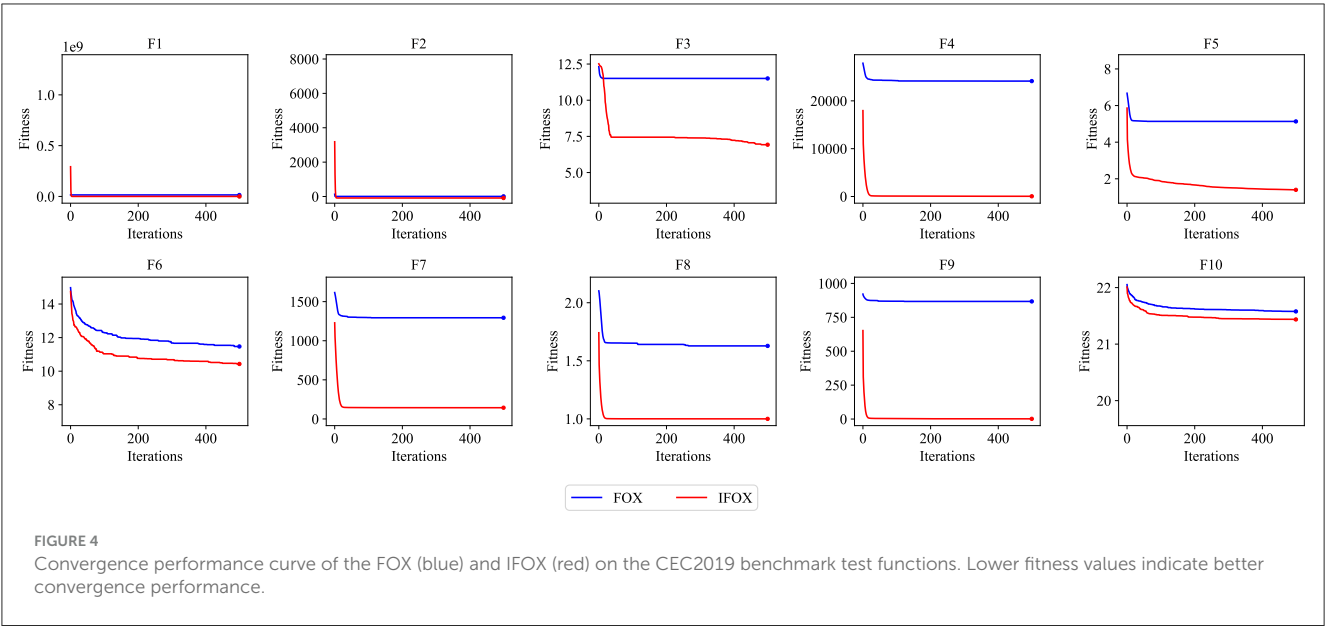Convergence performance curve of the FOX (blue) and IFOX (red) on the CEC2019 benchmark test functions. Lower fitness values indicate better convergence performance.

TABLE 2 Function-wise ranks, total rank, and average rank for each optimization algorithm across the 10 selected CEC2019 benchmark functions.

Function/optimizer	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	Total rank	Avg. rank
IFOX	1	3	1	2	2	1	1	1	1	5	18	1.8
FDO	3	1	4	1	3	4	4	6	4	1	31	3.1
ANA	6	1	4	3	1	6	3	4	3	1	32	3.2
FOX	1	4	2	5	5	5	2	2	2	6	34	3.4
LEO	4	5	3	4	4	2	5	3	5	3	38	3.8
GOOSE	6	6	4	6	6	3	6	5	6	4	52	5.2

Lower ranks indicate better performance.

assessment on the performance of the models, cross-validation was used on the basis of k. In particular, each of the datasets has been split into k equal size folds (10-folds were used in our experiments), and the test has been repeated on every fold. Mean results of each of the runs were used to compute final performance metrics. To facilitate later comparison and analysis, additional classifiers, including MLP, SVM, LR, and XGBoost (XGB), were executed on the same datasets. However, all experiments were conducted on an MSI GL63 8RD laptop equipped with an Intel® Core™ i7-8750H × 12 processor and 32 GB of memory. This consistent setup ensured a robust evaluation of the proposed ALC alongside the other classifiers under the same conditions.
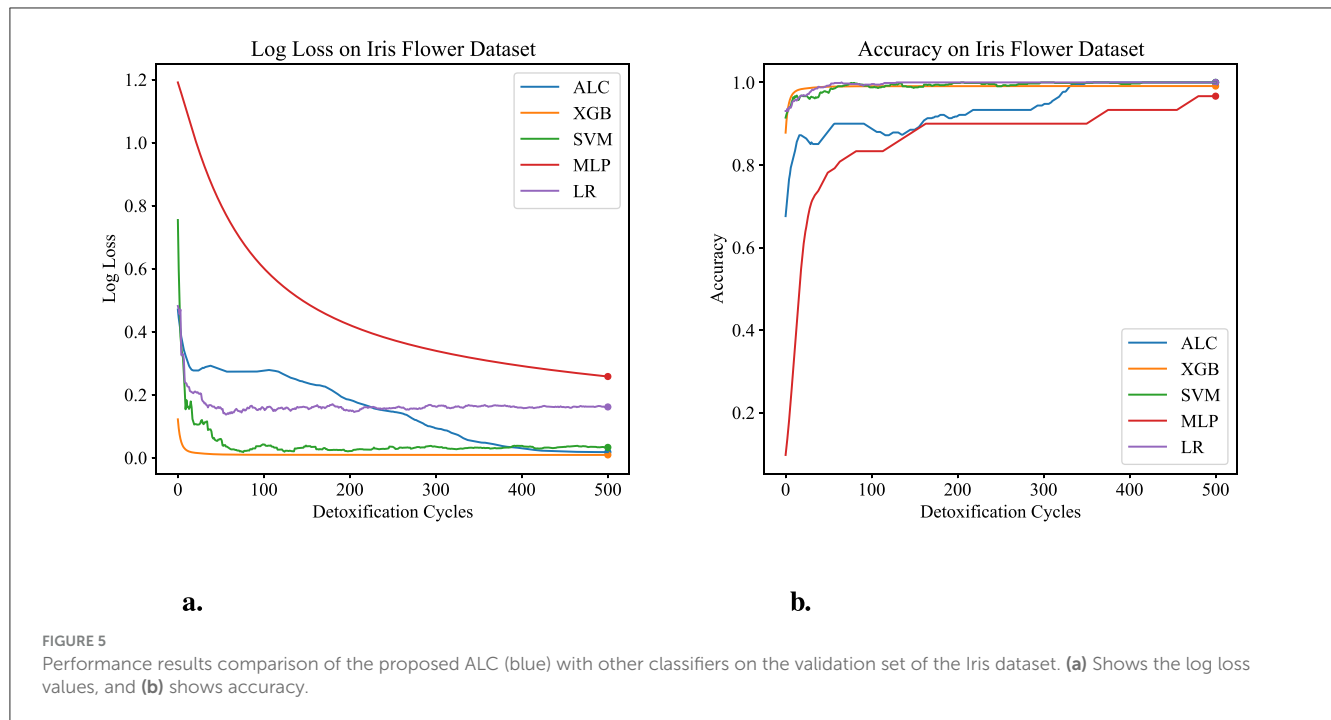
5.1 Performance metrics

To evaluate the performance of the proposed ALC, several metrics were employed, including log loss (cross-entropy loss), accuracy, precision, recall, F1-score, and training time. Initially, Log loss (Equation 10) quantifies the divergence between predicted probabilities and actual labels, where lower values indicate better predictive performance (Xue et al., 2023). The accuracy (Equation 11) measures the proportion of correctly classified instances, serving as a straightforward indicator of overall

TABLE 3 Cross-validation performance of the proposed ALC and other classifiers on the Iris Flower dataset (mean over 10-folds).

Metric	ALC	XGB	SVM	MLP	LR
Loss	0.0169	0.0085	0.0691	0.2417	0.0543
Accuracy	1.0000	1.0000	1.0000	0.9667	1.0000
Precision	1.0000	1.0000	1.0000	0.9694	1.0000
Recall	1.0000	1.0000	1.0000	0.9667	1.0000
F1-Score	1.0000	1.0000	1.0000	0.9664	1.0000
Overfitting	-0.0231%	-0.0144%	-0.0384%	-0.0689%	-0.0415%
Time (sec.)	2.12	0.91	4.42	4.18	4.31

correctness. Moreover, precision (Equation 12) evaluates the proportion of true positives among all positive predictions, emphasizing the model's ability to reduce false positives. In contrast, recall Equation 13 focuses on the proportion of true positives among all actual positive instances, highlighting the importance of minimizing false negatives. Furthermore, the F1-score (Equation 14), as the harmonic mean of precision and recall, provides a balanced assessment when class distributions are imbalanced (Naidu et al., 2023). Moreover, the overfitting gap

Performance results comparison of the proposed ALC (blue) with other classifiers on the validation set of the Iris dataset. **(a)** Shows the log loss values, and **(b)** shows accuracy.

defined as the difference between training and validation accuracy, provides insights into generalization. A smaller value indicate better generalization, while a larger value indicates overfitting, where the model excels on the training set but struggles with unseen data. Finally, the training time reflects the duration required to train the model, offering insight into its computational efficiency.

$$c\text{Log Loss} = -\frac{1}{n}\sum_{i=1}^{n}\left(y_i\log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\right) \quad (10)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (11)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (12)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (13)$$

$$\text{F1-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right) \quad (14)$$

where TP, TN, FP, and FN represent the true positive, true negative, false positive, and false negative counts, respectively. Additionally, y denotes the actual labels, while $\hat{y}$ represents the predicted labels.
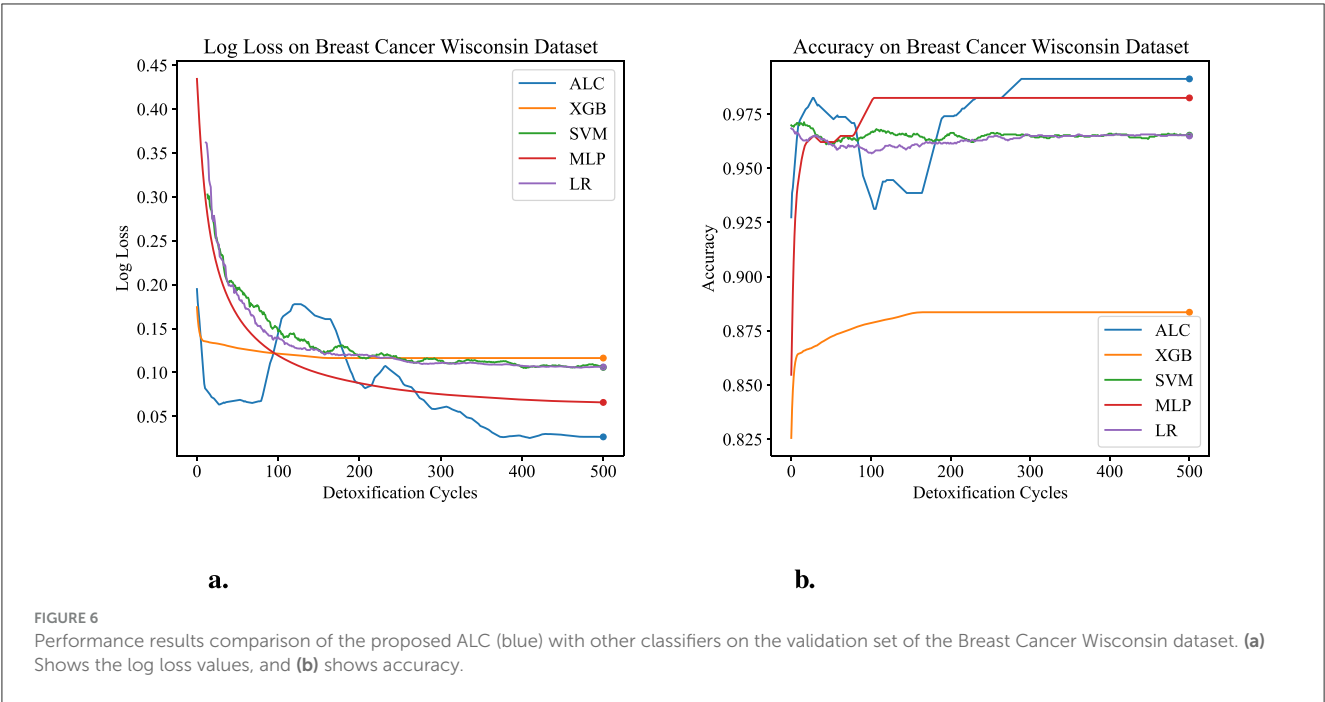
5.2 Convergence result of the training algorithm IFOX

Table 1 provides the empirical result of the convergence behavior of IFOX based on selected benchmark functions of the CEC2019 suite. Unlike the original FOX, IFOX invariably reaches lower average objective values in most of the test functions, with vastly less variance and greater stability, especially on high-dimensional and multimodal functions like F3 through F7.

TABLE 4 Cross-validation performance of the proposed ALC and other classifiers on the Breast Cancer Wisconsin dataset (mean over 10-folds).

Metric	ALC	XGB	SVM	MLP	LR
Loss	0.0261	0.1132	0.1203	0.0682	0.1319
Accuracy	0.9932	0.8927	0.9833	0.9833	0.9559
Precision	0.9943	0.9233	0.9833	0.9833	0.9638
Recall	0.9932	0.9233	0.9833	0.9833	0.9559
F1-Score	0.9932	0.9260	0.9833	0.9833	0.9631
Overfitting	-0.0029%	0.1178%	0.0213%	0.0086%	0.0237%
Time (sec.)	3.62	1.09	3.79	4.53	3.81

Furthermore, four recent and competitive optimizers were added to the test in order to prove the effectiveness of IFOX over its predecessor. They consist of GOOSE, ant nesting algorithm (ANA), lagrange elementary optimization (LEO), and FDO, which have shown rather good performance in the literature (Hamad and Rashid, 2024; Hama Rashid et al., 2021; Aladdin and Rashid, 2025; Abdullah and Ahmed, 2019). The relative performance assures that IFOX attains better convergence characteristics and final solution quality most of the time. Moreover, the gain noticed in performance increase is explained based on adding adaptive inertia control and better search dynamics in IFOX. It is a design that puts an emphasis on more extensive exploration during the initial phases and more targeted exploitation as optimization goes on. Additionally, convergence curves in Figure 4 demonstrate that IFOX converges more rapidly compared to the original FOX and, in addition, obtains better final values. This can be seen by the fact that the fitness trajectory flattened very fast as compared to other methods. Although formal theoretical demonstration of convergence is

FIGURE 6
Performance results comparison of the proposed ALC (blue) with other classifiers on the validation set of the Breast Cancer Wisconsin dataset. **(a)** Shows the log loss values, and **(b)** shows accuracy.

outside the scope of this paper, the similar results in independent runs (30 runs) give good evidence of the soundness and reliability of IFOX in varied and different optimization landscapes.

In order to compare performance on the whole, each optimization algorithm was ranked by the number of functions with the best objective value. The overall and the average ranking of all functions is given in Table 2. IFOX presented a minimum total (18) and average (1.8) rank and maintained a better performance as compared to the rest of the optimization algorithms. Due to IFOX's superior convergence and stability, it was chosen as the training algorithm in this study.

TABLE 5 Cross-validation performance of the proposed ALC and other classifiers on the Wine dataset (mean over 10-folds).

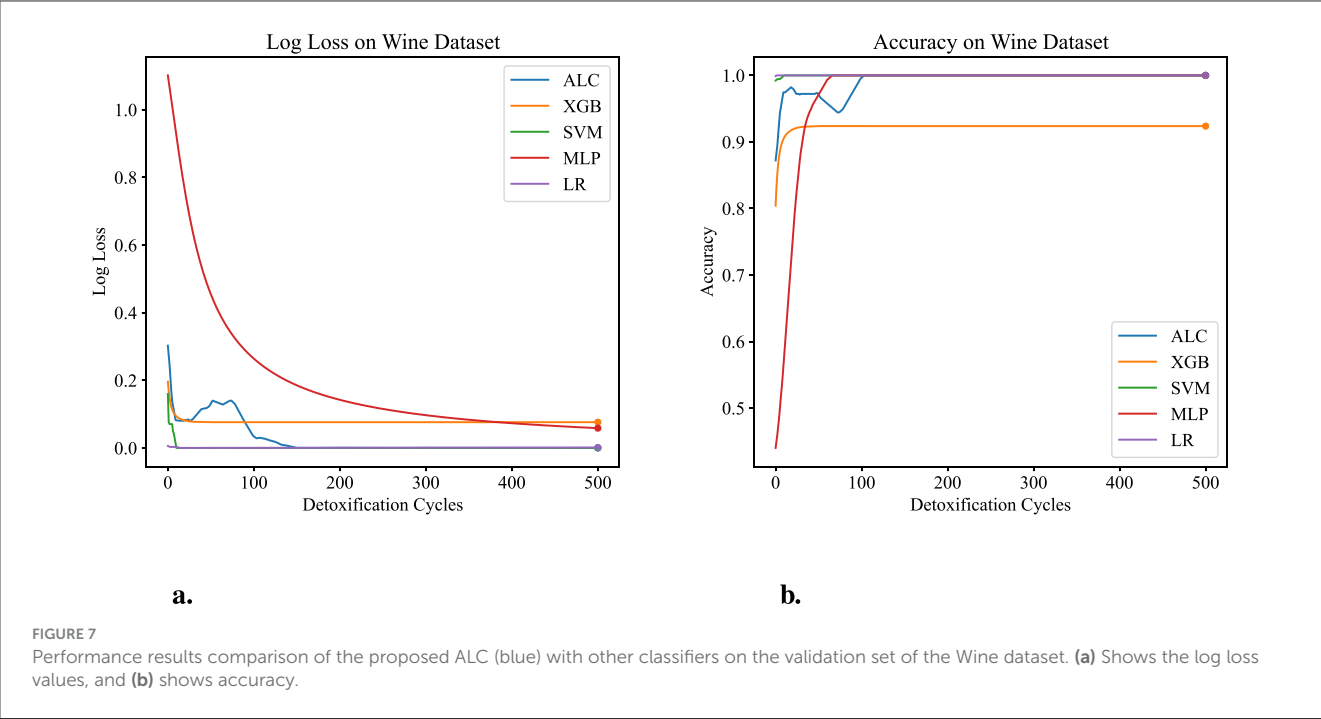| Metric	ALC	XGB	SVM	MLP	LR
Loss	0.0011	0.0691	0.0001	0.0568	0.0012
Accuracy	1.0000	0.9258	1.0000	1.0000	1.0000
Precision	1.0000	0.9534	1.0000	1.0000	1.0000
Recall	1.0000	0.9424	1.0000	1.0000	1.0000
F1-Score	1.0000	0.9429	1.0000	1.0000	1.0000
Overfitting	0.0000%	0.0675%	0.0000%	-0.0068%	0.0000%
Time (sec.)	2.38	1.17	3.89	3.89	3.92

5.3 Experimental results of ALC

The performance results of the proposed ALC are presents through this subsection, summarized in the figures and tables. Additionally, comparisons with other classifiers, including MLP, SVM, LR, and XGB, have been conducted on the five datasets described in Section 4.1. Table 3 presents the performance results of the proposed ALC and other classifiers on the Iris Flower dataset. Additionally, Figures 5a, b show the loss and accuracy, respectively, across the validation folds. The proposed ALC achieved 100% accuracy with a loss of 0.0169, an overfitting gap of -0.0231%, and a training time of 2.12 s. The XGB also achieved 100% accuracy with a loss of 0.0085, an overfitting gap of -0.0144%, and a training time of 0.91 s. Similarly, the SVM reached 100% accuracy with a loss of 0.0704, an overfitting gap of -0.0384%, and a training time of 4.42 s. The MLP attained 96.67% accuracy with a loss of 0.2417, an overfitting gap of -0.0714%, and a training time of 4.18 s. Lastly, the LR reached 100% accuracy with a loss of 0.0543, an overfitting gap of -0.0415%, and a training time of 4.31 s.

Table 4 presents the performance results of the proposed ALC and other classifiers on the Breast Cancer Wisconsin dataset. Figures 6a, b display the loss and accuracy, respectively, on the validation folds. The proposed ALC achieved 99.12% accuracy, with a loss of 0.0261 and an overfitting gap of -0.0029%, with a training time of 3.62 s. The XGB achieved 88.36% accuracy, with a loss of 0.1132 and an overfitting gap of 0.1178%, with a training time of 1.09 s. The SVM reached 98.25% accuracy, with a loss of 0.1105 and an overfitting gap of 0.0213%, with a training time of 3.79 s. The MLP achieved 98.25% accuracy, with a loss of 0.0682 and an overfitting gap of 0.0086%, with a training time of 4.53 s. Lastly, the LR achieved 96.49% accuracy, with a loss of 0.1319 and an overfitting gap of 0.0237%, with a training time of 3.81 s.

Table 5 presents the performance results of the proposed ALC and other classifiers on the Wine dataset. Figures 7a, b display the loss and accuracy, respectively, on the validation folds. The proposed ALC achieved 100% accuracy, with a loss of 0.0011 and an overfitting gap of 0.0000%, with a training time of 2.38 s.

FIGURE 7
Performance results comparison of the proposed ALC (blue) with other classifiers on the validation set of the Wine dataset. **(a)** Shows the log loss values, and **(b)** shows accuracy.

The XGB achieved 92.38% accuracy, with a loss of 0.0691 and an overfitting gap of 0.0675%, with a training time of 1.17 s. The SVM achieved 100% accuracy, with a loss of 0.0001 and an overfitting gap of 0.0000%, with a training time of 3.89 s. The MLP achieved 100% accuracy, with a loss of 0.0568 and an overfitting gap of -0.0068%, with a training time of 3.89 s. Lastly, the LR achieved 100% accuracy, with a loss of 0.0012 and an overfitting gap of 0.0000%, with a training time of 3.92 s.

Table 6 presents the performance results of the proposed ALC and other classifiers on the Voice Gender dataset. Figures 8a, b display the log loss and accuracy, respectively, on the validation folds. The proposed ALC achieved 97.63% accuracy, with a loss of 0.0613 and an overfitting gap of 0.0004%, with a training time of 3.21 s. The XGB achieved 92.79% accuracy, with a loss of 0.0706 and an overfitting gap of 0.0601%, with a training time of 1.19 s. The SVM achieved 97.32% accuracy, with a loss of 0.1932 and an overfitting gap of 0.0043%, with a training time of 4.35 s. The MLP achieved 98.26% accuracy, with a loss of 0.0622 and an overfitting gap of -0.0036%, with a training time of 12.59 s. Lastly, the LR achieved 98.11% accuracy, with a loss of 0.0601 and an overfitting gap of -0.0063%, with a training time of 4.61 s.

Table 7 presents the performance results of the proposed ALC and other classifiers on the MNIST dataset. Figures 9a, b display the log loss and accuracy, respectively, on the validation set. The proposed ALC achieved 99.75% accuracy on the validation set, with a loss of 0.0000 and an overfitting gap of 0.0025%, with a training time of 6.18 s. The XGB achieved 94.05% accuracy, with a loss of 0.0581 and an overfitting gap of 0.0571%, with a training time of 2.35 s. The SVM achieved 99.50% accuracy, with a loss of 0.0076 and an overfitting gap of 0.0050%, with a training time of 5.38 s. The MLP achieved 99.00% accuracy, with a loss of 0.0473 and an overfitting gap of 0.0100%, with a training time of 5.22 s. Lastly, the LR achieved 99.50% accuracy, with a loss of

TABLE 6 Cross-validation performance of the proposed ALC and other classifiers on the Voice dataset (mean over 10-folds).

Metric	ALC	XGB	SVM	MLP	LR
Loss	0.0613	0.0706	0.1932	0.0622	0.0601
Accuracy	0.9752	0.9164	0.9750	0.9800	0.9811
Precision	0.9750	0.9325	0.9736	0.9811	0.9811
Recall	0.9752	0.9105	0.9750	0.9800	0.9811
F1-Score	0.9752	0.9150	0.9750	0.9811	0.9811
Overfitting	0.0004%	0.0601%	0.0043%	-0.0036%	-0.0063%
Time (sec.)	3.21	1.19	4.35	12.59	4.61

0.0137 and an overfitting gap of 0.0050%, with a training time of 5.61 s.

In summary, the proposed ALC outperformed or matched other classifiers across the datasets tested, including Iris, Breast Cancer Wisconsin, Wine, Voice Gender, and MNIST. The results demonstrated superior loss, accuracy, precision, recall, and F1-scores, highlighting the reliability and generalization of the proposed ALC in achieving high classification performance. Furthermore, the proposed ALC exhibited minimal overfitting and efficient training times compared to other classifiers. However, the next section will provide a detailed analysis and interpretation of these results, and shedding light on limitations and imperfections of the proposed ALC.

6 Discussion

The results presented in Section 5.3, derived from experiments conducted on the datasets described in Section 4.1, highlight
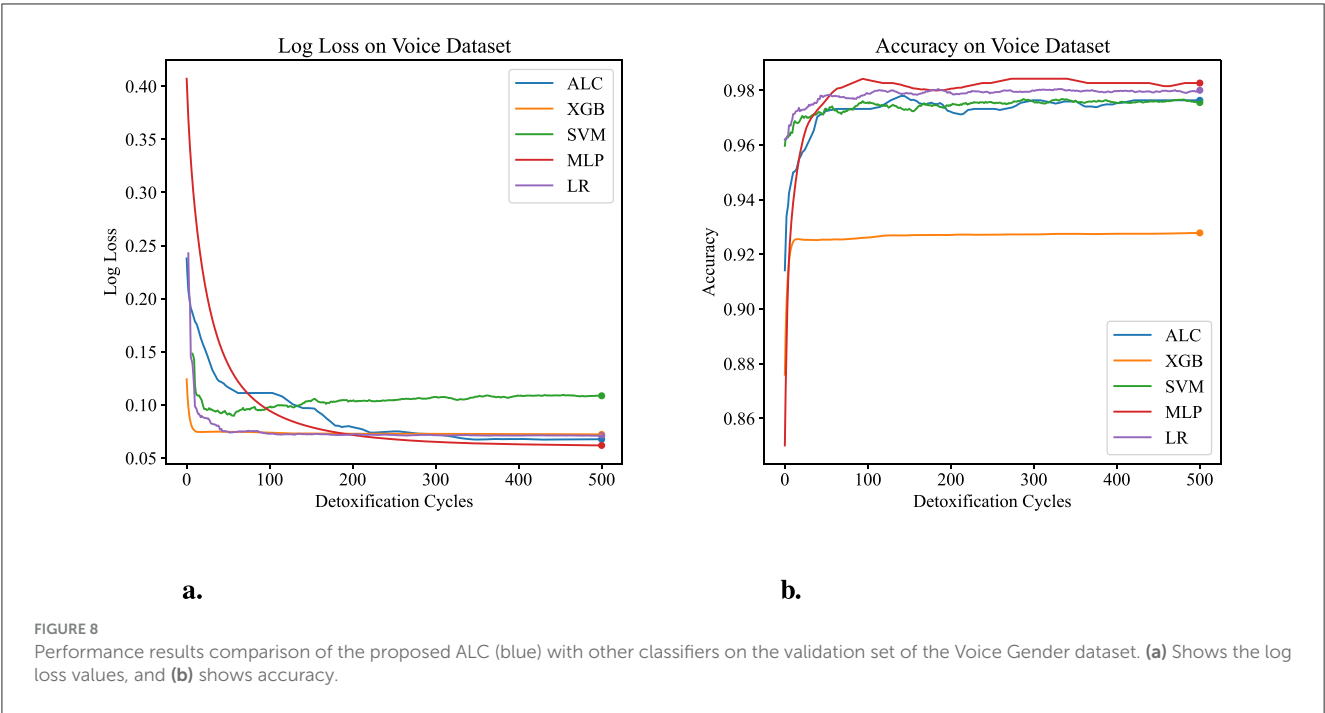
FIGURE 8
Performance results comparison of the proposed ALC (blue) with other classifiers on the validation set of the Voice Gender dataset. **(a)** Shows the log loss values, and **(b)** shows accuracy.

TABLE 7 Cross-validation performance of the proposed ALC and other classifiers on the MNIST dataset (mean over 10-folds).

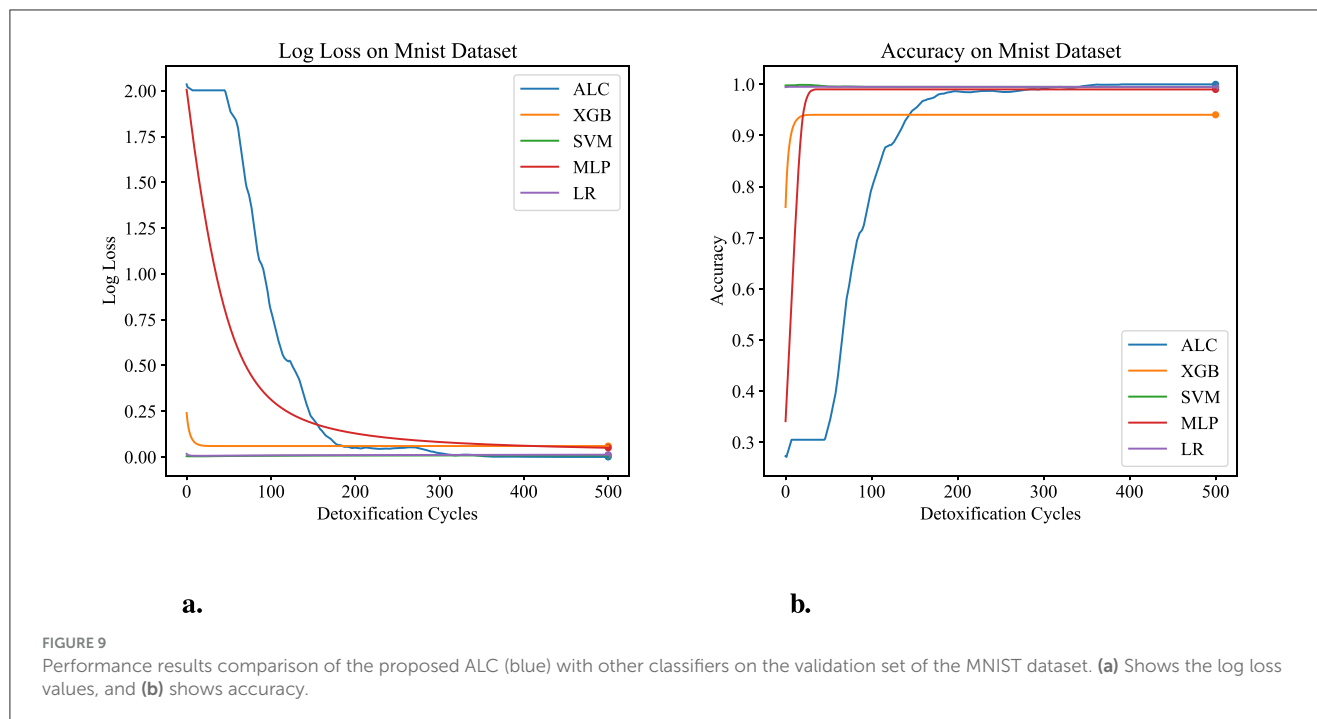| Metric	ALC	XGB	SVM	MLP	LR
Loss	0.0000	0.0581	0.0076	0.0473	0.0137
Accuracy	0.9975	0.9421	0.9967	0.9900	0.9967
Precision	0.9970	0.9828	0.9953	0.9906	0.9953
Recall	0.9967	0.9802	0.9967	0.9900	0.9967
F1-Score	0.9987	0.9800	0.9967	0.9900	0.9967
Overfitting	0.0025%	0.0571%	0.0050%	0.0100%	0.0050%
Time (sec.)	6.18	2.35	5.38	5.22	5.61

the superior performance of the proposed ALC compared to other classifiers. However, a more in-depth statistical analysis is necessary, particularly of the validation set results, as they are considered more reliable indicators of classifier performance due to being obtained from unseen data. The statistical analysis presented in Table 8 compare the performance of the proposed ALC with four classifiers—XGB, SVM, MLP, and LR—across the five datasets described in Section 4.1. The analysis focuses on four metrics: loss, accuracy, overfitting gap, and training time, with statistical significance determined using the Wilcoxon signed-rank test at a threshold of P-value $<$ 0.05. The Wilcoxon signed-rank test is used to compare paired samples, particularly when data may not follow a normal distribution. It assesses whether the differences between paired observations are statistically significant (Hodges et al., 2022). Hence, this analysis results provide insights into the strengths of the

proposed ALC in terms of its generalization, accuracy, and computational efficiency.

The loss metric, which is a primary indicator of classifier generalizability, demonstrates that the proposed ALC outperforms other classifiers in several datasets. Specifically, in the Iris Flower dataset, the proposed ALC showed statistically significant improvements in loss compared to SVM (P = 0.012), MLP (P = 0.006), and LR (P = 0.037), while its performance was comparable to XGB (P = 0.432), indicating XGB outperforms the proposed ALC. Similarly, in the Breast Cancer Wisconsin dataset, the proposed ALC showed significant improvements over XGB (P = 0.004), SVM (P = 0.001), MLP (P = 0.015), and LR (P = 0.000). In the Wine dataset, the proposed ALC demonstrated significant improvements compared to XGB (P = 0.022), MLP (P = 0.005), and LR (P = 0.031), but did not show statistically significant with SVM (P = 0.893). These trends were consistent in more complex datasets like Voice Gender and MNIST, where the proposed ALC achieved lower loss values compared to other classifiers in most cases (P < 0.05). The findings indicate that the proposed ALC offers better generalization across these datasets of varying complexity.

In terms of accuracy, the differences between the proposed ALC and other classifiers were generally less pronounced, as reflected by P-values exceeding 0.05 in most datasets. Notable exceptions include the Voice Gender dataset, where the proposed ALC significantly outperformed SVM (P = 0.002), and the Breast Cancer Wisconsin dataset, where the proposed ALC showed an advantage over XGB (P = 0.015). Furthermore, additional accuracy comparisons were conducted with other models discussed in the related work Section 2, as presented in Table 9, demonstrating the superiority of the proposed ALC. These results suggest that while accuracy remains an important

Performance results comparison of the proposed ALC (blue) with other classifiers on the validation set of the MNIST dataset. **(a)** Shows the log loss values, and **(b)** shows accuracy.

metric, it may not always effectively differentiate the performance of classifiers, particularly when accuracy levels are already high across classifiers (Qu et al., 2022).

The overfitting gap metric, which evaluates the difference between the performance on training and validation folds, reveals that the proposed ALC demonstrates superior generalization. In most datasets, significant improvements were observed, such as in the Breast Cancer Wisconsin and MNIST datasets, where all *P*-values were < 0.05. In contrast, the overfitting gap in the Wine dataset showed inconsistent patterns, with *P*-values largely exceeding the significance threshold ($P = 0.500$). These results support the ability of the proposed ALC to reduce the risk of overfitting. Furthermore, the training time metric is used to measure the speed of classifiers. The statistical results suggest that the proposed ALC is competitive and efficient. The training time differed considerably (at least) in datasets that are smaller, including Iris Flower ($P < 0.01$), Breast Cancer Wisconsin ($P = 0.002$), Wine ($P = 0.002$), and voice Gender ($P = 0.001$). But on a bigger and more complicated dataset such as MNIST, the training time of the proposed ALC was similar to that of other classifiers ($P > 0.90$). The advocated ALC had no significant differences with XGB since it utilized tree-based models, which tend to have short processing times.

6.1 Computational complexity and ablation analysis

The complexity of the proposed ALC is mostly influenced by matrix manipulation and optimization procedure. Suppose that there are n input samples, f features, p lobules, o output classes,

and the number of iterations of the optimization (i.e., detoxification cycles) = I. The initial large step is the product of the input toxin matrix $X \in \mathbb{R}^{n \times f}$ by the cofactor matrix $C \in \mathbb{R}^{f \times p}$ in Phase I and takes $\mathcal{O}(nfp)$ time. This is then summed with an element-wise ReLU activation of the resultant matrix whose cost is $\mathcal{O}(np)$. During Phase II, a similar model of conjugation is treated as the second matrix multiplication involving the activated toxin matrix and the vitamin matrix $V \in \mathbb{R}^{npo}$, which would lead to time complexity of $\mathcal{O}(npo)$. This last elimination step runs the softmax on each of the n output vectors, and costs $\mathcal{O}(no)$. The training is based on IFOX that successively optimizes the cofactor and vitamin matrices. Suppose every iteration uses the entire dataset, training will hence have time complexity $\mathcal{O}(I \cdot n \cdot (fp + po))$. So, this term dominates the overall time complexity of the ALC when training. Moreover, in the space complexity, the model will need $\mathcal{O}(nf)$ storage of the input data, $\mathcal{O}(fp)$ memory to hold the cofactor matrix C, $\mathcal{O}(po)$ memory to hold the vitamin matrix V, and $\mathcal{O}(np + no)$ intermediate activation and outputs. Thus, the overall space complexity is $\mathcal{O}(nf + fp + po + np + no)$. The parameter matrices and batch level intermediate results consume the most memory. Hence, the ALC has a scalable architecture whose complexity scales linearly with size of input and size of optimization steps and quadratically with size of internal representation (lobules).

The ablation study results were summarized in Table 10, where the significance of each component of ALC was revealed. The complete model (Phase I + Phase II) represented the optimal result, reaching 99.12 percent accuracy and demonstrating small overfitting. Withdrawing Phase II or replacing Phase I output freedom with a constant value resulted in significant accuracy declines (95.20% and 91.45%, respectively), and this fact shows that both steps are needed. The replacement of the cofactor matrix C by random numbers or an identity vitamin matrix also lowered

TABLE 8 Wilcoxon signed-rank test results comparing classifier pairs on validation set metrics across all datasets.

ALC vs.	Dataset	P-value			
		Loss	Accuracy	Overfitting	Time
XGB	Iris Flower	0.432	0.872	0.481	0.493
SVM	Iris Flower	0.012	0.950	0.008	0.025
MLP	Iris Flower	0.006	0.951	0.021	0.001
LR	Iris Flower	0.037	0.042	0.029	0.017
XGB	Breast Cancer	0.004	0.015	0.007	0.970
SVM	Breast Cancer	0.001	0.009	0.013	0.064
MLP	Breast Cancer	0.015	0.011	0.012	0.004
LR	Breast Cancer	0.000	0.020	0.001	0.005
XGB	Wine	0.022	0.763	0.012	0.974
SVM	Wine	0.893	0.004	0.706	0.002
MLP	Wine	0.005	0.681	0.023	0.003
LR	Wine	0.031	0.822	0.748	0.002
XGB	Voice Gender	0.011	0.019	0.005	0.951
SVM	Voice Gender	0.000	0.002	0.007	0.001
MLP	Voice Gender	0.851	0.804	0.029	0.002
LR	Voice Gender	0.019	0.781	0.003	0.001
XGB	MNIST	0.001	0.014	0.004	0.993
SVM	MNIST	0.003	0.043	0.012	0.945
MLP	MNIST	0.001	0.017	0.008	0.936
LR	MNIST	0.002	0.041	0.009	0.898

TABLE 9 Performance comparison (accuracy metric) of the proposed ALC with models discussed in the related work.

Classifier	Dataset	Accuracy	Ref.
ALC	Iris Flower	**1.0000**	Proposed
SVM		0.9600	Fan et al., 2024
ALC	Breast Cancer	0.9932	Proposed
RRNN		**0.9951**	Rajeswari and Sakthi Priya, 2025
ALC	Wine	**1.0000**	Proposed
SVM		0.8790	Cortez et al., 2009
MR		0.8645	Cortez et al., 2009
ANN		0.8675	Cortez et al., 2009
SVM		0.9830	Fan et al., 2024
ALC	Voice Gender	**0.9752**	Proposed
MLP		0.9674	Buyukyilmaz and Cibikdiken, 2016
ALC	MNIST	**0.9975**	Proposed
SVC		0.9780	Xiao et al., 2017
DT		0.8860	Xiao et al., 2017
KNN		0.9590	Xiao et al., 2017
MLP		0.9720	Xiao et al., 2017
OPIUM		0.9590	Cohen et al., 2017

Bold values indicate that the corresponding method achieved superior results compared to other competing algorithms.

performance indicating the need to learn both matrices. These findings demonstrate that all its components play a significant role in the work of the proposed ALC as a whole.

6.2 Failure case analysis

Although the proposed ALC can deliver good results irrespective of the data encountered, a few limitations can be associated with it, based on an application during certain situations. ALC does not utilize any mini-batch training mechanism, e.g., stochastic gradient descent (Wojtowytsch, 2023), which is normally applied to large-scale learning to minimize computing costs. This consequence can cause longer runtimes in full-batch training working with mass data. There is also slower convergence in the model in that it uses the stochastic IFOX that does not directly optimize training error when applied to cofactor and vitamin matrices. It could influence either the stability or efficiency of convergence. From a model behavior perspective, ALC might fail to perform well on datasets with poor non-linear structure, noisy or sparse features, or extreme class skew, where the biological metaphor might not find any useful patterns. In addition to that, errors can be propagated and replicated by the sequential dependency between Phase I and Phase II. These constraints

point to the directions of further research, such as utilizing mini-batch techniques, improving the IFOX, implementing hybrid optimization schemes, or reorganizing the bio-chemical paradigm to be less rigid and more flexible.

7 Conclusions

In conclusion, this paper suggests a novel supervised learning classifier, termed the artificial liver classifier (ALC), inspired by the human liver's detoxification function. The ALC is easy to implement, fast, and capable of reducing overfitting by simulating the detoxification function through straightforward mathematical operations. Furthermore, it introduces an improvement to the FOX optimization algorithm, referred to as IFOX, which is integrated with the ALC as training algorithm to optimize parameters effectively. Furthermore, the ALC was evaluated on five benchmark machine learning datasets: Iris Flower, Breast Cancer Wisconsin, Wine, Voice Gender, and MNIST. The empirical results demonstrated its superior performance compared to support vector machines, multilayer perceptron, logistic regression, XGBoost and other established classifiers. Despite these superiority, the ALC has limitations, such as longer training times on large datasets and slower convergence rates, which could be addressed in future work using methods like mini-batch training or parallel processing. Finally, this paper underscores the potential of biologically inspired models and encourages researchers to simulate natural functions to develop more efficient and powerful machine learning models.

TABLE 10 Ablation study results showing the contribution of Phase I, Phase II, and the respective matrices in the proposed ALC on the Breast Cancer Wilcoxon dataset.

Model variant	Accuracy	Loss	F1-Score	Overfitting	Time (sec.)
Full ALC (phase I + phase II)	99.12%	0.0261	0.9932	-0.0029%	3.62
Phase I only	95.20%	0.0745	0.9517	0.0121%	2.58
Phase II only	91.45%	0.1123	0.9140	0.0450%	2.89
Random cofactor matrix	88.36%	0.1341	0.8823	0.0663%	2.51
Identity vitamin matrix	93.62%	0.0897	0.9312	0.0387%	2.73

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

MJ: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. YA: Project administration, Supervision, Validation, Writing – review & editing. TR: Conceptualization, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Gen AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdalrdha, Z. K., Al-Bakry, A. M., and Farhan, A. K. (2023). "CNN hyper-parameter optimizer based on evolutionary selection and gow approach for crimes tweet detection," in *2023 16th International Conference on Developments in eSystems Engineering (DeSE)* (Cham: IEEE). doi: 10.1109/DeSE60595.2023.10469361

Abdullah, J. M., and Ahmed, T. (2019). Fitness dependent optimizer: inspired by the bee swarming reproductive process. *IEEE Access* 7, 43473–43486. doi: 10.1109/ACCESS.2019.2907012

Aladdin, A. M., and Rashid, T. A. (2025). Leo: Lagrange elementary optimization. *Neural Comput. Appl.* 37, 14365–14397. doi: 10.1007/s00521-025-11225-2

Alshayeji, M. H., Ellethy, H., Abed, S., and Gupta, R. (2022). Computer-aided detection of breast cancer on the wisconsin dataset: an artificial neural networks approach. *Biomed. Signal Process. Control* 71:103141. doi: 10.1016/j.bspc.2021.103141

Arora, A., Alsadoon, O. H., Khairi, T. W. A., and Rashid, T. A. (2020). "A novel softmax regression enhancement for handwritten digits recognition using tensor flow library," in *2020 5th International Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA)* (Sydney, NSW: IEEE), 1–9. doi: 10.1109/CITISIA50690.2020.9371821

Aula, S. A., and Rashid, T. A. (2025). FOX-TSA: navigating complex search spaces and superior performance in benchmark and real-world optimization problems. *Ain Shams Eng. J.* 16:103185. doi: 10.1016/j.asej.2024.103185

Azevedo, B. F., Rocha, A. M. A. C., and Pereira, A. I. (2024). Hybrid approaches to optimization and machine learning methods: a systematic literature review. *Mach. Learn.* 113, 4055–4097. doi: 10.1007/s10994-023-06467-x

Beam, K. S., and Zupancic, J. A. F. (2022). Machine learning: remember the fundamentals. *Pediatr. Res.* 93, 291–292. doi: 10.1038/s41390-022-02420-1

Bridle, J. S. (1990). "Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition," in *Neurocomputing*, eds. F. F. Soulié and J. Hérault (Springer Berlin: Heidelberg), 227–236. doi: 10.1007/978-3-642-76153-9_28

Buyukyilmaz, M., and Cibikdiken, A. O. (2016). "Voice gender recognition using deep learning," in *Proceedings of 2016 International Conference on Modeling, Simulation and Optimization Technologies and Applications (MSOTA2016)* (Xiamen: Atlantis Press), msota-16. doi: 10.2991/msota-16.2016.90

Cohen, G., Afshar, S., Tapson, J., and van Schaik, A. (2017). EMNIST: an extension of mnist to handwritten letters. *arXiv Preprint arXiv:1702.05373*. doi: 10.48550/arXiv.1702.05373

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decis. Support Syst.* 47, 547–553. doi: 10.1016/j.dss.2009.05.016

Donati, G., Angeletti, A., Gasperoni, L., Piscaglia, F., Croci Chiocchini, A. L., Scrivo, A., et al. (2020). Detoxification of bilirubin and bile acids with intermittent coupled plasmafiltration and adsorption in liver failure (hercole study). *J. Nephrol.* 34, 77–88. doi: 10.1007/s40620-020-00799-w

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., et al. (2021). Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int. J. Inf. Manage.* 57:101994. doi: 10.1016/j.ijinfomgt.2019.08.002

Elizabeth Rani, G., Sakthimohan, M., Abhigna Reddy, G., Selvalakshmi, D., Keerthi, T., and Raja Sekar, R. (2022). "Mnist handwritten digit recognition using machine learning," in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (Noida: IEEE), 768–772. doi: 10.1109/ICACITE53722.2022.9823806

Fan, S., Li, H., Guo, C., Liu, D., and Zhang, L. (2024). A novel cost-sensitive three-way intuitionistic fuzzy large margin classifier. *Inf. Sci.* 674:120726. doi: 10.1016/j.ins.2024.120726

Gao, Q., and Schweidtmann, A. M. (2024). Deep reinforcement learning for process design: review and perspective. *Curr. Opin. Chem. Eng.* 44:101012. doi: 10.1016/j.coche.2024.101012

Gibert-Ramos, A., Sanfeliu-Redondo, D., Aristu-Zabalza, P., Martínez-Alcocer, A., Gracia-Sancho, J., Guixé-Muntet, S., et al. (2021). The hepatic sinusoid in chronic liver disease: the optimal milieu for cancer. *Cancers* 13:5719. doi: 10.3390/cancers13225719

Goyal, S., Sharma, A., Gupta, P., and Chandi, P. (2021). *Assessment of Iris Flower Classification Using Machine Learning Algorithms*. Springer: Singapore, 641–649. doi: 10.1007/978-981-16-1048-6_50

Grzybowski, A., Pawlikowska-Łagód, K., and Lambert, W. C. (2024). A history of artificial intelligence. *Clin. Dermatol.* 42, 221–229. doi: 10.1016/j.clindermatol.2023.12.016

Guengerich, F. P. (2020). A history of the roles of cytochrome p450 enzymes in the toxicity of drugs. *Toxicol. Res.* 37, 1–23. doi: 10.1007/s43188-020-00056-z

Hama Rashid, D. N., Rashid, T. A., and Mirjalili, S. (2021). ANA: ANT nesting algorithm for optimizing real-world problems. *Mathematics* 9:3111. doi: 10.3390/math9233111

Hamad, R. K., and Rashid, T. A. (2024). Goose algorithm: a powerful optimization tool for real-world engineering challenges and beyond. *Evol. Syst.* 15, 1249–1274. doi: 10.1007/s12530-023-09553-6

Hodges, C. B., Stone, B. M., Johnson, P. K., Carter, J. H., Sawyers, C. K., Roby, P. R., et al. (2022). Researcher degrees of freedom in statistical software contribute to unreliable results: a comparison of nonparametric analyses conducted in SPSS, SAS, STATA, and R. *Behav. Res. Methods* 55, 2813–2837. doi: 10.3758/s13428-022-01932-2

Hoffmann, F., Bertram, T., Mikut, R., Reischl, M., and Nelles, O. (2019). Benchmarking in classification and regression. *WIREs Data Mining and Knowledge Discovery* 9:1318. doi: 10.1002/widm.1318

Ishibashi, H., Nakamura, M., Komori, A., Migita, K., and Shimoda, S. (2009). Liver architecture, cell function, and disease. *Semin. Immunopathol.* 31, 399–409. doi: 10.1007/s00281-009-0155-6

Jabber, S., Hashem, S., and Jafer, S. (2023). Analytical and comparative study for optimization problems. *Iraqi J. Comput. Commun. Control Syst. Eng.* 23, 46–57. doi: 10.33103/uot.ijccce.23.4.5

Jiang, T., Gradus, J. L., and Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behav. Ther.* 51, 675–687. doi: 10.1016/j.beth.2020.05.002

Jumaah, M. A., Ali, Y. H., Rashid, T. A., and Vimal, S. (2024). FOXANN: a method for boosting neural network performance. *J. Soft Comput. Comput. Appl.* 1:2. doi: 10.70403/3008-1084.1001

Jumin, E., Zaini, N., Ahmed, A. N., Abdullah, S., Ismail, M., Sherif, M., et al. (2020). Machine learning versus linear regression modelling approach for accurate ozone concentrations prediction. *Eng. Appl. Comput. Fluid Mech.* 14, 713–725. doi: 10.1080/19942060.2020.1758792

Kennedy, C. C., Brown, E. E., Abutaleb, N. O., and Truskey, G. A. (2021). Development and application of endothelial cells derived from pluripotent stem cells in microphysiological systems models. *Front. Cardiovasc. Med.* 8:625016. doi: 10.3389/fcvm.2021.625016

Khudhair, A. T., Maolood, A. T., and Gbashi, E. K. (2024a). A novel approach to generate dynamic s-box for lightweight cryptography based on the 3d hindmarsh rose model. *J. Soft Comput. Comput. Appl.* 1. doi: 10.70403/3008-1084.1003

Khudhair, A. T., Maolood, A. T., and Gbashi, E. K. (2024b). Symmetric keys for lightweight encryption algorithms using a pre—trained VGG16 model. *Telecom* 5, 892–906. doi: 10.3390/telecom5030044

Khudhair, A. T., Maolood, A. T., and Gbashi, E. K. (2024c). Symmetry analysis in construction two dynamic lightweight s-boxes based on the 2D tinkerbell map and the 2d duffing map. *Symmetry* 16:872. doi: 10.3390/sym16070872

Kolides, A., Nawaz, A., Rathor, A., Beeman, D., Hashmi, M., Fatima, S., et al. (2023). Artificial intelligence foundation and pre-trained models: fundamentals, applications, opportunities, and social impacts. *Simul. Model. Pract. Theory.* 126:102754. doi: 10.1016/j.simpat.2023.102754

Krebs, N. J., Neville, C., and Vacanti, J. P. (2007). *Cellular Transplants for Liver Diseases* (Amsterdam: Elsevier), 215–240. doi: 10.1016/B978-012369415-7/50013-2

Kumar, M., Mehta, U., and Cirrincione, G. (2024). Enhancing neural network classification using fractional-order activation functions. *AI Open* 5, 10–22. doi: 10.1016/j.aiopen.2023.12.003

Kumar, S., Yildiz, B. S., Mehta, P., Panagant, N., Sait, S. M., Mirjalili, S., et al. (2023). Chaotic marine predators algorithm for global optimization of real-world engineering problems. *Knowl. Based Syst.* 261:110192. doi: 10.1016/j.knosys.2022.110192

Lasalvia, M., Capozzi, V., and Perna, G. (2022). A comparison of PCA-LDA and PLS-DA techniques for classification of vibrational spectra. *Appl. Sci.* 12:5345. doi: 10.3390/app12115345

Liu, Y., Cao, B., and Li, H. (2020). Improving ant colony optimization algorithm with epsilon greedy and levy flight. *Complex Intell. Syst.* 7, 1711–1722. doi: 10.1007/s40747-020-00138-3

Maharjan, S., Alsadoon, A., Prasad, P., Al-Dalain, T., and Alsadoon, O. H. (2020). A novel enhanced softmax loss function for brain tumour detection using deep learning. *J. Neurosci. Methods* 330:108520. doi: 10.1016/j.jneumeth.2019.108520

Mohammed, H., and Rashid, T. (2022). Fox: a fox-inspired optimization algorithm. *Appl. Intell.* 53, 1030–1050. doi: 10.1007/s10489-022-03533-0

Molnar, C., Casalicchio, G., and Bischl, B. (2020). *Interpretable Machine Learning—A Brief History, State-of-the-Art and Challenges* (Springer International Publishing: New York), 417–431. doi: 10.1007/978-3-030-65965-3_28

Moradi, E., Jalili-Firoozinezhad, S., and Solati-Hashjin, M. (2020). Microfluidic organ-on-a-chip models of human liver tissue. *Acta Biomater.* 116, 67–83. doi: 10.1016/j.actbio.2020.08.041

Mutar, H., and Jawad, M. (2023). Analytical study for optimization techniques to prolong wsns life. *Iraqi J. Comput. Commun. Control Syst. Eng.* 23, 13–23. doi: 10.33103/uot.ijccce.23.2.2

Naidu, G., Zuva, T., and Sibanda, E. M. (2023). *A Review of Evaluation Metrics in Machine Learning Algorithms* (Springer International Publishing: New York), 15–25. doi: 10.1007/978-3-031-35314-7_2

Nikmaneshi, M. R., Firoozabadi, B., and Munn, L. L. (2020). A mechanobiological mathematical model of liver metabolism. *Biotechnol. Bioeng.* 117, 2861–2874. doi: 10.1002/bit.27451

Oladejo, S. O., Ekwe, S. O., Ajibare, A. T., Akinyemi, L. A., and Mirjalili, S. (2024). "Tuning SVMs' hyperparameters using the whale optimization algorithm," in *Handbook of Whale Optimization Algorithm Variants, Hybrids, Improvements, and Applications Book* (London: Elsevier), 495–521. doi: 10.1016/B978-0-32-395365-8.00042-7

Palanivinayagam, A., El-Bayeh, C. Z., and Damaševičius, R. (2023). Twenty years of machine-learning-based text classification: a systematic review. *Algorithms* 16:236. doi: 10.3390/a16050236

Parimala, N., and Muneeswari, G. (2023). "A review: binary classification and hybrid segmentation of brain stroke using transfer learning-based approach," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (New Delhi: IEEE), 1–6. doi: 10.1109/ICCCNT56998.2023.10307442

Qu, Y., Roitero, K., Barbera, D. L., Spina, D., Mizzaro, S., and Demartini, G. (2022). Combining human and machine confidence in truthfulness assessment. *J. Data Inf. Qual.* 15, 1–17. doi: 10.1145/3546916

Quan, Z., and Pu, L. (2022). An improved accurate classification method for online education resources based on support vector machine (SVM): algorithm and experiment. *Educ. Inf. Technol.* 28, 8097–8111. doi: 10.1007/s10639-022-11514-6

Rajeswari, V., and Sakthi Priya, K. (2025). Ontological modeling with recursive recurrent neural network and crayfish optimization for reliable breast cancer prediction. *Biomed. Signal Process. Control* 99:106810. doi: 10.1016/j.bspc.2024.106810

Sarker, I. H. (2021). Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* 2:160. doi: 10.1007/s42979-021-00592-x

Schlegel, A., Mergental, H., Fondevila, C., Porte, R. J., Friend, P. J., and Dutkowski, P. (2023). Machine perfusion of the liver and bioengineering. *J. Hepatol.* 78, 1181–1198. doi: 10.1016/j.jhep.2023.02.009

Schmidgall, S., Ziaei, R., Achterberg, J., Kirsch, L., Hajiseyedrazi, S. P., and Eshraghian, J. (2024). Brain-inspired learning in artificial neural networks: a review. *APL Mach. Learn.* 2. doi: 10.1063/5.0186054

Seliya, N., Abdollah Zadeh, A., and Khoshgoftaar, T. M. (2021). A literature review on one-class classification and its potential applications in big data. *J. Big Data* 8. doi: 10.1186/s40537-021-00514-x

Sidumo, B., Sonono, E., and Takaidza, I. (2022). An approach to multi-class imbalanced problem in ecology using machine learning. *Ecol. Inform.* 71:101822. doi: 10.1016/j.ecoinf.2022.101822

Sun, H., and Schanze, K. S. (2022). Functionalization of water-soluble conjugated polymers for bioapplications. *ACS Appl. Mater. Interfaces* 14, 20506–20519. doi: 10.1021/acsami.2c02475

Tan, Y., An, K., and Su, J. (2024). Review: mechanism of herbivores synergistically metabolizing toxic plants through liver and intestinal microbiota. *Comp. Biochem. Physiol. C Toxicol. Pharmacol.* 281:109925. doi: 10.1016/j.cbpc.2024.109925

Tufail, S., Riggs, H., Tariq, M., and Sarwat, A. I. (2023). Advancements and challenges in machine learning: a comprehensive review of models, libraries, applications, and algorithms. *Electronics* 12:1789. doi: 10.3390/electronics12081789

Waheed, Z., and Humaidi, A. (2023). Whale optimization algorithm enhances the performance of knee-exoskeleton system controlled by smc. *Iraqi J. Comput. Commun. Control Syst. Eng*. 23, 125–135. doi: 10.33103/uot.ijccce.23.2.10

Watson, D. S. (2023). On the philosophy of unsupervised learning. *Philos. Technol*. 36:28. doi: 10.1007/s13347-023-00635-6

Wojtowytsch, S. (2023). Stochastic gradient descent with noise of machine learning type part I: Discrete time analysis. *J. Nonlinear Sci*. 33:45. doi: 10.1007/s00332-023-09903-3

Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint* arXiv:1708.07747. doi: 10.48550/arxiv.1708.07747

Xue, Y., Jin, G., Shen, T., Tan, L., And Wang, L. (2023). Template-guided frequency attention and adaptive cross-entropy loss for UAV visual tracking. *Chin. J. Aeronaut*. 36, 299–312. doi: 10.1016/j.cja.2023.03.048

Zhang, A., Meng, K., Liu, Y., Pan, Y., Qu, W., Chen, D., et al. (2020). Absorption, distribution, metabolism, and excretion of nanocarriers *in vivo* and their influences. *Adv. Colloid Interface Sci*. 284:102261. doi: 10.1016/j.cis.2020. 102261

Zhao, Z., Alzubaidi, L., Zhang, J., Duan, Y., and Gu, Y. (2024). A comparison review of transfer learning and self-supervised learning: definitions, applications, advantages and limitations. *Expert Syst. Appl*. 242:122807. doi: 10.1016/j.eswa.2023. 122807