IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 26, 2021, accepted May 7, 2021, date of publication May 17, 2021, date of current version May 25, 2021.
*Digital Object Identifier 10.1109/ACCESS.2021.3081366*

# Multiple Filter-Based Rankers to Guide Hybrid Grasshopper Optimization Algorithm and Simulated Annealing for Feature Selection With High Dimensional Multi-Class Imbalanced Datasets

**ABDULRAUF GARBA SHARIFAI[1,2], (Graduate Student Member, IEEE), AND ZURINAHNI BINTI ZAINOL[1]**
[1]School of Computer Sciences, Universiti Sains Malaysia, Gelugor 11800, Malaysia
[2]Department of Computer Sciences, Yusuf Maitama Sule University, Kano 700252, Nigeria
Corresponding author: Zurinahni Binti Zainol (zuri@usm.my)

This work was supported by the Universiti Sains Malaysia under Grant 1001/PKOMP/8014084.

**ABSTRACT** DNA microarray data analysis is infamous due to a massive number of features, imbalanced class distribution, and limited available samples. In this paper, we focus on high-dimensional multi-class imbalanced problems. The high dimensional and multi-class imbalanced problem has posed acute challenges for the conventional classifiers to effectively perform classification tasks on both the minority and majority classes. Numerous efforts have been devoted to addressing either high dimensionality dataset or class imbalance problems. Nonetheless, few methods have been proposed to address the intersection of multi-class imbalanced and high-dimensional problems concurrently due to their intricate interactions. This paper presents novel hybrid algorithms for feature selection with the high dimensional multi-class imbalanced problem using multiple filter-based rankers (MFR) and hybrid Grasshopper optimization algorithm (GOA). The Simulated Annealing (SA) algorithm is incorporated into GOA. SA is used to enhance the best solution found by the GOA algorithm. The aim of using the SA here is to tackle the slow convergence and improve the exploitation by searching the high-quality regions found by the GOA. The experimental results confirm the effectiveness of the proposed methods in improving the classification performance in terms of area under the curve (AUC) compared to other well-known methods, which guarantees the ability of the proposed methods in searching the feature space and identifying very robust and discriminative features that best predict the minority class.

**INDEX TERMS** Grasshopper optimization algorithm, simulated annealing, tournament selection, multiple filter-based rankers, feature selection, hybrid filter-wrapper, high-dimensionality, imbalanced class distribution.

## I. INTRODUCTION

Over the last decades, rapid technological developments have enabled researchers to analyse a massive amount of data from various application domains such as biomedical, information retrieval, and text classification [1]. The characteristics of these datasets are a massive number of features with limited available samples and imbalanced class distribution; these open challenges have degraded the classification performance of most learning algorithms [2]. The imbalanced class distribution has posed acute challenges in numerous applications, including bioinformatics (i.e., diagnosis of rare diseases) [3]. The class imbalance occurs when at least one class is under-represented with a fewer number of samples (i.e., minority class) while other classes contain the most significant part of the remaining samples (i.e., majority class) [4]. Most real-world datasets are affected by the class imbalance problem due to the number of majority class examples (negative class) outnumbered the number of minority class examples (positive class). However, conventional

The associate editor coordinating the review of this manuscript and approving it for publication was Lefei Zhang.

classifiers are appropriate for the balanced training set. In the case of a class imbalance situation, they showed erroneous classification results, i.e., a high recognition for the negative class samples, while the positive class samples are considered noise data or ignored [5].

In the past decade, significant research efforts have been committed to overcoming the imbalanced class problem [6], [7], include resampling methods [8]–[11], ensemble learning [12]–[15], cost-sensitive learning [16]–[18], one class learning [19]–[21], and active learning [22]–[24]. However, most of the existing approaches paid more attention to the development of class imbalanced techniques without considering the effect of other data complexity embedded in the data structure that degenerates the classification performance of the learning algorithm [25]. Recent studies validate this claim that the classification performance of the existing class imbalance techniques can drastically degrade if directly applied to a dataset with more than thousands of features and limited available samples. These findings demonstrate that high dimensionality interferes with the performance of imbalanced class techniques [26]. The native feature selection methods have been proposed by implicitly or explicitly assumed that the number of class samples is equally distributed. However, in real-world applications, most high-dimensional datasets are affected by skewed classes, i.e., one class is under-represented with much fewer examples while the other classes contain the most significant part of the remaining examples [27]. However, for skewed datasets, the native feature selection technique inclines to select features representing the negative classes rather than those features describing the positive class example; the reduct features use for the next classification task will be difficult to achieve optimal solution using biased selected features [28]. Most previous techniques, such as cost-sensitive learning and sampling methods, are often insufficient to improve classifier performance when learning from high-dimensional multi-class imbalanced datasets [27]. The main problem is that high dimensionality and class imbalance learning problem need to be addressed concurrently.

In machine learning, the classification task is broadly classified into a binary classification and multi-class problem, for the binary classification where the samples are divided into the majority and minority class. However, most real-world applications involved the classification of more than two classes, where each class contains a small portion of the samples [29]. The multi-class imbalanced classification is associated with different classification difficulties because the interrelation between classes is no longer apparent [30]. A class could be a minority class compare to other classes, and it could be a majority class compare with other classes or a balanced class compare with the remaining classes [31]. The multi-class imbalanced problems are much more challenging to address than the binary scenario since the decision boundary involves distinguishing between more classes. Unfortunately, directly applying the proposed methods for dealing with the binary class problem to address the multi-class imbalanced problem, the learning models may produce erroneous classification results [32], [33].

Lately, applying feature selection approaches to overcome imbalanced class problems has become a well-known technique among machine learning researchers [34]–[36]. The feature selection technique for the class imbalance problem aims to identify the possible combinations of features that best predict the minority class [37]. Feature selection techniques can be classified into the filter, wrapper, and embedded approaches. The filter methods carry out the feature selection process as a pre-processing step without involving the learning algorithm. The filter method used general characteristics of the training sets to assess the significance of each feature subset (i.e., distances between classes or statistical dependencies) [38], [39]. The wrapper approach generates a subset of features where the induction algorithm is used as a black box to evaluate the significance of each feature subset based on the classification accuracy of the induced classifiers [39]. The embedded feature selection method differs from other feature selection approaches, and the embedded feature selection includes interaction with the classifiers [40]. In the embedded method, searching for the best feature subsets is constructed into the classifier construction. This approach carries out feature selection as part of the classifier construction process that saves the time required for two induction processes as in the wrapper method [41]. Examples of embedded feature selection methods include CART, C4.5, and random forest. Some embedded techniques perform feature ranking according to the significance of each feature using Logistic Regression t predict the probabilities of the classes based on the input features [42].

Numerous studies have used sequential search methods for feature selection problems. However, sequential searching for all possible combinations of features will guarantee optimal subsets to be selected at the cost of high computational expensive or even computationally impractical for high dimensional datasets [43].In contrast to the sequential search techniques, meta-heuristics are suitable approaches for solving complex optimization problems. Metaheuristic methods can search for the best (near-optimal) solutions without increasing the computational complexity [44]. The metaheuristics are appropriate methods to generate a feature subset for the high-dimensional multi-class imbalanced dataset. The metaheuristic is a higher-level heuristic algorithm designed to search and generate a heuristic that can provide better solutions to the optimization process [45]. Metaheuristics have been used to tackle the feature selection problems, and the obtained result is shown to be better than other methods [46], [47]. Numerous metaheuristics based optimization techniques have been proposed, such as Simulated Annealing (SA), Iterated Local Search (ILS) [48], the particle swarm optimization (PSO) [49], the genetic algorithm (GA) [50], Harmony search (HS), Grasshopper optimization algorithm (GOA) [51], Slime Mould Algorithm (SMA) [52], Heap-based optimizer (HBO) [53], Harris hawks optimization algorithm (HHO) [54], Marine

Predators Algorithm (MPA) [55], Equilibrium Optimizer (EO) [56], Manta-Ray Foraging Optimization (MRFO) [57] and Archimedes optimization algorithm (AOA) [58] have been used in the literature to search feature subset space for selecting (sub)optimal feature set.

The metaheuristics are broadly divided into two categories; trajectory-based (e.g., local search) methods that are exploitation-oriented techniques, and population-based methods (e.g., evolutionary and swarm intelligence techniques) that are exploration-oriented techniques. A global exploration of the feature search space and local exploitation, searching for a specific region that has been previously explored. Exploration and exploitation are two conflicting criteria to be considered when developing or employing metaheuristics optimization. A proper balance between exploitation and exploration will improve the performance of the metaheuristic optimization process. It has been proven that using a hybrid approach that combined the advantages of multiple methods can significantly improve the performance of the learning model, and in most cases, hybrid algorithms significantly outperformed the performance of an individual algorithm. This paper uses the GOA algorithm with SA to formulate a novel hybrid model to enhance the classification performance for high-dimensional multi-class imbalanced problems [59].

SA [60] is a single solution-based metaheuristic algorithm proposed by Kirkpatrick *et al.* [60] and can be regarded as a hill-climbing-based technique that is iteratively attempting to enhance a candidate solution in terms of the objective criterion of the problem at hand. The solution is accepted if the chosen move enhances the solution. In other words, the worse move is accepted with a specific probability to help the SA to avoid being stuck in local optima. The Boltzmann probability is used to find the possibility of accepting a worse solution, and the $\mathrm{P} = \mathrm{e} - \theta/\mathrm{T}$ where $\theta$ is the difference of evaluation of the objective criterion between the trail solution $Sol_{trial}$) (and the best solution ($Sol_{best}$). The $T$ is the current temperature that occasionally approaches zero during the search process according to a particular cooling schedule.

GOA [51]@comm proposed by Mirjalili *et al.*, is a recent optimization algorithm that mimics grasshoppers' behaviour. The GOA is a population-based method that can successfully explore vast search spaces to locate the best (near-optimal) solutions. The GOA has fewer adjustable parameters to be set and has a fast convergence speed. Therefore, GOA has been successfully adapted for numerous optimization problems, include cloud logistics [61], clustering [62], image processing [63], global optimization problems [64], and classification tasks [65]. Due to the efficacy of GOA in various applications, that motivates numerous researchers towards enhancing the optimization ability of the basic GOA by hybridizing it with other meta-heuristic algorithms or local search algorithms [63], [66], [67] and modification of some components of GOA [66], [68], [69], have been proposed to address different complex optimization problems. However, the hybrid of GOA with the SA algorithm is not yet proposed and GOA-SA is not yet investigated for high-dimensional multi-class imbalanced problems. The robustness of GOA and SA algorithms could be integrated to yield a hybrid method that takes advantage of both approaches to achieve outstanding results better than using individual algorithm independently. This hybridization is to improve the exploitation ability of the GOA. The tournament selection (TS) mechanism is used rather than a random selection (RS) to maintain the population diversity and to improve the exploration ability of the proposed method.

Tharwat *et al.* [70] developed a modified multi-objective GOA (MOGOA) with an external archive for constrained and unconstrained problems. Mirjalili *et al.* [71] developed the multi-objective GOA and revealed that the proposed algorithm could tackle several benchmark problems effectively and with better performance in terms of accuracy of Pareto optimal solutions and the related distribution. Luo *et al.* [65] incorporated three strategies to balance exploitation and exploration of GOA. (i) The Gaussian mutation is used to boost the population diversity, (ii) Levy-flight is adopted to boost the randomness of the search agent movement, (iii) opposition-based learning is applied to enhance the search agent in the solution space of the GOA algorithm. Ewees *et al.* [72] improve the exploration ability of GOA using opposition-based learning (OBL). The proposed method has been evaluated on twenty-three benchmark functions. Liang *et al.* [66] GOA is modified for multilevel Tsallis cross-entropy. The levy flight is utilized to improve the GOA to achieve a proper balance between exploration and exploitation. Jia *et al.* [63] proposed a hybrid algorithm of GOA and Differential Evolution (DE) to mitigate the slow convergence speed and balance between exploration and exploitation of the GOA. Amaireh *et al.* [73] proposed a hybrid algorithm of GOA and Antlion Optimization (ALO). ALO has strong exploitation capability, and GOA has good exploration capability. The proposed algorithm aims to cope with the drawbacks of both algorithms.

This paper proposes a hybrid filter-wrapper algorithm for feature selection with high dimensional multi-class imbalanced datasets. This paper comprises two phases filter-based and wrapper-based approaches. In the filter-based approach, where the top-ranked features from each filter method were selected to form a new feature list, the feature occurrence threshold value computed from the top-ranking features, the features those satisfied the threshold value are selected and used as input to the wrapper method. While in the wrapper approach, a hybrid approach of the global search algorithm (GOA) with local search algorithm (SA) is proposed.

The main contributions of this paper focus on proposing a hybrid filter-wrapper method to improve predictive performance for the classification of high-dimensional multi-class imbalanced datasets and improve the exploitation capability of the GOA algorithm. To enhance the exploitation, the SA is employed in a pipeline mode after the GOA terminates to tackle the slow convergence and improve the exploitation after the GOA algorithm finds the best solution.


**FIGURE 1.** (a) Real grasshopper (b) Life cycle of grasshopper.

To preserve the diversity of the algorithm, tournament selection is employed to select search agents from the population to give a chance to all individuals to be selected. The proposed hybrid method, namely a high-level transmit hybrid (HTH). The proposed approaches are evaluated on nine benchmark datasets with varying high dimensionality, imbalance ratio, and number samples. Experimental results show that the proposed methods can achieve better results than CBR-PSO, SMOTE-BOOST, RUS-BOOST, GOA, PSO, GWO, and GA on most datasets. In literature, numerous metaheuristic hybridizations have been proposed, but this is the first time a hybrid method of SA and GOA algorithm is used for feature selection with the high-dimensional multi-class imbalanced problems.

The rest of this paper is structured as follows: Section 2 provides the related works. The concepts of GOA and SA algorithms are provided in Section 3. Section 4 represents results, discussion, and analysis. In section 5, conclusions and future work are given.

## II. RELATED WORKS

Considering the specific focus of this paper, we shall cover only relevant articles that use feature selection approaches to tackle high-dimensional multi-class imbalanced datasets. Lately, feature reduction techniques such as feature selection have been applied to tackle the imbalanced class problem at the feature level [6] because most of the high dimensional data sets have imbalanced class problems [6], [74], such as bioinformatics [75], text categorization [76], microarray data set [77].

In practice, many real-world datasets comprise more than two class labels with skewed class distributions, such as biomedical datasets [78]–[80]. Various binary classification methods have been proposed, but few methods have been extended to cope with the high-dimensional multi-class imbalanced problem. Classification for high-dimensional multi-class imbalanced datasets has drawn considerable attention from many researchers. Numerous studies presented many decomposition approaches to cope with the multi-class imbalanced problem, including the one versus all (OVA) method [81], [82], one versus one (OVO) technique [83], [84], error-correcting output codes [85], and decision directed acyclic graph (DDAG) [86]. These decomposition approaches have been used to cope with the classification of imbalanced datasets. Statnikov *et al.* [87] investigate OVA and OVO decomposing approaches on imbalanced data after conducting extensive experiments using these strategies. The authors reported that OVA achieved better results compared with other decomposition approaches. Several studies investigated feature selection for high-dimensional multi-class imbalanced datasets. The experimental results demonstrated that dealing with multi-class is more demanding than binary class imbalanced data. García *et al.* [33] suggested that the learning algorithm performance drastically deteriorates when the number of classes increase; this demonstrates that performing classification task on multi-class imbalanced dataset degenerate the classification performance of the most classifiers. Yang *et al.* [88] proposed an iterative ensemble feature selection (IEFS) on microarray data. The filter-based techniques coupled with sampling methods are iteratively used to improve the binary classification using the One-Versus-All (OVA) approach. The obtained results demonstrate that IEFS achieves better results than other methods in terms of classification accuracy and AUC. Du *et al.* [89] employed a genetic algorithm for multi-class imbalanced class distribution. A fitness function was formulated using the G-mean metric instead of classification accuracy to discriminate features proportion of both classes. The obtained results demonstrate that the proposed method has achieved promising results than other conventional feature selection techniques.

Yu *et al.* [90] proposed an ensemble approach to handle multi-class imbalanced datasets using a one versus all (VOA) approach to adapt multi-class into many binary classes combined for feature subspace that create multiple different training features subsets. Firstly, two methods have been proposed: decision threshold adjustment and random undersampling into each training set to handle the imbalanced class problem using the SVM classifier as a base classifier. The proposed method has been assessed on eight benchmark datasets. The obtained results demonstrate the effectiveness of the proposed approach to handle high-dimensional imbalanced datasets than other techniques.

Arias-Michel *et al.* [91] proposed dynamic selection techniques to handle multi-class imbalanced problems. Five pre-processing techniques and fourteen dynamic selection strategies were employed. The effectiveness of the proposed method was evaluated on twenty-six multi-class imbalanced data sets. The obtained results indicate that the dynamic ensemble had achieved comparable results based on AUC and G-mean performance metrics similar to the static ensemble technique. Hosseini and Moattar [92] proposed an evolutionary feature subset selection algorithm using three-phase approaches. At the first stage, a features weighted approach was applied to identify high-ranking features. In the second stage, the feature subsets were formulated and assessed using a multivariate interaction information method. Finally, the optimal feature subsets were identified and extracted using a dominant/dominated relationship.

Sun *et al.* [93] introduced an ensemble method that hybridized the multi-objective Ant colony optimization algorithm to under-sample the majority class samples and feature selection method to select the best features. The bootstrap approach is used in the original feature set to create various sample subsets. The V-statistic is used to assess the dataset distribution and a fitness criterion of GA to under-sampling the majority class. Two fitness functions were formulated using G-mean and F1 measures. The obtained results show that the proposed method outperforms many approaches in terms of G-mean and F1 metrics. Aydogan *et al.* [94] presented a hybrid approach using PSO and rough set theory (CBR-PSO) for feature selection. The CBR-PSO is used to identify the relevant features, and the obtained reduct features can improve the classification performance of the model. The CBR-PSO handles the multi-class problem using the OVO strategy to decompose the multi-class datasets into a binary classification problem. The obtained results reveal that the proposed method has achieved promising results than other approaches for high-dimensional problems. Notwithstanding, the algorithms mentioned above have achieved promising results. Still, there is no consistent winning approach for all datasets. The performance of most class imbalanced techniques is affected by the high-dimensionality problem.

Metaheuristics have been used for feature selection with high-dimensional multi-class imbalanced datasets, but the current methods showed some shortcomings. The high-dimensional multi-class imbalanced datasets have a large search space. Most bio-inspired optimization techniques did not achieve optimal performance in large-scale applications such as high dimensional multi-class imbalanced datasets to the premature convergence and stagnation problem [6], [95]. Improving meta-heuristic optimization via hybridization can improve the classification performance for the high-dimensional multi-class imbalanced datasets. Therefore, there is room for further improvement of the existing approaches for the high-dimensional multi-class imbalanced problem. That motivated this research to proposed hybrid multiple filter-based rankers and memetic algorithms for high-dimensional multi-class imbalanced datasets in the subsequent section.

## III. MATERIALS AND METHOD

### A. GRASSHOPPER OPTIMIZATION ALGORITHM

GOA is a recent bioinspired optimization technique [51] proposed by S Mirjalili *et al.* it imitates grasshoppers swarming behaviour. The GOA algorithm is broadly classified into two types of behaviours, exploration and exploitation. During the exploration, the search agents tend to move swiftly to explore the whole feature space, while during the exploitation, the search agents tend to move sluggishly to exploit around a neighbourhood of the current solution in the search space. These two attributes enable the GOA to handle complex optimization problems and outperform numerous well-regarded techniques on various complex optimization problems [71].

#### 1) SOLUTION CONSTRUCTION PHASE

The GOA algorithm mathematically is represented to mimic the swarming behaviour of the grasshoppers. The location of *ith* grasshopper towards the target solution is expressed as $X_i$ as shown in Eq. (1).

$$Xi = Si + Gi + Ai \tag{1}$$

where $\mathrm{S_i}S_i$ is the social interaction, $G_i$ is the gravity force on *ith* grasshopper, and $A_i$ denotes the wind advection. Th $S_i$ works as the significant component during the grasshopper movement towards the target, and it can be computed as shown in Eq. (2)

$$S_i = \sum_{\substack{j=1 \\ j \neq i}}^{N} s\left(d_{ij}\right)\hat{d}_{ij} \tag{2}$$

where $d_{ij}$ is the Euclidian distance between the *ith* and the *jth* grasshopper, the distance between grasshoppers can be computed as $d_{ij} = \left|x_j - x_i\right|$ and $\hat{d}_{ij} = \frac{x_j - x_i}{d_{ij}}$ is a unit vector from the *ith* grasshopper to *jth* grasshopper.

The s function, which represents the strength of the social force, and can be computed as shown in Eq. (3)

$$s\left(r\right) = fe^{\frac{-d}{l}} - e^{-r} \tag{3}$$

where $f$ defines the attraction power, and $l$ is the attractive length scale. The gravity force $G$ can be calculated, as shown in Eq. (4)

$$G_i = -g\hat{e}_g \tag{4}$$

where $g$ represents the gravitational constant and $\hat{e}_g$ represents the unity vector towards the centre of the earth.

The wind advection $A_i$ can be computed, as shown in Eq. (5)

$$\mathrm{A}_i = u\hat{e_w} \tag{5}$$

where $u$ indicates a constant drift and the $\hat{e_w}$ is a unit vector in the direction of the wind.

GOA mathematical model, shown in Eq. (1), can be formulated, as shown in Eq. (6).

$$X_i^d = c\left(\sum_{\substack{j=1 \\ j \neq i}}^{N} c\frac{ub_d - lb_d}{2} s\left(\left|X_j^d - X_i^d\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \hat{T}_d \tag{6}$$

where $\mathrm{ub_d}$ and $\mathrm{lb_d}$ indicate the upper bound and the lower bound in the Dth dimension, $Td$ shows the $Dth$ dimension value in the target grasshopper. The parameter c is a decreasing coefficient that is used to decrease the comfort, attraction, and the gravity strength is considered to be zero, assuming that the wind is moving towards a target $\mathrm{T_d}$ in Eq. (6). The parameter c is applied two times in Eq. (6) to control the convergence rate of grasshoppers and balance the exploration and exploitation. The exterior c controls the movements of grasshoppers in the direction of the target. The interior c shrinks the impact of the comfort, attraction, and repulsion strength between grasshoppers per the number of iterations to reduce the comfort, attraction, and repulsion regions.

**Algorithm 1** The Main Steps of the GOA Algorithm

Initialize a set of random solutions xi (i = 1, 2, 3, . . . , n) as an initial population
Initialize the GOA parameters cMax, cMin, and Max number of iterations
Evaluate the fitness of all individuals
T = the best solution
  **while** (k < maximum number of iterations) **do**
    Update c using equation (7)
    **for each** solution in the population, **do**
      Standardize the distance between grasshopper into [1, 4]
      Update the position vectors using equation (8)
      Update the step vectors according to equation (10)
    **end for**
    there is a better solution, update
    k = k +
  **end while**
**Output the T**

### 2) COEFFICIENT PARAMETE

Parameter c decreases the comfort zone proportional to the number of iterations.

$$c = cmax - l\frac{cMax - cMin}{L} \tag{7}$$

cMax denotes the c maximum, and cMin denotes the $c$ minimum value, l represents the current iteration, and $L$ is the maximum bound of iterations.

### 3) TERMINATION PHAS

The overall operations are repeated until the cessation criteria are reached. The stopping criteria are a computational time constraint, the maximum number of iterations(i.e., MaxIter) is met, the number of halted generations, or the best fitness value is obtained.

Algorithm 1 presents the pseudo-code of the conventional GOA algorithm. It can be observed that the GOA algorithm randomly generates its initial population and assesses each search agent using an objective criterion upon the optimization process begin. After obtaining the best target solution, the GOA randomly performs the following steps until the cessation condition is met. Firstly, c minimum (cMin) and c maximum (cMax) coefficient parameter values are updated. Secondly, a random value is generated, according to the generated random value, the GOA algorithm updates the position of the search agents, according to Eq. (8). Thirdly, the current search agents are stopped from going outside the boundaries. Finally, the algorithm outputs the best target solution found so far.

The GOA is a bioinspired optimization method, as stated earlier. What guarantees the GOA algorithm convergence is the use of the best target solution achieved so far to update the position of the rest solutions. Nevertheless, this approach might cause search agents to be stuck in local optima.

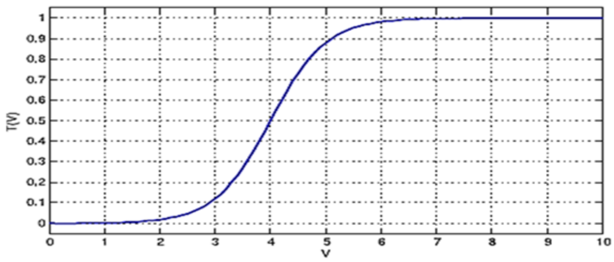


**FIGURE 2.** Sigmoidal Transfer function.

A random variable is used to switch between the two equations to update the position of search agents. The parameter c is used to balance the between exploration and exploitation. This parameter seamlessly decreases the magnitude of changes in the solutions and facilitates the convergence/exploitation corresponding to the number of iterations.

### 4) BINARY GOA ALGORITHM

Selecting the best feature subset has been reported as an NP-hard problem [96]. Identifying the best possible combination of features is a complex optimization problem, especially in high-dimensional feature search space. Therefore, according to the feature selection problem NP-hard nature, the search space can be applied by binary values. Consequently, some of the GOA algorithm equations need to be adjusted. GOA algorithm, each solution updates its position according to its current position, the position of the best grasshopper found so far (target), and all other grasshoppers' position, according to Eq. (8). The best technique to adapt the optimization method from Eq. (8) in the continuous form to the binary version without adjusting its components is by utilizing transfer functions [97].

The transfer function is applied to Eq. (8) and re-defined $\Delta$ in Eq. (9) as the probability of changing the position of grasshoppers.

$$\Delta X = c\left(\sum_{\substack{j=1 \\ j\neq i}}^{N} c\frac{\mathrm{ub_d}-\mathrm{lb_d}}{2}\mathrm{s}\left(\left|\mathrm{X_j^d}-\mathrm{X_i^d}\right|\right)\frac{\mathrm{x_j}-\mathrm{x_i}}{\mathrm{d_{ij}}}\right) \tag{8}$$

A sigmoidal function is a popular transfer function proposed in [98], as shown in Eq. (9).

$$\Delta X = \frac{1}{1 + \mathrm{e}^{-\Delta \mathrm{X_t}}} \tag{9}$$

The position of the current grasshopper is updated, as represented in Eq. (10), via the probability value $T(\Delta \mathrm{X_t})$ obtained from Eq. (9).

$$X_{t+1}^{k}(t+1) = \begin{cases} 1 & \mathit{if}\ \mathit{rand} < (\Delta X_{t+1}) \\ 0 & \mathit{if}\ \mathit{rand} \geq (\Delta X_{t+1}) \end{cases} \tag{10}$$

The sigmoidal function enables grasshoppers to move in line with a binary search space, as shown in Fig. 2.

## B. SIMULATED ANNEALING

Simulated annealing (SA) [60] proposed by Kirkpatrick *et al.* SA is a single-solution metaheuristic algorithm based on the hill-climbing method that emulates the annealing process to solve a combinatorial optimization problem. SA algorithm inspired by the process of solid-state annealing, in which a solid is melted and then cooled down gradually in order to gain ideal crystal structures that can-shaped as a state of minimum energy. The cooling process is required to be done slowly [99]. The idea of underlying SA is to resolve the combinatorial optimization problem analogous to the stable annealing process. The candidate solution of the given problem is analogous to the physical system states, and the global optimum solutions are analogous to the meta-stable states of the physical system [99]. SA typically begins by indiscriminately generate a random solution and initializes the temperature T. Then, at each iteration, a solution $s'$ is indiscriminately selected in the neighbourhood $N(s)$ of the current solution s. The solution $s$ is accepted as a new current solution depending on the value of $T$ and the values of the objective function of $s'$ and s,$s$, denoted by $f(s')$ and $f(s)$, respectively. If $f(s') \geq f(s)$, then, the solution $s'$ is accepted and replaces by solution s. On the other hand, if $f(s') < f(s)$, then, a solution s will be accepted, with a Boltzmann probability.

$$P = \begin{cases} 1, & \text{if } f(s') < f(s), \\ exp\left(\frac{f(s) - f(s')}{T}\right), & \text{if } f(s') \geq f(s) \end{cases} \tag{11}$$

where the $f(s') - f(s)$ represents the difference between the best fitness solution and the generated neighbour solution. Additionally, the temperature $T$, which occasionally decreases during the search process based on the cooling plan. In this research, the initial temperature is selected to be $2 *|N|$, where |N| represents the number of features for each data set, and the cooling scheduled is preferred $\gamma = 0.93$

$$T_{i+1} = \gamma * T_i \tag{12}$$

where $i = 0, 1, \ldots N$ [100].

### C. TOURNAMENT SELECTIO

The tournament selection (TS) is a simple and easy to implement selection mechanism applied to select the finest individual from the populations in evolutionary algorithms. It was introduced by [101] Grefenstette *et al.* TS selection is the most widely used selection mechanism in evolutionary algorithms [102]. InTS, n individual solutions are chosen indiscriminately from the population P(n; these obtained solutions are determined by comparing against other solutions, and a tournament t is held to decide a victor. The tournament requires to generate a random number in the interval of [0 1], it compared with a selection probability that used to adapt the selection pressure parameter (often set to 0.5) if the generated number is higher, it indicates that the individual with the most robust fitness value will be selected or else the low candidate solution is selected. The TS mechanism allows the most robust individual to be favoured. The selection pressure parameter determines the rate of convergence of the evolutionary algorithms [102]. Evolutionary algorithms can identify the best or (near-optimal) solutions over various selection pressures.

**Algorithm 2** Pseudo-Code of the SA Algorithm

```
T0 = 2* |N| is the number of attributes for each data set
BestSol ← Si′
δ (BestSol) ← δ (Si)// δ demonstrates the quality of the solution
While T > Ti
Generate at random a new solution in the neighbour of Si′
Calculate δ (TrialSol)
  If δ (TrialSol) > δ (BestSol)
    Si′ ← TrialSol; BestSol ← TrialSol;
    δ (Si′) ← δ (TrialSol); δ (BestSol)← δ (TrialSol)
  else if (δ (BestSol) = δ (TrialSol))
    calculate |TrialSol| < |BestSol|
    Si′ ← TrialSol; BestSol ← TrialSol
    δ (Si′) ← δ (TrialSol); δ (BestSol)← δ (TrialSol)
  end if
  else // accepting the worse solution
    Calculate θ = δ (BestSol) - δ (TrialSol)
    Generate a random number, P = |0, 1| ;
    if (P ≤e^(-θ/T))
    Si′ ← TrialSol; δ (Si′) ← δ (TrialSol);
  end if
  end if
  T = 0.93 ∗ T; // update temperature
end while
Output BestSol
```

### D. THE PROPOSED FEATURE SELECTION APPROACH FOR HIGH DIMENSIONAL MULTI-CLASS IMBALANCED DATASETS

In this paper, the proposed algorithm for high dimensional multi-class imbalanced datasets consists of two stages, a filter-based approach and the wrapper-based approach. The filter-based and wrapper-based methods are comprehensively discussed in the subsequent subsections. The filter-based approach is multiple filter-based rankers (MFR), and the wrapper approach is the proposed hybrid GOA with the SA algorithm.

#### 1) FILTER-BASED APPROACH: MULTIPLE FILTER-BASED RANKERS (MFR)

Filter-based methods are computationally efficient compared to wrapper methods;nevertheless, filter-based methods suffer severely from the feature interaction problem. The filter-based method can achieve the best result on a specific data set while its performance drastically deteriorates on another data set. Therefore, selecting an optimal filter method for a particular data set will be a demanding task due to the inadequate a priori knowledge of the dataset [103].

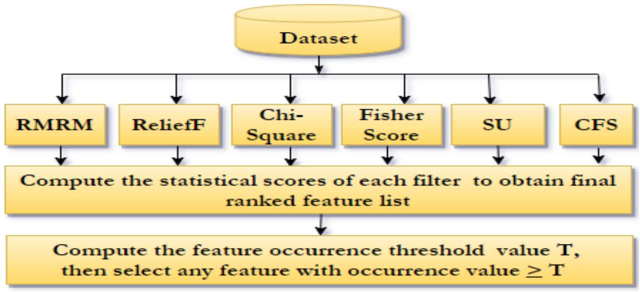


**FIGURE 3. Flowchart of the proposed multiple filter-based rankers (MFR) approach.**

Therefore, using a single filter-based method for feature selection requires conducting extensive trial and error tests to select the best filter-based method for a specific dataset. This problem leads to high computational complexity because a feature selection problem is considered a computationally expensive problem. The limitations above motivated many researchers to combine the outputs of multiple filter-based methods to mitigate inconsistency and the local optima problem induced by a single-based filter method and enhance filter-based technique robustness and stability. Fig. 3 demonstrates the schematic illustration of multiple filter-based rankers for high-dimensional multi-class imbalanced datasets. It can be observed in Fig. 3, six most commonly used filters were used to generate feature ranking, and the top $N$ ranking features from each filter-based were selected and then aggregated to formulate a new ranking features list. A feature occurrence threshold value T is computed to determine the number of feature occurrences among the six filter-based rankers and set to 3 (T = 3). The features that satisfied the threshold criteria are identified and selected as the final feature ranked list.

In this paper, we utilized six filter-based methods, include ReliefF, Symmetric uncertainty (SU), Max-Relevance Min-Redundancy (MRMR), Fisher Score, Chi-square, and Correlation-based feature selection (CFS). These six filter-based rankers were employed to identify and select relevant and non-redundant features that best predict minority class.

### 2) THE PROPOSED WRAPPER APPROACH

The problem of feature selection with the high-dimensional multi-class imbalanced problem is expressed as a multi-objective optimization problem to select the best features that predict the minority class. Mathematically, the whole training set of matrix S can be expressed as $S \in \mathbb{R}^{Y \times Z}$ with their class label C, where Y indicates the number of samples, andZ indicates the number of attributes. In general, the attribute vector $a_i = \{a_1, a_2, a_3, \ldots, a_Z\}$ represents the corresponding Z attributes an $S_{ij}$ d means $jth$ sample and $ith$ attribute, as demonstrated in Eq. (13). For example, the attribute vector of the first sample is represented as $s_{i,1} = \{s_{i,1}, s_{i,2}, s_{i,3}, \ldots s_{i,z}\}$.

The problem of feature selection is considered a difficult combinatorial optimization problem. The whole feature set (i.e., candidate solution s) is represented using a binary string of length S, $s_{i,1} = \{s_{i,1}, s_{i,2}, s_{i,3}, \ldots s_{i,z}\}$. Our goal is to identify and extract discriminative features of dimension z, where $z \subseteq Z$ such that a candidate subset is set to 1 if a feature is selected or 0 if a feature is excluded. Fig. 4 shows the representation of grasshopper i with dimension Z, where Z represents the number of features.

$$S_{ij} = \begin{bmatrix} S_{1,1} & s_{1,2} & s_{1,3} & \cdots & s_{1,z} \\ s_{2,1} & s_{2,2} & s_{2,3} & \cdots & s_{2,z} \\ S_{3,1} & s_{3,2} & s_{3,3} & \cdots & s_{3,z} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{y,1} & s_{y,2} & s_{y,3} & \cdots & s_{y,z} \end{bmatrix} \tag{13}$$

The wrapper-based method is proposed with Z attributes. Our goal is to identify z attributes with the highest predictive accuracy in class imbalanced settings and identify the best features from both minority and majority classes. Because of that, the fitness criterion needs to be formulated to realize these objectives [104]. Each solution is assessed based on the fitness function formulated using the kNN classifier to achieve the highest AUC and identify the reduct features from both the minority and majority classes that best predict the minority class.

The fitness function is an essential part of the feature selection process. Most conventional feature selection methods have been proposed without considering the effects of imbalanced class distribution [28]. The fitness function of the conventional feature selection techniques is formulated using a combination of classification error rate or overall accuracy and length of features, as shown inEq. (14).

$$AUC_{ff} = \alpha\gamma R(D). + \beta \cdot \left(\frac{|\text{N-R}|}{|N|}\right) \tag{14}$$

where $\gamma R(D)$ presents the overall classification accuracy rate obtained using kNN classifier, $|R|$ denotes the number of features selected and |N| represents the entire feature set, $\alpha$ and $\beta$ are two parameters corresponding to the importance of classification quality and length of the features, $\alpha\epsilon\ |0, 1|$ and $\beta = (1-\alpha)$ adopted from [105]. The overall accuracy metric is broadly used as an evaluation metric for classification tasks [106]. Nonetheless, for imbalanced datasets, overall accuracy rate and classification error rate metrics are not appropriate evaluation measures. The classification results obtained using overall accuracy or error rate are biased toward the majority class and ignore the minority class. In a highly imbalanced data set, where the negative samples outnumbered the positive samples, the classifier can achieve the highest accuracy rate of up to 99Still, the positive class recognition rate is insignificant or ignored, as shown in Fig. 5. Therefore, the overall accuracy metric is not suitable for imbalanced datasets.

The AUC is an evaluation metric that measures the balance between predictive accuracy on both positive and negative classes. [107]. AUC measures the whole two-dimensional regions beneath the ROC curve.AUC measure of performance of the entire classification thresholds [35], [108], [109]. The

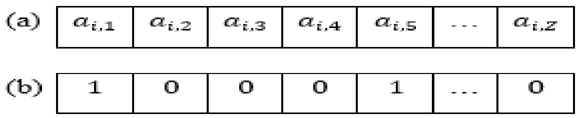


**FIGURE 4.** **(a) Chromosome coding of the attribute vector of grasshopper *i* (b) the binary representation of the feature vector of grasshopper.**

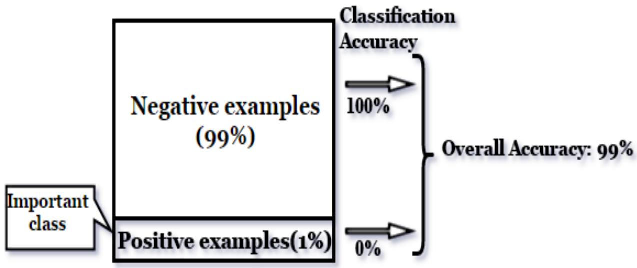


**FIGURE 5.** **Schematic illustration of class imbalance problems.**

best way to understands the AUC is via a probability measure that ranks a random minority sample significantly higher than a random majority sample, the model that achieves the highest AUC shows good recognition of the minority class sample. The AUC assesses the predictive algorithm discriminating capability between the true positive rate *TP*R and false positive rates *FPR* (without considering the model misclassification costs).

According to the confusion matrix, the AUC measure can be defined as follows:

$$AUC = \frac{1 + TP - FP}{2} \tag{15}$$

The fitness function formulated for a binary classification problem using the AUC metric can be formulated, as shown in Eq. (16).

$$AUC_{ff} = \alpha \cdot AUC + \beta \cdot \left(\frac{|\text{N-R}|}{|N|}\right) \tag{16}$$

The binary classification problem can be extended to deal with multi-class imbalance problems. MAUC is defined for the multi-class problem as shown in Eq. (17),

$$\text{MAUC} = \frac{\sum_{\text{i}=1}^{\text{m}} \text{AUC}_{\text{i}}}{\text{M}} \tag{17}$$

where *i* represents the index of the class under consideration, the fitness function for the multi-class classification problem is formulated, as shown in Eq. (18).

$$MAUC_{ff} = \alpha \cdot MAUC + \beta \cdot \left(\frac{|\text{N-R}|}{|N|}\right) \tag{18}$$

### 3) HYBRID GOA-SA APPROACHES

The GOA is a population-based that are exploration-oriented approach, i.e., exploring the whole unexplored search space regions. In contrast, the local search techniques are exploitation-oriented approaches, i.e., intensify searching for specific regions that have been explored [110]. The native GOA has achieved promising results on complex

**TABLE 1.** **Confusion matrix.**

| | Predicted positive class 1 | Predictive negative class 2 |
|---|---|---|
| Actual positive class | TP (True Positive) | FN (False Negative) |
| Actual negative class | FP (False Positive) | TN (True Negative) |

optimization problems than other well-known optimization algorithms. However, the original GOA uses a blind operator to perform local exploitation irrespective of the current solution fitness value and explored solutions. The GOA exploitation (as in Eq. 8) relies on computing the distance between the search agent and the best-explored solution so far. The hybridization of population-based with local-based search techniques to explore the neighbourhood around the best-explored solution might enhance the classification performance of the classifier [48]. Because of this reason, the GOA algorithm is integrated with the local search method SA algorithm to produce a hybrid GOA-SA model. The GOA exploration (as in Eq. 8) relies on switching each solution position based on a randomly selected solution. It has been proven that applying another selection mechanism, such as the TS mechanism, might enhance the exploration capability of the feature search space because the TS mechanism allows the worst individuals to be randomly repositioned around the best solution found so far, which could enhance the GOA population diversity [100].

The GOA is a global search technique, and the SA is a local search technique. Hybridization of two algorithms is proposed, namely, the High-Level Transmit Hybrid (HTH). In the HTH, the SA uses after the GOA algorithm obtained the best solution. After that, the solution serves as the initial input to the SA algorithm to improve the GOA final solutions. The best solutions obtained by the GOA algorithm are then passed to the SA to accelerate the search, tackle the slow convergence, and improve the exploitation capability of the current best solution until a local optimum is reached. The GOASA employs the RS mechanism to explore the feature space. Therefore, the TS mechanism is used to enhance the proposed approaches exploration capability and supplement the SA local search algorithm to strike a proper balance between exploitation and exploration of the search space. The proposed algorithm is called GOASAT, and the flowchart of the proposed MFR-GOASAT algorithm is illustrated in Fig 6.

## IV. EXPERIMENTAL RESULTS AND ANALYSIS

This section experimentally investigates the efficacy of proposed MFR-GOA, MFR-GOASA, and MFR-GOASAT algorithms for high dimensional multi-class imbalanced data sets. Table 2 shows the characteristics of nine datasets used to assess the effectiveness of the proposed methods. The proposed algorithms are used to search for the best feature subsets with the highest MAUC using a kNN classifier (K = 5) [111]. The kNN was selected because of its simplicity and broadly employed in similar existing researches. Since most of the benchmark datasets have a limited sample size and a thousand features; therefore, each dataset is

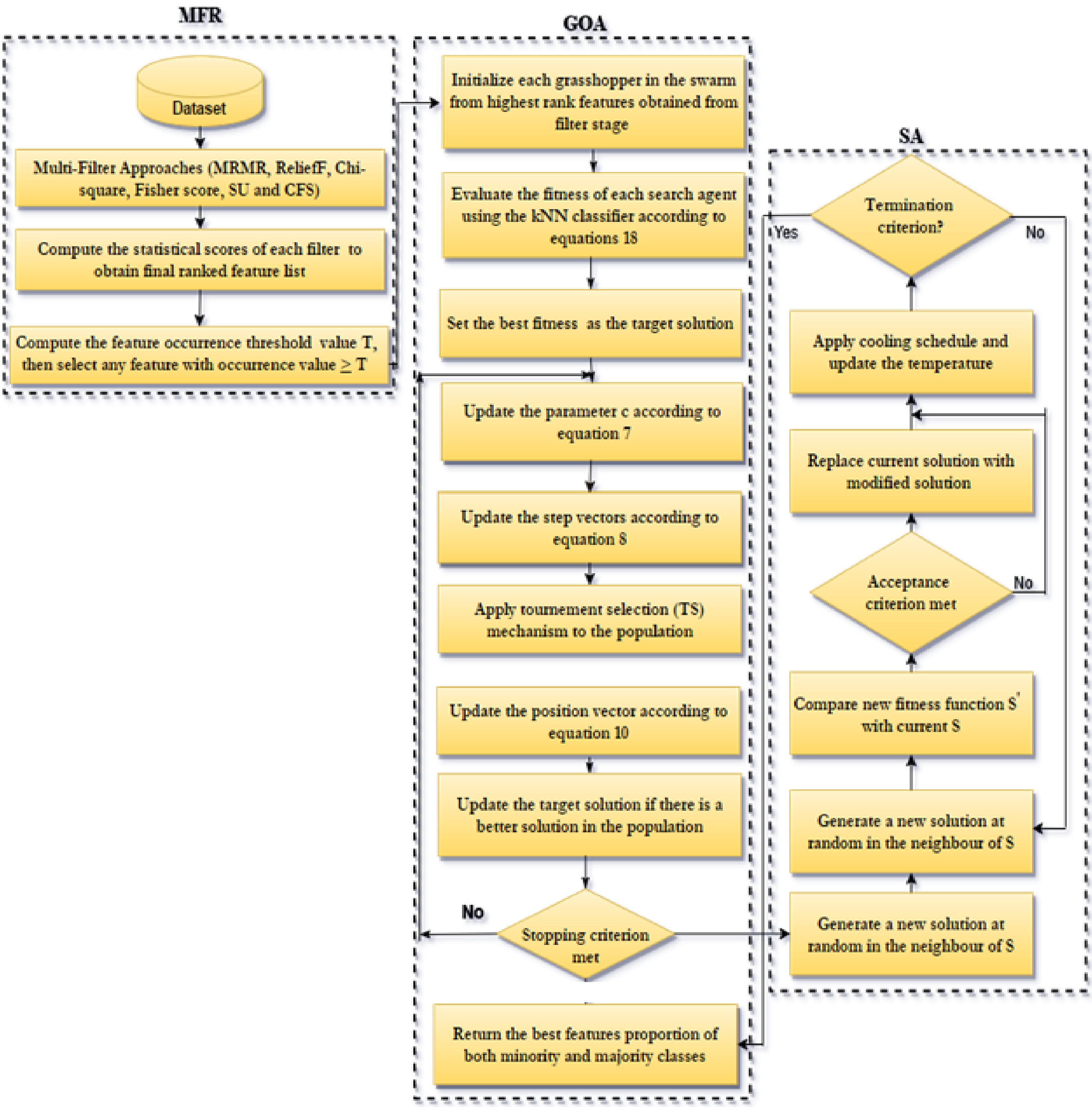


**FIGURE 6.** Schematic illustration of the proposed hybrid MFR-GOASAT approach.

divided into five cross-validations to prevent feature selection bias [65], [66]. The experiments of the proposed methods were performed on MATLAB 2020a software. The experiments were carried out on an Intel Core i5-4300U CPU @ 1.90GHz CPU and 8 GB RAM in Microsoft Windows 10 Pro platform. The maximum number of iterations (L) is set to 100, and the population size (N) is set to 50. The $\alpha$ and $\beta$ values in Eq. (15) are set to 0.9 and 0.1.

### *A. DATASET*

In order to assess the effectiveness of the proposed approach, we conducted numerous experiments on nine benchmark datasets that were downloaded from the machine learning repository [70], [71], [85]. Table 2 presents the general characteristic of datasets in terms of the number of features, the number of samples, the number of classes, and the imbalance ratio (IR) in each dataset.

**TABLE 2.** Data sets characteristics.

| Data sets | #Features | #Samples | #Classes | IR |
|---|---|---|---|---|
| CLL-SUB-111 | 11,340 | 111 | 3 | 4.63 |
| Brain_Tumor1 | 5920 | 90 | 5 | 15.00 |
| TOX-171 | 5748 | 171 | 4 | 1.15 |
| GLIOMA | 4433 | 50 | 4 | 2.29 |
| 11_Tumour | 12,533 | 174 | 11 | 4.5 |
| Lung cancer | 12600 | 203 | 5 | 23.16 |
| SRBCT | 2308 | 83 | 4 | 2.63 |
| 9_Tumour | 5726 | 60 | 9 | 4.5 |
| Brain_Tumor2 | 10367 | 50 | 4 | 2.14 |

## B. RESULTS AND DISCUSSION

This section reports the results obtained from the proposed methods. All experiments were performed using fitness criterion MAUC formulated using AUC in terms of multi-class imbalanced settings to show the efficacy of the proposed methods of handling the classification task for high dimensional multi-class imbalanced datasets. Comparing the proposed MFR-GOA, MFR-GOASA, GOASAT, and MFR-GOASAT algorithms, experiments are performed to assess the effectiveness of hybridization of SA with conventional GOA and the use of the TS mechanism rather than the RS mechanism. In order to determine the best method among the proposed methods, the performance of all four methods is evaluated in one table. The best approach among the proposed methods is also compared against other well-known state-of-the-art methods:

- MAUC accuracy rate using selected features proportion of both minority and majority class.
- To compare the proposed method best, mean, and worst fitness values against other well-known states of the art methods.
- Statistically compared the Wilcoxon signed-rank test of MFR-GOASAT against other approaches.

### 1) COMPARISON BETWEEN MFR-GOA, MFR-GOASA, MFR-GOASAT AND GOASAT

The effectiveness of MFR-GOA, GOASA, and MFR-GOASAT and GOASAT are evaluated based on MAUC predictive accuracy and computational time, as stated in this section. It can be recalled that in GOASA, SA is used after the MFR-GOA found the optimal solutions, then the SA utilizes the best solutions locates by MFR-GOA to improve the exploitation capability until local optima are reached. Table 3 shows that the hybrid models outperform the conventional ones for the MAUC accuracy rate. The traditional hybrid MFR-GOA algorithm does not outperform the MFR-GOASA on nine datasets regarding the MAUC accuracy rate. The MFR-GOASA outperforms MFR-GOA on nine datasets, and the MAUC accuracy rate between the two models varies from 2% to 6%. It can be observed on the Brain-Tumor2 dataset, and the MFR-GOASA method achieves the MAUC accuracy rate of 97.3% while the MFR-GOA achieves the MAUC accuracy rate of 87.5% result. The obtained results from Table 3 show that the MFR-GOASA method outperforms MFR-GOA on most datasets. When comparing the MFR-GOASA and MFR-GOASAT, it can be observed that the MFR-GOASAT outperforms the MFR-GOASA on most datasets. The GOASA achieves the best MAUC predictive accuracy on two datasets (i.e., 9_Tumor and 11_Tumor), while the MFR-GOASAT achieves the best predictive accuracy on four datasets (CLL-SUB_111, Glioma, Brain_Tumor1, and Brain_Tumor2) while MFR-GOASA and MFR-GOASAT tie to achieve the best results on two datasets (SRBCT and TOX-171). To evaluate the effect of using the MFR method, the MFR-GOASAT was compared with the GOASA-T algorithm. It can be observed that MFR-GOASAT achieves the best MAUC predictive accuracy on eight out of nine datasets (9_Tumor and 11_Tumor, CLL-SUB_111, Glioma, Brain_Tumor1, Brain_Tumor2, SRBCT, and TOX-171). The obtained results show that the filter approach (i.e., MFR) finds the discriminative features that use the as strong initial stage to wrapper method to find a most informative subset of features. The obtained results have shown that the hybridization with the filtering-based method (i.e., MFR) has improved classification performance and reduce the computational complexity of the wrapper method.

The experiments were conducted among proposed methods to evaluate the effect of improving conflicting optimization criteria of exploitation, exploration. It can be observed that employing the TS mechanism performs a supplementary task in improving exploration in the MFR-GOA algorithm along with SA, which improves exploitation capability. The performance of the proposed methods is improved inprogressive order: MFR-GOA < MFR-GOASA < MFR-GOASAT. It can be seen that MFR-GOASAT is a robust algorithm that achieves the proper balance between exploitation and exploration when searching for the global optimum. SA improved the exploitation, and the TS mechanism improved the population diversity in MFR-GOASAT, which supplemented the SA role. It can be seen that MFR-GOASAT achieves outstanding results on five datasets in terms of MAUC accuracy rate. The obtained results showed that MFR-GOASAT had achieved a proper balance between exploration and exploitation because of the used SA algorithm and TS mechanism.

Table 4 reports the average execution time (in seconds) needed for the convergence of each algorithm. All methods were formulated using the same parameter settings and evaluated on the same benchmark datasets. Table 4 reports that the MFR-GOA has the best total execution time than other method; the MFR-GOA has a better full execution time on nine datasets. When comparing the full execution time of MFR-GOASA and MFR-GOASAT, note that MFR-GOASAT uses the TS mechanism while the MFR-GOASA uses the RS mechanism. The MFR-GOASAT has a better execution time on seven out of nine datasets. On the other hand, the MFR-GOASA has a better total execution time on two out of nine datasets. This findingelucidated that the TS mechanism requires less computational time to convergence the proposed algorithm than the RS mechanism. When comparing the full execution time of MFR-GOASAT and GOASAT, the experimental results show that MFR-GOASAT has a better execution time.

**TABLE 3.** Experimental results of MFR-GOA, MFR-GOASA, MFR-GOASAT, and GOASAT in terms of MAUC accuracy rate.

| Dataset | Method | | | |
|---|---|---|---|---|
| | MFR-GOA | MFR-GOASA | MFR-GOASA-T | GOASA-T |
| CLL-SUB-111 | 93.6±19.52 | 95.5±9.45 | **95.6**±9.77 | 94.9±8.23 |
| TOX-171 | 91.8±5. 04 | **96.8**±12.26 | **96.8**±11.20 | 92.5±13.53 |
| GLIOMA | 92.6±16.35 | 96.5±9.54 | **97.3**±6.60 | 93.1±1.34 |
| 9_Tumour | 87.4±1.32 | **97.9**±3.33 | 96.8±8.75 | 95.2±6.67 |
| 11_Tumour | 93.1±11.28 | **98.7**±6.94 | 98.5±12.42 | 97.3±8.01 |
| SRBCT | 98.3±8.29 | **100** | **100** | 98.1±18.14 |
| Lung cancer | 95.4±3.91 | 97.3±8.21 | 98.2±5.43 | **99.3**±7.42 |
| Brain_Tumor1 | 88.7±8.78 | 93.5±5.37 | **96.4**±7.37 | 89.8±16.71 |
| Brain_Tumor2 | 87.5±8.21 | 95.3±14.45 | **97.3**±5.32 | 92.3±14.34 |

**TABLE 4.** Average convergence time (in seconds) for proposed methods in terms of MAUC accuracy rate.

| Dataset | Method | | | |
|---|---|---|---|---|
| | MFR-GOA | MFR-GOASA | MFR-GOASA-T | GOASA-T |
| CLL-SUB-111 | **139.57**±0.12 | 783.75±1.52 | 689.12±10.87 | 1131.12±0.83 |
| TOX-171 | **131.23**±1.11 | 693.84±2.79 | 676.65±15.18 | 983.58±0.68 |
| GLIOMA | **82.60**±0.83 | 428.34±3.12 | 412.34±25.21 | 872.34±1.02 |
| 9_Tumour | **217.12**±1.27 | 834.29±2.48 | 793.09±31.60 | 1156.10±6.36 |
| 11_Tumour | **144.46**±1.11 | 788.35±3.32 | 772.63±17.32 | 1098.23±7.95 |
| SRBCT | **91.38**±0.82 | 344.82±21.21 | 341.56±26.52 | 683.57±12.22 |
| Lung cancer | **125.84**±1.26 | 633.26±2.51 | 590.81±1.27 | 824.91±10.50 |
| Brain_Tumor1 | **78.72**±2.71 | 413.91±24.15 | 401.16±1.08 | 901.48±18.90 |
| Brain_Tumor2 | **85.51+12.42** | 516.42±0.35 | 452.53±0.85 | 889.62±12.89 |

This demonstrates the importance of using the filtering-based approach to finds discriminative features that use a strong initial stage to wrapper method to find a most informative subset of features.

### 2) COMPARISON OF THE MFR-GOASAT APPROACH WITH OTHER EXISTING METHODS

The experimental results showed that the MFR-GOASAT method outperforms other methods to achieve promising results on most datasets. Therefore, in this subsection, the best-proposed approach is compared with other well-known state-of-the-art methods for high dimensional multi-class imbalanced datasets regarding the MAUC accuracy rate. Table 5 shows the experimental results of MFR-GOASAT, ECOC-MDC, CBR-PSO, SMOTEBOOST, RUSBOOST, PSO, GA, GOA, and GWO. In order to demonstrate the significance of applying data pre-processing (i.e., feature selection and class imbalanced techniques) in the classification of high dimensional multi-class imbalanced datasets, the MAUC predictive accuracy using the full features is reported. From the experimental results obtained, the MAUC predictive accuracy using the full features shows poor performance on all datasets compared to the proposed high-dimensional multi-class imbalanced methods. Furthermore, MFR-GOASAT achieves the best MAUC accuracy rate on six out of nine datasets (i.e., Brain-Tumour1, Brain-Tumour2, Glioma, SRBCT, 9_Tumour, and CLL-SUB-111), which were significantly better than the MAUC accuracy achieved by other approaches. Furthermore, the CBR-PSO achieves the best MAUC accuracy rate on two out of nine datasets (i.e., Lung cancer and TOX-171), which were significantly better than the MAUC accuracy rate achieved by other approaches. Finally, the ECOC-MDC outperforms other methods to achieve the best MAUC accuracy rate on one out of nine datasets (i.e., 11_Tumour), which was significantly higher than the MAUC accuracy rate achieved by other approaches. Therefore, from the experimental results, it can be observed that the MFR-GOASAT algorithm that was proposed in this paper could identify and select the best feature subset proportion of both minority and majority class along with enhancing the MAUC accuracy rate for high dimensional multi-class imbalanced datasets.

### 3) STATISTICAL ANALYSIS

The Wilcoxon signed-rank test is used to determine significant differences between the proposed MFR-GOASAT method against other methods in the literature. The statistical analysis aims to evaluate if the results from the two approaches are statistically independent. The null hypothesis indicates the significant difference between the proposed MFR-GOASAT method and each other method. If the p-values exceed 5%, the null hypothesis is retained, which implies no significant improvement of using the MFR-GOASAT metho; otherwise is rejected at a significance level below 5%.

As shown in Table 6, the Wilcoxon signed-rank test was calculated using pairwise comparisons for the MAUC accuracy rate. The resulting test of p-values is below the significance level of 5% for the seven approaches. Therefore, there are statistically significant differences between the MFR-GOASAT and other methods (i.e., ECOC-MDC, SMOTEB, RUSB, PSO, GA, GWO, and Full). However, the Wilcoxon signed-rank test result for the one approach (i.e., CBR-PSO) is above the significance level of 5%. Statistically, there is no significant difference between the MFR-GOASAT and that of (i.e., the CBR-PSO) method.

Table 7 reports the comparative assessment of the proposed MFR-GOASAT method against state-of-the-art methods for high dimensional imbalanced datasets (i.e., PSO, GA, and GWO). The best-obtained result on each dataset is marked in **boldface**. The comparison of the proposed approaches is performed in terms of fitness criteria. The fitness value is one of the standard performance evaluation measures for the feature selection techniques for the high dimensional imbalanced approach [100]. It can be seen from Table 5 the MFR-GOASAT proposed in this paper outperforms PSO,

**TABLE 5. Experimental results of MFR-GOASAT in comparison with other approaches in terms of the MAUC metric.**

| Datasets | MFR-GOASAT | ECOC-MDC | CBR-PSO | SMOTE BOOST | RUS BOOST | PSO | GWO | GA | GOA | Full |
|---|---|---|---|---|---|---|---|---|---|---|
| Brain_Tumour1 | **96.4** | 85.1 | 91.7 | 50.4 | 54.1 | 88.1 | 87.6 | 89.3 | 91.2 | 81.5 |
| Brain_Tumour2 | **97.7** | 85.0 | 94.9 | 62.5 | 64.6 | 78.6 | 87.5 | 91.9 | 85.6 | 73.3 |
| Lung Cancer | 98.2 | 91.2 | **98.3** | 60.1 | 62.2 | 89.7 | 92.1 | 89.9 | 91.4 | 88.1 |
| SRBCT | **100** | 91.6 | 98.0 | 56.5 | 60.8 | 91.4 | 98.3 | 86.2 | 97.6 | 92.1 |
| 9_Tumours | **96.8** | 29.5 | 86.2 | 45.4 | 79.9 | 50.2 | 65.5 | 68.1 | 71.0 | 43.4 |
| 11_Tumours | 98.5 | **100** | 92.8 | 51.3 | 54.3 | 94.4 | 99.6 | 94.4 | 98.3 | 88.7 |
| Glioma | **97.3** | 76.1 | 92.0 | 60.0 | 59.3 | 93.3 | 91.4 | 92.5 | 90.8 | 87.1 |
| TOX-171 | 96.2 | 71.8 | **97.4** | 52.7 | 49.5 | 57.3 | 60.9 | 57.7 | 61.5 | 51.3 |
| CLL-SUB-11 | **95.6** | 86.4 | 88.0 | 49.0 | 65.7 | 71.6 | 74.6 | 76.3 | 74.7 | 72.3 |

**TABLE 6. Wilcoxon tests for the MFR-GOASAT against other methods in terms of MAUC accuracy rate the "Yes" indicates the there is a statistical difference at p = 0.05 (95% confidence); otherwise, "No".**

| Evaluation Metric | Comparison | Hypothesis | p-value | Significant difference |
|---|---|---|---|---|
| MAUC | MFR-GOASAT vs ECOC-MDC | **Rejected for MFR-GOASAT at 5%** | 0.0078 | Yes |
| | MFR-GOASAT vs CBR-PSO | Not rejected | 0.0547 | No |
| | MFR-GOASAT vs. SMOTEBOOST | **Rejected for MFR-GOASAT at 5%** | 0.0039 | Yes |
| | MFR-GOASAT vs. RUSBOOST | **Rejected for MFR-GOASAT at 5%** | 0.0052 | Yes |
| | MFR-GOASAT vs. -PSO | **Rejected for MFR-GOASAT at 5%** | 0.0061 | Yes |
| | MFR-GOASAT vs. GA | **Rejected for MFR-GOASAT at 5%** | 0.0075 | Yes |
| | MFR-GOASAT vs. GWO | **Rejected for MFR-GOASAT at 5%** | 0.0059 | Yes |
| | MFR-GOASAT vs. GOA | **Rejected for MFR-GOASAT at 5%** | 0.0064 | Yes |
| | MFR-GOASAT vs. Full | **Rejected for MFR-GOASAT at 5%** | 0.0039 | Yes |

**TABLE 7. Comparison of fitness values results obtained between the proposed method in terms of MAUC against other methods.**

| Dataset | Mean | | | | | Best | | | | | Worst | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MFR-GOASAT | PSO | GA | GWO | GOA | MFR-GOASAT | PSO | GA | GWO | GOA | MFR-GOASAT | PSO | GA | GWO | GOA |
| Brain_Tumour1 | 91.23 | **92.47** | 91.87 | 91.07 | 89.13 | **95.20** | 94.01 | 93.31 | 94.07 | 92.11 | 82.34 | **81.01** | 81.97 | 88.58 | 86.23 |
| Brain_Tumour2 | **93.25** | 90.24 | 91.43 | 92.04 | 93.94 | **97.66** | 95.83 | 96.26 | 96.1 | 95.83 | 89.92 | 88.23 | 89.32 | **88.19** | 89.51 |
| Lung Cancer | 92.1 | **93.53** | 92.42 | 91.92 | 92.10 | **97.81** | 96.93 | 94.24 | 95.97 | 97.18 | **79.13** | 83.53 | 81.03 | 87.03 | 84.12 |
| SRBCT | **96.23** | 92.34 | 94.09 | 95.92 | 91.02 | **99.23** | 97.51 | 99.01 | 98.31 | 98.94 | 94.34 | 90.34 | 87.29 | 90.07 | **84.67** |
| 9_Tumours | 89.03 | 72.42 | 88.56 | 69.47 | 79.10 | **97.78** | 79.93 | 93.91 | 70.71 | 89.53 | 83.15 | 68.81 | 86.16 | **64.16** | 81.81 |
| 11_Tumours | **91.18** | 90.23 | 89.07 | 89.23 | 89.33 | **95.41** | 94.9 | 93.21 | 94.28 | 91.54 | 87.23 | 88.51 | **78.12** | 86.19 | **81.12** |
| Glioma | 91.16 | 95.5 | **91.56** | 94.91 | 94.91 | 99.11 | 98.01 | 97.23 | **100** | 99.22 | 92.45 | 91.67 | **85.29** | 93.07 | 85.29 |
| TOX-171 | **91.47** | 88.09 | 79.22 | 78.63 | 78.63 | **98.86** | 94.09 | 92.12 | 87.49 | 95.23 | 85.34 | 82.88 | 74.81 | **73.08** | 77.26 |
| CLL-SUB-11 | 88.34 | 78.91 | **91.51** | 84.23 | 74.23 | 94.23 | 88.91 | **95.05** | 93.15 | 92.26 | 78.34 | 71.1 | 75.24 | **64.16** | 65.19 |
| Average | **92.00** | 88.19 | 89.97 | 87.49 | 86.93 | **97.25** | 93.35 | 94.93 | 92.23 | 94.68 | 85.80 | 82.90 | 82.14 | **81.61** | 81.68 |

GA, GOA, and GWO in best fitness function values, mean fitness function values on most datasets, and it is not worse than any other method on most datasets. The proposed MFR-GOASAT achieves outstanding results because of the usage of multiple filter-based rankers approach coupled with hybrid GOA algorithm with SA local search to improve the exploitatio capability and TS mechanism to preserve the population diversity of the feature search space.

The MFR-GOASAT proposed in this paper achieves promising results on most datasets compared to other methods due to the hybridization of filter-wrapper approaches and the enhanced conventional GOA algorithm. MFR-GOASAT achieves a proper balance between exploitation and exploration in all the performed iterations. The experimental results show that MFR-GOASAT is an effective method on large and limited dimensional datasets. Lung cancer and 11_Tumour are large datasets, and the fitness value of the MFR-GOASAT is significantly higher than in comparison with other methods. It can be observed that the MFR-GOASAT performs better than PSO, GA, GOA, and GWO based on the best and worst obtained fitness function.

## V. CONCLUSION

High dimensionality and imbalanced class distribution are two serious issues that degrade classification performance most conventional classifiers' classification performance. In the literature, numerous methods have been proposed to deal with either the high dimensionality or imbalanced class problem separately, but high dimensionality interferes in the class imbalanced techniques. In this paper, multiple filter-based rankers coupled with hybrid metaheuristic techniques using the GOA algorithm were proposed (MFR-GOA, MFR-GOASA, and MFR-GOASAT). The proposed methods incorporate SA into the GOA algorithm. SA was utilized in the proposed methods, a hybrid model called High-Level Transmit Hybrid (HTH) was proposed.

In the HTH, SA was used to search the neighbourhood of the best-found solution after each iteration of the GOA algorith; two models were proposed using these algorithms, namely MFR-GOASA and MFR-GOASAT. The TS mechanism was applied to select the search agents rather than the RS mechanism (MFR-GOASAT) to provide a chance for the weak solutions to be selected in the searching process, maintain the population diversity and improve the exploration capabilities of the GOA algorithm. The three proposed MFR-GOA, MFR-GOASA, and MFR-GOASAT methods have been evaluated based on MAUC and total execution time (in seconds). The obtained results demonstrate that MFR-GOASAT achieves the best performance among the proposed methods. The experimental results obtained are inprogressive order: MFR-GOA<MFR-GOASA<MFR-GOASAT.

Furthermore, The best method among the proposed methods (i.e., MFR-GOASAT) was compared against state-of-the-art methods (i.e., include ECOC-MDC, CBR-PSO, SMOTEBOOST, RUSBOOST, PSO, GA, GOA, GWO, and Full features) in terms of MAUC predictive accuracy and non-parametric statistical tests to assess the statistical significance of the proposed method against other methods. In terms of the MAUC accuracy rate, the proposed MFR-GOASAT achieves the best MAUC accuracy rate on six out of nine datasets, and in terms of the Wilcoxon statistical test, the obtained results show that the proposed method is statistically significant against seven other approaches. Therefore, from the experimental results obtained categorically, we can deduce that the proposed methods achieve the proper balance between the two conflicting objectives of the metaheuristic (exploration and exploitation) and improve the classification performance for high dimensional multi-class imbalanced datasets.

Future works suggest that the fitness function for the high-dimensional multi-class imbalanced problem will be modified to reduce the computational complexity of the current approaches. It would be suggesting that to hybridize the GOA algorithm with other global optimization approaches such as the cuckoo search algorithm.

## REFERENCES

[1] A. G. Sharifai and Z. Zainol, "The correlation-based redundancy multiple-filter approach for gene selection," *Int. J. Data Mining Bioinf.*, vol. 23, no. 1, pp. 62–78, 2020.

[2] W.-J. Lin and J. J. Chen, "Class-imbalanced classifiers for high-dimensional data," *Briefings Bioinf.*, vol. 14, no. 1, pp. 13–26, Jan. 2013.

[3] V. A. Fajardo, D. Findlay, R. Houmanfar, C. Jaiswal, J. Liang, and H. Xie, "VOS: A method for variational oversampling of imbalanced data," 2018, *arXiv:1809.02596*. [Online]. Available: http://arxiv.org/abs/1809.02596

[4] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 4, pp. 463–484, Jul. 2012.

[5] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[6] J. L. Leevy, T. M. Khoshgoftaar, R. A. Bauder, and N. Seliya, "A survey on addressing high-class imbalance in big data," *J. Big Data*, vol. 5, no. 1, p. 42, Dec. 2018.

[7] S. Wang, L. L. Minku, and X. Yao, "A systematic study of online class imbalance learning with concept drift," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4802–4821, Oct. 2018.

[8] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002.

[9] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.* Springer, 2005, pp. 878–887.

[10] S.-J. Yen and Y.-S. Lee, "Cluster-based under-sampling approaches for imbalanced data distributions," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 5718–5727, Apr. 2009.

[11] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw., IEEE World Congr. Comput. Intell.*, Mar. 2008, pp. 1322–1328.

[12] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," in *Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery*, 2003, pp. 107–119.

[13] D. Tao, X. Tang, X. Li, and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1088–1099, Jul. 2006.

[14] F. S. Hanifah, H. Wijayanto, and A. Kurnia, "SMOTE bagging algorithm for imbalanced dataset in logistic regression analysis (case: Credit of bank X)," *Appl. Math. Sci.*, vol. 9, no. 138, pp. 6857–6865, 2015.

[15] G.-Z. Li, H.-H. Meng, W.-C. Lu, J. Y. Yang, and M. Yang, "Asymmetric bagging and feature selection for activities prediction of drug molecules," *BMC Bioinf.*, vol. 9, no. 6, p. S7, 2008.

[16] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, Jan. 2006.

[17] C. Elkan, "The foundations of cost-sensitive learning," in *Proc. Int. Joint Conf. Artif. Intell.*, vol. 17, no. 1. Trenton, NJ, USA: Lawrence Erlbaum Associates, 2001, pp. 973–978.

[18] C. Ling and V. Sheng, "Cost-sensitive learning and the class imbalance problem," *Encyclopedia Mach. Learn.*, vol. 2011, pp. 231–235, Jan. 2008.

[19] K. Hempstalk, E. Frank, and I. H. Witten, "One-class classification by combining density and class probability estimation," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2008, pp. 505–519.

[20] H. J. Shin, D.-H. Eom, and S.-S. Kim, "One-class support vector machines—An application in machine fault detection and classification," *Comput. Ind. Eng.*, vol. 48, no. 2, pp. 395–408, Mar. 2005.

[21] K.-K. Seo, "An application of one-class support vector machines in content-based image retrieval," *Expert Syst. Appl.*, vol. 33, no. 2, pp. 491–498, Aug. 2007.

[22] S. Ertekin, J. Huang, L. Bottou, and L. Giles, "Learning on the border: Active learning in imbalanced data classification," in *Proc. 16th ACM Conf. Conf. Inf. Knowl. Manage. (CIKM)*, 2007, pp. 127–136.

[23] S. Ertekin, J. Huang, and C. L. Giles, "Active learning for class imbalance problem," in *Proc. 30th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 2007, pp. 823–824.

[24] J. Attenberg and S. Ertekin, "Class imbalance and active learning," in *Imbalanced Learning: Foundations, Algorithms, and Applications*, 2013, pp. 101–149.

[25] H. Yu and J. Ni, "An improved ensemble learning method for classifying high-dimensional and imbalanced biomedicine data," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 11, no. 4, pp. 657–666, Jul. 2014.

[26] C. Zhang, Y. Zhou, J. Guo, G. Wang, and X. Wang, "Research on classification method of high-dimensional class-imbalanced datasets based on SVM," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 7, pp. 1765–1778, Jul. 2019.

[27] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, p. 27, Dec. 2019.

[28] M. Liu, C. Xu, Y. Luo, C. Xu, Y. Wen, and D. Tao, "Cost-sensitive feature selection by optimizing F-measures," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1323–1335, Mar. 2018.

[29] M. Wasikowski and X.-W. Chen, "Combating the small sample class imbalance problem using feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1388–1400, Oct. 2010.

[30] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, Nov. 2016.

[31] S. Wang, Z. Li, W. Chao, and Q. Cao, "Applying adaptive over-sampling technique based on data density and cost-sensitive SVM to imbalanced learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–8.

[32] A. Fernández, V. López, M. Galar, M. J. del Jesus, and F. Herrera, "Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches," *Knowl.-Based Syst.*, vol. 42, pp. 97–110, Apr. 2013.

[33] S. García, Z.-L. Zhang, A. Altalhi, S. Alshomrani, and F. Herrera, "Dynamic ensemble selection for multi-class imbalanced datasets," *Inf. Sci.*, vols. 445–446, pp. 22–37, Jun. 2018.

[34] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowl.-Based Syst.*, vol. 82, pp. 29–40, Jul. 2015.

[35] A. Moayedikia, K.-L. Ong, Y. L. Boo, W. G. Yeoh, and R. Jensen, "Feature selection for high dimensional imbalanced class data using harmony search," *Eng. Appl. Artif. Intell.*, vol. 57, pp. 38–49, Jan. 2017.

[36] G. A. Sharifai and Z. Zainol, "Feature selection for high-dimensional and imbalanced biomedical data based on robust correlation based redundancy and binary grasshopper optimization algorithm," *Genes*, vol. 11, no. 7, p. 717, Jun. 2020.

[37] P. Zhou, X. Hu, P. Li, and X. Wu, "Online feature selection for high-dimensional class-imbalanced data," *Knowl.-Based Syst.*, vol. 136, pp. 187–199, Nov. 2017.

[38] M. Liu and D. Zhang, "Feature selection with effective distance," *Neurocomputing*, vol. 215, pp. 100–109, Nov. 2016.

[39] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, p. 94, 2016.

[40] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, Jan. 2003.

[41] I. Guyon, "Practical feature selection: From correlation to causality," in *Mining Massive Data Sets for Security: Advances in Data Mining, Search, Social Networks and Text Mining, and Their Applications to Security*. IOS Press, 2008, pp. 27–43.

[42] Q. Cheng, P. K. Varshney, and M. K. Arora, "Logistic regression for feature selection and soft classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 491–494, Oct. 2006.

[43] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Comput. Statist. Data Anal.*, vol. 143, Mar. 2020, Art. no. 106839.

[44] A. Rouhi and H. Nezamabadi-Pour, "A hybrid-ensemble based framework for microarray data gene selection," *Int. J. Data Mining Bioinf.*, vol. 19, no. 3, pp. 221–242, 2017.

[45] O. Sahin and B. Akay, "Comparisons of metaheuristic algorithms and fitness functions on software test data generation," *Appl. Soft Comput.*, vol. 49, pp. 1202–1214, Dec. 2016.

[46] L. Brezočnik and V. Podgorelec, "Applying weighted particle swarm optimization to imbalanced data in software defect prediction," in *Proc. Int. Conf. New Technol., Develop. Appl.*, 2018, pp. 289–296.

[47] H. B. Nguyen, B. Xue, I. Liu, and M. Zhang, "PSO and statistical clustering for feature selection: A new representation," 2014, pp. 569–581.

[48] C. Blum, J. Puchinger, G. R. Raidl, and A. Roli, "Hybrid metaheuristics in combinatorial optimization: A survey," *Appl. Soft Comput.*, vol. 11, no. 6, pp. 4135–4151, Sep. 2011.

[49] Eberhart and Y. Shi, "Particle swarm optimization: Developments, applications and resources," in *Proc. Congr. Evol. Comput.*, 2001, pp. 81–86.

[50] D. Whitley, "A genetic algorithm tutorial," *Statist. Comput.*, vol. 4, no. 2, pp. 65–85, Jun. 1994.

[51] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: Theory and application," *Adv. Eng. Softw.*, vol. 105, pp. 30–47, Mar. 2017.

[52] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Gener. Comput. Syst.*, vol. 111, pp. 300–323, Oct. 2020.

[53] Q. Askari, M. Saeed, and I. Younas, "Heap-based optimizer inspired by corporate rank hierarchy for global optimization," *Expert Syst. Appl.*, vol. 161, Dec. 2020, Art. no. 113702.

[54] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Future Gener. Comput. Syst.*, vol. 97, pp. 849–872, Aug. 2019.

[55] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine predators algorithm: A nature-inspired Metaheuristic," *Expert Syst. Appl.*, vol. 152, Aug. 2020, Art. no. 113377.

[56] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: A novel optimization algorithm," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105190.

[57] W. Zhao, Z. Zhang, and L. Wang, "Manta ray foraging optimization: An effective bio-inspired optimizer for engineering applications," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103300.

[58] F. A. Hashim, K. Hussain, E. H. Houssein, M. S. Mabrouk, and W. Al-Atabany, "Archimedes optimization algorithm: A new metaheuristic algorithm for solving optimization problems," *Int. J. Speech Technol.*, vol. 51, no. 3, pp. 1531–1551, Mar. 2021.

[59] W. Pei, B. Xue, L. Shang, and M. Zhang, "New fitness functions in genetic programming for classification with high-dimensional unbalanced data," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, Jun. 2019, pp. 2779–2786.

[60] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[61] L. Xu, Y. Tu, and Y. Zhang, "A grasshopper optimization-based approach for task assignment in cloud logistics," *Math. Problems Eng.*, vol. 2020, pp. 1–10, Apr. 2020.

[62] S. Łukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, "Data clustering with grasshopper optimization algorithm," in *Proc. Federated Conf. Comput. Sci. Inf. Syst.*, Sep. 2017, pp. 71–74.

[63] H. Jia, C. Lang, D. Oliva, W. Song, and X. Peng, "Hybrid grasshopper optimization algorithm and differential evolution for multilevel satellite image segmentation," *Remote Sens.*, vol. 11, no. 9, p. 1134, May 2019.

[64] S. Arora and P. Anand, "Chaotic grasshopper optimization algorithm for global optimization," *Neural Comput. Appl.*, vol. 31, no. 8, pp. 4385–4405, Aug. 2019.

[65] J. Luo, H. Chen, Q. Zhang, Y. Xu, H. Huang, and X. Zhao, "An improved grasshopper optimization algorithm with application to financial stress prediction," *Appl. Math. Model.*, vol. 64, pp. 654–668, Dec. 2018.

[66] H. Liang, H. Jia, Z. Xing, J. Ma, and X. Peng, "Modified grasshopper algorithm-based multilevel thresholding for color image segmentation," *IEEE Access*, vol. 7, pp. 11258–11295, 2019.

[67] H. Jia, C. Lang, D. Oliva, W. Song, and X. Peng, "Hybrid grasshopper optimization algorithm and differential evolution for multilevel satellite image segmentation," *Remote Sens.*, vol. 11, no. 9, p. 1134, 2019.

[68] M. A. Taher, S. Kamel, F. Jurado, and M. Ebeed, "Modified grasshopper optimization framework for optimal power flow solution," *Electr. Eng.*, vol. 101, no. 1, pp. 121–148, Apr. 2019.

[69] A. Sezavar, H. Farsi, and S. Mohamadzadeh, "A modified grasshopper optimization algorithm combined with CNn for content based image retrieval," *Int. J. Eng.*, vol. 32, no. 7, pp. 924–930, 2019.

[70] A. Tharwat, E. H. Houssein, M. M. Ahmed, A. E. Hassanien, and T. Gabel, "MOGOA algorithm for constrained and unconstrained multi-objective optimization problems," *Int. J. Speech Technol.*, vol. 48, no. 8, pp. 2268–2283, Aug. 2018.

[71] S. Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, and I. Aljarah, "Grasshopper optimization algorithm for multi-objective optimization problems," *Appl. Intell.*, vol. 48, no. 4, pp. 805–820, Apr. 2018.

[72] A. A. Ewees, M. A. Elaziz, and E. H. Houssein, "Improved grasshopper optimization algorithm using opposition-based learning," *Expert Syst. Appl.*, vol. 112, pp. 156–172, Dec. 2018.

[73] A. A. Amaireh, A. S. Al-Zoubi, and N. I. Dib, "Sidelobe-level suppression for circular antenna array via new hybrid optimization algorithm based on antlion and grasshopper optimization algorithms," *Prog. Electromagn. Res.*, vol. 93, pp. 49–63, 2019.

[74] S. Maldonado, R. Weber, and F. Famili, "Feature selection for high-dimensional class-imbalanced data sets using support vector machines," *Inf. Sci.*, vol. 286, pp. 228–246, Dec. 2014.

[75] A. Fernández, S. del Río, N. V. Chawla, and F. Herrera, "An insight into imbalanced big data classification: Outcomes and challenges," *Complex Intell. Syst.*, vol. 3, no. 2, pp. 105–120, Jun. 2017.

[76] G. I. Winata and M. L. Khodra, ‘‘Handling imbalanced dataset in multi-label text categorization using bagging and adaptive boosting,’’ in *Proc. Int. Conf. Electr. Eng. Informat. (ICEEI)*, Aug. 2015, pp. 500–505.
[77] A. Reyes-Nava, H. Cruz-Reyes, R. Alejo, E. Rendón-Lara, A. Flores-Fuentes, and E. Granda-Gutiérrez, ‘‘Using deep learning to classify class imbalanced gene-expression microarrays datasets,’’ in *Proc. Iberoamerican Congr. Pattern Recognit.* Springer, 2018, pp. 46–54.
[78] X.-M. Zhao, X. Li, L. Chen, and K. Aihara, ‘‘Protein classification with imbalanced data,’’ *Proteins,Struct., Function, Bioinf.*, vol. 70, no. 4, pp. 1125–1132, Dec. 2007.
[79] F. Herrera, F. Charte, A. J. Rivera, and M. J. Del Jesus, ‘‘Multilabel classification,’’ in *Multilabel Classification*. Springer, 2016, pp. 17–31.
[80] M. R. Bakhtiarizadeh, M. Rahimi, A. Mohammadi-Sangcheshmeh, V. Shariati J, and S. A. Salami, ‘‘PrESOgenesis: A two-layer multi-label predictor for identifying fertility-related proteins using support vector machine and pseudo amino acid composition approach,’’ *Sci. Rep.*, vol. 8, no. 1, p. 9025, Dec. 2018.
[81] R. Rifkin and A. Klautau, ‘‘In defense of one-vs-all classification,’’ *J. Mach. Learn. Res.*, vol. 5, pp. 101–141, Dec. 2004.
[82] J. A. Sáez, M. Galar, J. Luengo, and F. Herrera, ‘‘Analyzing the presence of noise in multi-class problems: Alleviating its influence with the one-vs-one decomposition,’’ *Knowl. Inf. Syst.*, vol. 38, no. 1, pp. 179–206, Jan. 2014.
[83] T. Hastie and R. Tibshirani, ‘‘Classification by pairwise coupling,’’ in *Proc. Adv. Neural Inf. Process. Syst.*, 1998, pp. 507–513.
[84] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera, ‘‘An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes,’’ *Pattern Recognit.*, vol. 44, no. 8, pp. 1761–1776, Aug. 2011.
[85] T. G. Dietterich and G. Bakiri, ‘‘Solving multiclass learning problems via error-correcting output codes,’’ *J. Artif. Intell. Res.*, vol. 2, pp. 263–286, Jan. 1995.
[86] B. Kijsirikul and N. Ussivakul, ‘‘Multiclass support vector machines using adaptive directed acyclic graph,’’ in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2002, pp. 980–985.
[87] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, ‘‘A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis,’’ *Bioinformatics*, vol. 21, no. 5, pp. 631–643, Mar. 2005.
[88] J. Yang, J. Zhou, Z. Zhu, X. Ma, and Z. Ji, ‘‘Iterative ensemble feature selection for multiclass classification of imbalanced microarray data,’’ *J. Biol. Res.-Thessaloniki*, vol. 23, no. S1, p. 13, May 2016.
[89] L.-M. Du, Y. Xu, and H. Zhu, ‘‘Feature selection for multi-class imbalanced data sets based on genetic algorithm,’’ *Ann. Data Sci.*, vol. 2, no. 3, pp. 293–300, Sep. 2015.
[90] H. Yu, S. Hong, X. Yang, J. Ni, Y. Dan, and B. Qin, ‘‘Recognition of multiple imbalanced cancer types based on DNA microarray data using ensemble classifiers,’’ *BioMed Res. Int.*, vol. 2013, pp. 1–13, Jul. 2013.
[91] R. Arias-Michel, M. Garcia-Torres, C. E. Schaerer, and F. Divina, ‘‘Feature selection via approximated Markov blankets using the CFS method,’’ in *Proc. Int. Workshop Data Mining Ind. Appl. (DMIA)*, Sep. 2015, pp. 38–43.
[92] E. S. Hosseini and M. H. Moattar, ‘‘Evolutionary feature subsets selection based on interaction information for high dimensional imbalanced data classification,’’ *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105581.
[93] M. Sun, K. Liu, Q. Wu, Q. Hong, B. Wang, and H. Zhang, ‘‘A novel ECOC algorithm for multiclass microarray data classification based on data complexity analysis,’’ *Pattern Recognit.*, vol. 90, pp. 346–362, Jun. 2019.
[94] E. K. Aydogan, M. Ozmen, and Y. Delice, ‘‘CBR-PSO: Cost-based rough particle swarm optimization approach for high-dimensional imbalanced problems,’’ *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6345–6363, Oct. 2019.
[95] S. Gu, R. Cheng, and Y. Jin, ‘‘Feature selection for high-dimensional classification using a competitive swarm optimizer,’’ *Soft Comput.*, vol. 22, no. 3, pp. 811–822, Feb. 2018.
[96] M. Mafarja, I. Aljarah, A. A. Heidari, A. I. Hammouri, H. Faris, A. M. Al-Zoubi, and S. Mirjalili, ‘‘Evolutionary population dynamics and grasshopper optimization approaches for feature selection problems,’’ *Knowl.-Based Syst.*, vol. 145, pp. 25–45, Apr. 2018.
[97] S. Mirjalili and A. Lewis, ‘‘S-shaped versus V-shaped transfer functions for binary particle swarm optimization,’’ *Swarm Evol. Comput.*, vol. 9, pp. 1–14, Apr. 2013.
[98] J. Kennedy and R. C. Eberhart, ‘‘A discrete binary version of the particle swarm algorithm,’’ in *Proc. IEEE Int. Conf. Syst., Man, Cybernetics. Comput. Cybern. Simul.*, Oct. 1997, pp. 4104–4108.
[99] A.-R. Hedar and R. Ismail, ‘‘Simulated annealing with stochastic local search for minimum dominating set problem,’’ *Int. J. Mach. Learn. Cybern.*, vol. 3, no. 2, pp. 97–109, Jun. 2012.
[100] M. M. Mafarja and S. Mirjalili, ‘‘Hybrid whale optimization algorithm with simulated annealing for feature selection,’’ *Neurocomputing*, vol. 260, pp. 302–312, Oct. 2017.
[101] J. Grefenstette, ‘‘Optimization of control parameters for genetic algorithms,’’ *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-16, no. 1, pp. 122–128, Jan. 1986.
[102] G. Sanchita and D. Anindita, ‘‘Evolutionary algorithm based techniques to handle big data,’’ in *Techniques and Environments for Big Data Analysis*. Springer, 2016, pp. 113–158.
[103] M. B. Dowlatshahi, V. Derhami, and H. Nezamabadi-Pour, ‘‘A novel three-stage filter-wrapper framework for miRNA subset selection in cancer classification,’’ *Informatics*, vol. 5, no. 1, p. 13, 2018.
[104] B. Wutzl, K. Leibnitz, F. Rattay, M. Kronbichler, M. Murata, and S. M. Golaszewski, ‘‘Genetic algorithms for feature selection when classifying severe chronic disorders of consciousness,’’ *PLoS ONE*, vol. 14, no. 7, Jul. 2019, Art. no. e0219683.
[105] O. A. Alomari, A. T. Khader, M. A. Al-Betar, and M. A. Awadallah, ‘‘A novel gene selection method using modified MRMR and hybrid bat-inspired algorithm with $\beta$-hill climbing,’’ *Int. J. Speech Technol.*, vol. 48, no. 11, pp. 4429–4447, Nov. 2018.
[106] A. Tharwat, ‘‘Classification assessment methods,’’ *Appl. Comput. Informat.*, vol. 17, no. 1, pp. 168–192, Jul. 2020.
[107] Q. Gu, L. Zhu, and Z. Cai, ‘‘Evaluation measures of the classification performance of imbalanced data sets,’’ in *Proc. Int. Symp. Intell. Comput. Appl.* Springer, 2009, pp. 461–471.
[108] P. Yang, W. Liu, B. B. Zhou, S. Chawla, and A. Y. Zomaya, ‘‘Ensemble-based wrapper methods for feature selection and class imbalance learning,’’ in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Springer, 2013, pp. 544–555.
[109] L. Yin, Y. Ge, K. Xiao, X. Wang, and X. Quan, ‘‘Feature selection for high-dimensional imbalanced data,’’ *Neurocomputing*, vol. 105, pp. 3–11, Apr. 2013.
[110] M. Črepinšek, S.-H. Liu, and M. Mernik, ‘‘Exploration and exploitation in evolutionary algorithms: A survey,’’ *ACM Comput. Surveys*, vol. 45, no. 3, pp. 1–33, Jun. 2013.
[111] E. Emary, H. M. Zawbaa, and A. E. Hassanien, ‘‘Binary ant lion approaches for feature selection,’’ *Neurocomputing*, vol. 213, pp. 54–65, Nov. 2016.

**ABDULRAUF GARBA SHARIFAI** (Graduate Student Member, IEEE) received the B.Sc. degree in computer science from Bayero University, Kano, in 2010, and the M.Eng. degree in computer engineering from the Liaoning University of Technology (LUT), China, in 2014. He is currently pursuing the Ph.D. degree with the School of Computer Sciences, Universiti Sains Malaysia. He is also a Lecturer with the Department of Computer Science, Yusuf Maitama Sule University, Kano. He has published numerous international journals and conferences. His research interests include evolutionary algorithms, data mining, feature selection, class imbalance, bioinformatics, and computational intelligence.

**ZURINAHNI BINTI ZAINOL** received the B.Sc. degree in computer science and the master's degree in artificial intelligence from Universiti Kebangsaan Malaysia (UKM) and the Ph.D. degree in computer sciences from the University of Hull, U.K., in 2012. She is currently an Associate Professor with the School of Computer Sciences, Universiti Sains Malaysia. She also holds the position of the deputy dean for academic, career, and international. She has published many high-quality articles in international journals, conferences, and book chapters. Her research interests include the theory of database design and management document schema design, data model, software development using formal specification, and artificial intelligence.