Research Article

Enhanced Feature Selection Based on Integration Containment Neighborhoods Rough Set Approximations and Binary Honey Badger Optimization

Rodyna A. Hosny, Mohamed Abd Elaziz, and Rehab Ali Ibrahim

1Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt
2Academy of Scientific Research and Technology (ASRT), Ajman University, Cairo, Egypt
3Faculty of Computer Science & Engineering, Galala University, Suze 435611, Egypt
4Artificial Intelligence Research Center (AIRC), College of Engineering and Information Technology, Ajman University, Ajman, UAE

Correspondence should be addressed to Mohamed Abd Elaziz; abd_el_aziz_m@yahoo.com

Received 17 October 2021; Revised 29 November 2021; Accepted 14 December 2021; Published 10 March 2022

Academic Editor: Radu-Emil Precup

Copyright © 2022 Rodyna A. Hosny et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article appoints a novel model of rough set approximations (RSA), namely, rough set approximation models build on containment neighborhoods RSA (CRSA), that generalize the traditional notions of RSA and obtain valuable consequences by minifying the boundary areas. To justify this extension, it is integrated with the binary version of the honey badger optimization (HBO) algorithm as a feature selection (FS) approach. The main target of using this extension is to assess the quality of selected features. To evaluate the performance of BHBO based on CRS, a set of ten datasets is used. In addition, the results of BHOB are compared with other well-known FS approaches. The results show the superiority of CRS over the traditional RS approximations. In addition, they illustrate the high ability of BHBO to improve the classification accuracy overall the compared methods in terms of performance metrics.

1. Introduction

In recent days, the high dimensionality [1] became a big problem [2] in different fields such as human activities recognition [3], silicon-on-insulator FinFETs [4], nonlinear servo systems [5], computer vision [6], processing of IoT data [7], and feature selection. Moreover, utilizing the techniques of feature selection (FS) appeared more in many problems like the metaheuristic search techniques [8–10], for example, salp swarm algorithm (SSA) [11], grey-wolf optimization algorithm (GWO) [12], conflict monitoring optimization [13], runner-root algorithm (RRA) [14], boosting arithmetic optimization algorithm [15], electric fish-based arithmetic optimization algorithm [16], fractional calculus-based slime mould algorithm [17], and moth-flame optimization (MFO) [18]. In [19], the salp swarm algorithm was hybridized with the particle swarm optimization algorithm, where the hybridized algorithm is noted as SSAPSO, which is efficient in the procedures of exploitation and exploration. Khotimah et al. [20] hybridized the genetic algorithm (GA) with the naïve Bayes classification (NBC) to perform the exploration procedure through the classification of some incomplete data experiments. Other enhancements on other metaheuristic algorithms like the grey wolf optimization algorithm (GWO) have been performed, for example, in [21], the authors enhanced the exploration and the exploitation activity of GWO and used the new algorithm for selecting features for the galaxies images then classifying them. They utilized the opposition-based learning (OBL) and the chaotic logistic map for recommending solutions for avoiding the drawbacks of the random solutions. In addition to the above, utilizing the operators of DE with GWO as local operators improves the exploitation ability of GWO, and such hybrid solutions are updated. In addition, the
disruption operator (DO) is efficient for the exploration procedure and keeps the diversity in the population of the solutions. Yet, in the data analysis problems, the uncertainty is a big problem which may cause defect in the problem solution.

Recently, the rough set theory (RS) [22, 23] has been used as an efficient tool in solving such problem. Many recent applications included utilizing RS for reducing the dimension such as feature selection [24], pattern recognition [24], and machine learning [25]. In such context, more work is in the literature review as in [26], in which the authors developed a filter feature selection method which was based on utilizing the rough set theory for identifying all class documents. A multiplication of the class by a parameter in which the documents that are proved to be existed inside is performed, and then the authors utilized the new technique for text classification. In [27], Nabwey et al. introduced using the rough set with the hypergraph for determining the relevant subset of features and utilized the proposed technique for the wart treatment prediction. In [28], Zhao et al. proposed using the rough set theory, especially the representative entropy, and the proposed method called classified nested equivalence class (CNEC) has the ability for computing the information entropy and significance for performing the feature selection. The authors tested the proposed method on the KDD Cup competition and some datasets from the UCI repository. But, the rough set theory has a drawback in dealing the feature selection that it selects only one feature in each iteration and it has a difficulty in dealing with the real-world applications which require that RS performs features discretization into many partitions. Moreover, RS is used with the metaheuristic techniques that have the ability for avoiding the RS obstacles. From the examples in the literature, in [29], the authors introduced using RS with the binary whale optimization algorithm and tested the performance on 32 datasets taken from the repository of UCI machine learning. Ibrahim et al. in [30] combined the runner-root algorithm (RRA) with RS and also with NRS and utilized the proposed technique for the galaxies images classification after selecting the relevant features. Acharjya [31] merged RS with artificial bee colony (ABC) algorithm for the application of the hepatitis disease diagnosis. Jothi [32] combined RS with the firefly-based quick reduct and used the proposed technique, RS firefly-based quick reduct (TRSFFQR), for dealing the MRI brain images. Tawhid et al. [33] combined two metaheuristic algorithms, binary particle swarm optimization and flower pollination algorithm, with RS, and then found the binary version of the combination and used it for solving some binary problems. Patra and Barman [34] proposed using RS with developing the hyperspectral band selection method for hyperspectral band selection. Patra et al. [35] presented a multiobjective FS method depending on the cultural algorithm combined with RS and compared their method with other methods. Sahlol et al. [36] proposed combining RS with the whale optimization algorithm (WOA) and used the modified algorithm for the recognition of the handwritten Arabic optical. Jothi et al. [37] combined RS with the Jaya optimization and used the modified method as feature selection method after that using it for acute lymphoblastic leukemia classification. In [38], Mafarja and Mirjalili combined the ant lion optimization (ALO) with RS for solving some classification problems. Reddy et al. [39] combined RS with the fuzzy rule for extracting the features from the dataset of the heart disease, then selecting the relevant features for performing classification. Zu et al. [40] introduced the neighborhood RS combined with the fish swarm algorithm for solving the feature selection of some datasets, where the proposed method depended also on combining the tolerance rough set (TRS) and firefly algorithm (FA). Such techniques which are used for solving feature selection problems use RS, especially elementary sets according to some classes called equivalence ones where the equivalence can be applied just for complete data, so it may be not suitable for many cases.

The issue of incomplete knowledge became a pivotal problem for several researchers, essentially in the domain of information system. There are different methods to understand indistinctness and doubtful knowledge, one of them is rough set theory (RST). The explanation of RST counts on relating one subset with two sets called upper and lower approximations that are employed to set the boundary region and accuracy degree of that subset. RST was initiated by Pawlak [41] which has been generalized by many methods as [42–54]. What interests us around those ways whose ideas that are motivated by topology, for instance, the methods of structure lower and upper approximations utilizing diverse kinds of neighborhoods, such as \( zR, zR \) neighborhoods [53, 55], \( (z)R, R(z) \) neighborhoods [47], and \( zRz, R(z)z \) neighborhoods [50, 54]. Neighborhoods are significant approaches to decrease the boundary region and enhance the accuracy measure. The wish of maximizing the accuracy degree of any subset is a prime motivation factor for introducing C-neighborhoods system [44] that exemplifies a beneficial tool to increase the lower approximation and decrease the upper approximation compared with \( R \)-neighborhoods given in [44, 46]. In addition, C-neighborhoods preserved most characteristics of Pawlak’s approximations compared with the other systems of neighborhoods.

Furthermore, metaheuristic techniques, MHT, have more drawbacks like that more MHT depend on the evolutionary process like the evolutionary algorithms or extension for them, and also, the solutions quality is influenced by the stuck local optima. In addition, these MHT cannot solve all problems with the same efficiency according to the no free lunch theorem [56]. Therefore, this motivated us to propose an alternative FS method that depends on a recent efficient MHT called honey badger optimization (HBO) [57] algorithm that simulates the behaviour of honey badger to catch its prey. According to the mathematical model of these behaviours, HBO has been applied to solve global optimization problem and engineering problems [57]. In addition, we developed an extension of RST, named containment neighborhoods RSA (CRSA), as fitness value. According to our Knowledge, this is the first time the HBA is combined with extension rough set and used as FS method. In addition, the CRSA is a new extension of RST that is not used yet.
The HBOCRSA is used for solving the feature selection, which is initialized where the dataset is split into two parts, training and testing. After that, initialize the first set of candidate solutions by utilizing the containment neighborhoods RSA (CRSA) as a new extension of knowledge which represents a fitness function to use the training part for assessing each solution quality. Then, reaching the finest solution and using the operators of the proposed method for modernizing the current agents, as the stopping condition is met, the updating process is performed; then, obtaining the optimal solution can be used to remove the irrelevant features and evaluate the classification performance, then reducing the testing set of features.

The main contribution of this study can be summarized as follows:

(1) Propose an extension of rough set approximation (RSA) named containment neighborhoods RSA (CRSA). The new extension generalizes the traditional concepts of RSA and obtains valuable consequences by minifying the boundary areas.
(2) Propose a FS approach which combines the binary version of HBO (BHBO) and the new RS approximations (i.e., CRSA).
(3) Assess the performance of the proposed FS approach using different datasets; as well as, compare the results of well-known FS method with the results of developed method. Moreover, evaluate the performance of the competitive FS method based on CRSA with traditional RS approximation.

The rest of this paper is organized as follows: in Section 2, the basic notation about multiknowledge rough set is given. Section 3 presents the extension of the rough set approximations based on containment neighborhoods (CRSA). In Section 4, the steps of the proposed feature selection method are introduced. Experimental results and discussion are given in Section 5. The conclusion and future work are discussed in Section 6.

2. Preliminaries

Let $\Lambda$ be a universe (nonempty finite set), and $R$ be any relation on $\Lambda$, i.e., $R \subseteq \Lambda \times \Lambda$. A form $(w, z) \in R$ means that $w$ is in relation $R$ with $z$, which is abridged as $wRz$.

**Definition 1.** Let $\Lambda$ be a universe. A binary relation $R$ on $\Lambda$ is called \cite{41}:

1. Equivalence if it is transitive ($yRw$ whenever $yRz$ and $zRw$), symmetric ($wRz$ if $zRw$), and reflexive ($wRw$ for every $w \in \Lambda$).
2. Tolerance if it is both reflexive and symmetric.
3. Dominance if it is both reflexive and transitive.

Neighborhood systems have been adopted to characterize relationships between objects in database methods for the target of approximate retrieval.

**Definition 2.** Let $R$ be any binary relation on a universe $\Lambda$. Then, the $R$-neighborhoods of $z \in \Lambda$ are presented as follows:

1. Reference \cite{55} $zR = \{ y \in \Lambda : zRy \}$
2. Reference \cite{55} $Rz = \{ y \in \Lambda : yRz \}$
3. Reference \cite{54} $z \in \Lambda$ if $zR \subseteq Rz$
4. Reference \cite{47} $\langle z \rangle R = \cap \{ yR : z \in yR \}$
5. Reference \cite{47} $R \langle z \rangle = \cap \{ Ry : z \in Ry \}$
6. Reference \cite{50} $\langle z \rangle R \cap R \langle z \rangle$

**Proposition 1.** Let $R$ be any binary relation on $\Lambda$, then

1. If $R$ is a reflexive relation on $\Lambda$, then $\langle z \rangle R \subseteq zR$, $R \langle z \rangle \subseteq Rz$ and $R \langle z \rangle R \subseteq zRz$.
2. If $R$ is a symmetric relation on $\Lambda$, then $zR = Rz = zRz$ and $\langle z \rangle R = R \langle z \rangle = R \langle z \rangle R$.

By applying $zR_{\ell}$ neighborhoods, Abu-Donia \cite{42} debated three categories of upper and lower approximations of any set w.r.t finite class of binary relations $R_{\ell}$, $\ell \in \{1, 2, \ldots, n\}$.

**Definition 3.** \cite{42}. Let each $R_{\ell}$, $\ell \in \{1, 2, \ldots, n\}$ be a binary relation on a universe $\Lambda$. If $M \subseteq \Lambda$, then the $n$ lower and $n$ upper approximations of $M$ are given by

1. $\bigwedge_{\ell=1}^{n} R_{\ell} (M) = \{ z \in \Lambda : \cap_{\ell=1}^{n} R_{\ell} (M) \}$
2. $\bigvee_{\ell=1}^{n} R_{\ell} (M) = \{ z \in \Lambda : (\cap_{\ell=1}^{n} R_{\ell} (M)) \cap M \neq \emptyset \}$

Abu-Donia \cite{43} employed the neighborhood $\langle z \rangle R$ to characterize other distinct approximations of any set w.r.t reflexive, tolerance, dominance, and equivalence relations. Furthermore, he deduced the KRA (knowledge based on the rough approximation) approach, which generalized previous RS methods.

**Definition 4.** \cite{43}. Suppose each $R_{\ell}$, $\ell \in \{1, 2, \ldots, n\}$ is a binary relation on a universe $\Lambda$. Then, the $n$ lower and $n$ upper approximations of a subset $M$ of $\Lambda$ are given by

1. $\bigwedge_{\ell=1}^{n} \text{appr} (M) = \{ z \in \Lambda : \cap_{\ell=1}^{n} \langle z \rangle R_{\ell} \subseteq M \}$
2. $\bigvee_{\ell=1}^{n} \text{appr} (M) = \{ z \in \Lambda : (\cap_{\ell=1}^{n} \langle z \rangle R_{\ell}) \cap M \neq \emptyset \}$

3. Rough Set Approximation Models Based on Containment Neighborhoods (CRSA)

**Definition 5.** Suppose $\{ R_{\ell} : \ell = 1, 2, \ldots, n \}$ is a finite class of binary relations on $\Lambda$ and $z \in \Lambda$. For each $\ell$, the $C$-neighborhoods of $z \in \Lambda$ are offered as follows:

1. $zC_{\ell} = \{ y \in \Lambda : yR_{\ell} \subseteq zR_{\ell} \}$
2. $C_{\ell}z = \{ y \in \Lambda : R_{\ell}y \subseteq R_{\ell}z \}$
3. $zC_{\ell}z = zC_{\ell} \cap C_{\ell}z$
4. $\langle z \rangle C_{\ell} = \{ y \in \Lambda : \langle y \rangle R_{\ell} \subseteq \langle z \rangle R_{\ell} \}$
5. $C_{\ell} \langle z \rangle = \{ y \in \Lambda : R_{\ell} \langle y \rangle \subseteq R_{\ell} \langle z \rangle \}$
6. $C_{\ell} \langle z \rangle C_{\ell} = \langle z \rangle C_{\ell} \cap C_{\ell} \langle z \rangle$
Proposition 2. Suppose \( \mathcal{R}_\ell: \ell = 1, 2, \ldots, n \) is a finite class of binary relations on \( \Lambda \). If \( z \in \Lambda \), then

1. For each \( \ell \), \( z \in zC_\ell \), \( z \in zC_\ell \), and \( z \in zC_\ell \)
2. For each \( \ell \), \( z \in zC_\ell \), \( z \in zC_\ell \), and \( z \in zC_\ell \)

Proof. Direct to prove. \( \square \)

Lemma 1. Suppose \( \mathcal{R}_\ell: \ell = 1, 2, \ldots, n \) is a finite class of binary relations on \( \Lambda \) and \( w, z \in \Lambda \). For each \( \ell \), then the following results hold:

1. \( w \in zC_\ell \iff wC_\ell \subseteq zC_\ell \)
2. \( w \in zC_\ell \iff wC_\ell \subseteq zC_\ell \)
3. \( w \in zC_\ell \iff wC_\ell \subseteq zC_\ell \)
4. \( w \in z(z) \iff w \subseteq z(z) \)
5. \( w \in z(z) \iff w \subseteq z(z) \)
6. \( w \in z(z) \iff w \subseteq z(z) \)

Proof. We shall prove (1) and the else outcomes are similar.

Let \( \ell \in \{1, 2, \ldots, n\} \) and \( y \in wC_\ell \). Then, \( yR_\ell \subseteq wR_\ell \). Since \( w \in zC_\ell \), then \( wR_\ell \subseteq wR_\ell \). Hence, \( yR_\ell \subseteq zR_\ell \) and so \( y \in zC_\ell \), i.e., \( wC_\ell \subseteq zC_\ell \) for each \( \ell \). The other side, in view of Proposition 2, \( w \in wC_\ell \) for each \( \ell \). Since \( wC_\ell \subseteq zC_\ell \), then \( w \in zC_\ell \). \( \square \)

Corollary 1

(1) \( w \in \cap_{\ell=1}^n (zC_\ell) \iff \cap_{\ell=1}^n (wC_\ell) \subseteq \cap_{\ell=1}^n (zC_\ell) \)
(2) \( w \in \cap_{\ell=1}^n (zC_\ell) \iff \cap_{\ell=1}^n (wC_\ell) \subseteq \cap_{\ell=1}^n (zC_\ell) \)
(3) \( w \in \cap_{\ell=1}^n (zC_\ell) \iff \cap_{\ell=1}^n (wC_\ell) \subseteq \cap_{\ell=1}^n (zC_\ell) \)
(4) \( w \in \cap_{\ell=1}^n (zC_\ell) \iff \cap_{\ell=1}^n (wC_\ell) \subseteq \cap_{\ell=1}^n (zC_\ell) \)
(5) \( w \in \cap_{\ell=1}^n (zC_\ell) \iff \cap_{\ell=1}^n (wC_\ell) \subseteq \cap_{\ell=1}^n (zC_\ell) \)
(6) \( w \in \cap_{\ell=1}^n (zC_\ell) \iff \cap_{\ell=1}^n (wC_\ell) \subseteq \cap_{\ell=1}^n (zC_\ell) \)

Proposition 3. Suppose \( \mathcal{R}_\ell: \ell = 1, 2, \ldots, n \) is a finite class of reflexive relations on \( \Lambda \) and \( z \in \Lambda \). Then, for each \( \ell \), the following results hold:

1. \( zC_\ell \subseteq z\mathcal{R}_\ell \)
2. \( zC_\ell \subseteq z\mathcal{R}_\ell \)
3. \( zC_\ell \subseteq z\mathcal{R}_\ell \)
4. \( zC_\ell \subseteq z\mathcal{R}_\ell \)
5. \( zC_\ell \subseteq z\mathcal{R}_\ell \)
6. \( zC_\ell \subseteq z\mathcal{R}_\ell \)

Proof. We prove (1). Let \( w \in zC_\ell \), for every \( \ell \in \{1, 2, \ldots, n\} \). Hence, \( w\mathcal{R}_\ell \subseteq z\mathcal{R}_\ell \), for any \( \ell \in \{1, 2, \ldots, n\} \). Since every \( \mathcal{R}_\ell \in \{1, 2, \ldots, n\} \) is a reflexive relation, then \( w \in w\mathcal{R}_\ell \), so that \( w \in z\mathcal{R}_\ell \), i.e., \( zC_\ell \subseteq z\mathcal{R}_\ell \). Analogously, we can prove other consequences. \( \square \)

Corollary 2. Consider \( \mathcal{R}_\ell: \ell = 1, 2, \ldots, n \) is a finite collection of reflexive relations on a universe \( \Lambda \) and \( z \in \Lambda \). Then, the following results hold:

1. \( \cap_{\ell=1}^n zC_\ell \subseteq \cap_{\ell=1}^n z\mathcal{R}_\ell \)
2. \( \cap_{\ell=1}^n zC_\ell \subseteq \cap_{\ell=1}^n z\mathcal{R}_\ell \)
3. \( \cap_{\ell=1}^n zC_\ell \subseteq \cap_{\ell=1}^n z\mathcal{R}_\ell \)
4. \( \cap_{\ell=1}^n zC_\ell \subseteq \cap_{\ell=1}^n z\mathcal{R}_\ell \)
5. \( \cap_{\ell=1}^n zC_\ell \subseteq \cap_{\ell=1}^n z\mathcal{R}_\ell \)
6. \( \cap_{\ell=1}^n zC_\ell \subseteq \cap_{\ell=1}^n z\mathcal{R}_\ell \)

Example 1. If \( \Lambda = \{\beta_1, \beta_2, \beta_3, \beta_4\} \) and \( \Delta \) is an identity relation on \( \Lambda \), suppose \( \mathcal{R}_1, \mathcal{R}_2 \) are two reflexive relations on \( \Lambda \) defined as \( \mathcal{R}_1 = \Delta \cup \{\beta_1, \beta_2, \beta_3, \beta_4\} \) and \( \mathcal{R}_2 = \Delta \cup \{\beta_1, \beta_2, \beta_3, \beta_4\} \). Then, \( \beta_1 \mathcal{R}_1 = \{\beta_1, \beta_2\}, \beta_2 \mathcal{R}_1 = \{\beta_1, \beta_2\}, \beta_3 \mathcal{R}_1 = \{\beta_1, \beta_3\}, \beta_4 \mathcal{R}_1 = \{\beta_1, \beta_4\}, \beta_1 \mathcal{R}_2 = \{\beta_1, \beta_2\}, \beta_2 \mathcal{R}_2 = \{\beta_1, \beta_2\}, \beta_3 \mathcal{R}_2 = \{\beta_3, \beta_4\}, \beta_4 \mathcal{R}_2 = \{\beta_3, \beta_4\} \).
Remark 1. With reflexive relations $R_1, R_2$ on $\Lambda$, we get

1. Generally, the opposite of Proposition 3 may not be valid as we consider example 1.

2. Generally, the opposite of Corollary 2 may not be valid as we consider example 1, where $\beta_1 R_1 = \beta_2 R_1$, $\beta_1 R_2 = \beta_2 R_2$, and $\beta_1 C_1 = \beta_2 C_1$, and $\beta_1 C_2 = \beta_2 C_2 = \{\beta_1, \beta_2\}$.

3. If $\ell \in \{1, 2\}$, then the concepts of $z C_\ell, \langle z \rangle C_\ell$ (resp. $C_\ell z, C_\ell \langle z \rangle C_\ell$) are incomparable. See example 1, although $\langle z \rangle R_\ell \subseteq z R_\ell, z R_\ell \subseteq \langle z \rangle R_\ell$, and $z R_\ell \subseteq z R_\ell$, as it is shown in Proposition 1.

Proposition 4. Assume that $[R_\ell : \ell = 1, 2, \ldots, n]$ is a finite collection of transitive relations on a universe $\Lambda$ and $z \in \Lambda$. Then, for each $\ell$, the following results hold:

1. $z R_\ell \subseteq z C_\ell$
2. $R_\ell z \subseteq C_\ell z$
3. $z R_\ell z \subseteq C_\ell z$
4. $\langle z \rangle R_\ell \subseteq \langle z \rangle C_\ell$
5. $R_\ell \langle z \rangle \subseteq \langle z \rangle C_\ell$
6. $\langle z \rangle R_\ell \subseteq \langle z \rangle C_\ell$

Proof. We prove (1) and (2) and the rest of the proof is similar. Let $\ell \in \{1, 2, \ldots, n\}$. Suppose $w \in z R_\ell$, then $z R_\ell w$. We shall prove $w \in z C_\ell$, i.e., $w R_\ell z \subseteq z R_\ell$. Let $y \in w R_\ell$. Then $w R_\ell y$. Since each $R_\ell$ is a transitive relation on $\Lambda$, then $y \in z R_\ell$ and so $w R_\ell \subseteq z R_\ell$, i.e., $w \in z C_\ell$. Consequently, $z R_\ell \subseteq z C_\ell$.

Corollary 3. Assume that $[R_\ell : \ell = 1, 2, \ldots, n]$ is a finite collection of transitive relations on a universe $\Lambda$ and $z \in \Lambda$. Then, the following results hold:

1. $\cap _{\ell \in \Lambda} R_\ell \subseteq \cap _{\ell \in \Lambda} z C_\ell$
2. $\cap _{\ell \in \Lambda} R_\ell z \subseteq \cap _{\ell \in \Lambda} C_\ell z$
3. $\cap _{\ell \in \Lambda} z R_\ell \subseteq \cap _{\ell \in \Lambda} \langle z \rangle C_\ell$

Example 2. Let $\Lambda = \{\beta_1, \beta_2, \beta_3, \beta_4\}$. Suppose $R_1, R_2$ are two transitive relations on $\Lambda$ defined as $R_1 = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ and $R_2 = \{\beta_1, \beta_2, \beta_3, \beta_4\}$. Hence,

$\beta_1 R_1 = \beta_1 C_1 = \beta_2 C_1 = \beta_3 C_1 = \beta_4 C_1 = \beta_1, \beta_2 C_1 = \beta_1, \beta_3 C_1 = \beta_1, \beta_4 C_1 = \beta_1, \beta_1 C_2 = \beta_1, \beta_2 C_2 = \beta_1, \beta_3 C_2 = \beta_2, \beta_4 C_2 = \beta_4, \beta_1 \subseteq \beta_2 \subseteq \beta_3 \subseteq \beta_4$. See example 1, although $\langle z \rangle R_\ell \subseteq z R_\ell, z R_\ell \subseteq \langle z \rangle R_\ell, z R_\ell \subseteq \langle z \rangle R_\ell$, as it is shown in Proposition 1.
\[ C_2(\beta_1)C_2 = [\beta_1, \beta_2] \text{, } C_2(\beta_2)C_2 = [\beta_2, \beta_1] \text{, } C_2(\beta_3)C_2 = [\beta_3, \beta_1] \]

**Remark 2.** With transitive relations \( R_1, R_2 \) on \( \Lambda \) (see example 2), we have

1. If \( \ell \in \{1, 2\} \), then the concepts of \( zR_\ell (z) \) (resp. \( R_\ell z, R_\ell (z)R_\ell \)) are incomparable.
2. If \( \ell \in \{1, 2\} \), then the concepts of \( \{zC_\ell \} \) are incomparable.
3. Mainly, the converse of Proposition 4 and Corollary 3 cannot be valid as we see in example 2.

In view of Propositions 3 and 4, we have

**Proposition 5.** Suppose that \( \{ R_\ell : \ell = 1, 2, \ldots, n \} \) is a finite class of dominance relations on \( \Lambda \) and \( z \in \Lambda \). Then, for each \( \ell \), the following results hold:

1. \( zR_\ell = zC_\ell \)
2. \( R_\ell z = C_\ell z \)
3. \( zR_\ell z = zC_\ell z \)
4. \( \{zC_\ell \} \) are incomparable.
5. \( \{zC_\ell \} \) are incomparable.
6. \( \{zC_\ell \} \) are incomparable.

**Corollary 4.** Assume that \( \{ R_\ell : \ell = 1, 2, \ldots, n \} \) is a finite class of dominance relations on \( \Lambda \) and \( z \in \Lambda \). Then,

1. \( \cap_n \ell zR_\ell = \cap_n \ell zC_\ell \)
2. \( \cap_n \ell zR_\ell z = \cap_n \ell zC_\ell z \)
3. \( \cap_n \ell zR_\ell z = \cap_n \ell zC_\ell z \)
4. \( \cap_n \ell (z)R_\ell = \cap_n \ell (z)C_\ell \)
5. \( \cap_n \ell zR_\ell (z) = \cap_n \ell zC_\ell (z) \)
6. \( \cap_n \ell zR_\ell (z) = \cap_n \ell zC_\ell (z) \)

**Proposition 6.** Suppose that \( \{ R_\ell : \ell = 1, 2, \ldots, n \} \) is a finite class of dominance relations on \( \Lambda \) and \( x, y \in \Lambda \). Then, for each \( \ell \), \( y \in xR_\ell x \) iff \( yR_\ell y = xR_\ell y \).

**Proof.** (\( \implies \)) Let \( y \in xR_\ell x \) and \( z \in yR_\ell y \), then \( y \in xR_\ell z \) and \( z \in yR_\ell y \), i.e., \( xR_\ell y \), \( yR_\ell x \), \( yR_\ell z \), and \( zR_\ell z \). Since each relation is transitive, then \( xR_\ell z \) and \( zR_\ell z \). Hence, \( yR_\ell y \subseteq xR_\ell y \). Also, if \( y \in xR_\ell x \), then \( y \in xR_\ell y \). Consequently, \( y \in xR_\ell y \) for every \( y \in xR_\ell x \).

(\( \implies \)) Suppose \( yR_\ell y = xR_\ell y \). Since each relation is reflexive, then \( y \in xR_\ell x \).

**Remark 3.** Suppose \( \Lambda \) is a universe. If \( \{ R_\ell : \ell = 1, 2, \ldots, n \} \) represents class of reflexive (resp. transitive) relations, then Proposition 5 will not be correct, as shown in Examples 1 and 2.

**Proposition 7.** If \( \{ R_\ell : \ell = 1, 2, \ldots, n \} \) is a finite collection of symmetric relations on \( \Lambda \) and \( z \in \Lambda \), then \( zC_\ell = C_\ell z = \cap_n \ell zC_\ell \) for all \( \ell \).

**Proof.** Straightforward.

Depending on containment neighborhoods generated from the collection of finite binary relations, we propose in the next part some new sorts of lower and upper approximations.

**Definition 6.** Suppose that every \( R_\ell, \ell \in \{1, 2, \ldots, n\} \) is a relation on \( \Lambda \) and \( z \in \Lambda \). Based on containment neighborhoods \( zC_\ell z = C_\ell z \cap C_\ell z \), then for all \( \ell \), \( (n)_E aL(M), (n)_E aU(M) \) stands for lower and upper approximations of a set \( M \), respectively, defined as the following:

1. The \( n_E a \) lower approx. operator of \( M \) is
   \[ (n)_E aL(M) = \{z \in \Lambda : \cap_n \ell zC_\ell z \subseteq M\}. \]
2. The \( n_E a \) upper approx. operator of \( M \) is
   \[ (n)_E aU(M) = \{z \in \Lambda : \cap_n \ell zC_\ell z \cap M \neq \emptyset\}. \]
3. The \( n_E a \) boundary of \( M \) is
   \[ n_E aB(M) = (n)_E aU(M) - (n)_E aL(M). \]
4. The \( n_E a \) accuracy measure of any set \( M \) is
   \[ (n)_E aE(a)(M) = \frac{(n)_E aL(M)}{(n)_E aU(M)}. \]

where the cardinality of any set is denoted by the symbol [.]..
Example 3. Consider $R_1, R_2$ are two binary relations on $\Lambda = \{\beta_1, \beta_2, \beta_3, \beta_4\}$ such that

$R_1 = \{ (\beta_1, \beta_1), (\beta_2, \beta_2), (\beta_2, \beta_3), (\beta_3, \beta_3), (\beta_3, \beta_4), (\beta_4, \beta_4) \}$

$R_2 = \{ (\beta_1, \beta_2), (\beta_1, \beta_3), (\beta_2, \beta_2), (\beta_2, \beta_3), (\beta_2, \beta_4), (\beta_3, \beta_3), (\beta_3, \beta_4), (\beta_4, \beta_4) \}$

If $M = \{\beta_2\}$, then $(\cap_{\alpha \in \varnothing} E_{\cap_{\alpha \in \varnothing}} L(M)) \subseteq M$, but $M \not\subseteq \{\beta_2\}$. Therefore, $M \not\subseteq \{\beta_2\}$. Also, $\emptyset \not\subseteq \{\beta_2\}$.

Proposition 9. If each $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$ is a dominance relation on $\Lambda$ and $M \subseteq \Lambda$, then the coming statements hold:

1. $M \subseteq (\cap_{\alpha \in \varnothing} E_{\cap_{\alpha \in \varnothing}} L(M))$
2. $M \not\subseteq (\cap_{\alpha \in \varnothing} E_{\cap_{\alpha \in \varnothing}} L(M))$
3. $(\cap_{\alpha \in \varnothing} E_{\cap_{\alpha \in \varnothing}} L(M)) \subseteq M$
4. $(\cap_{\alpha \in \varnothing} E_{\cap_{\alpha \in \varnothing}} L(M)) \not\subseteq M$

Proof. Let $\alpha \in M$. We shall prove that $\cap_{\alpha \in \alpha} E_{\cap_{\alpha \in \alpha}} L(M)$, i.e., $\cap_{\alpha \in \alpha} (yC_{\alpha} y) \in M \not\subseteq \emptyset$. The result follows by a similar argument to the previous one.
(3) Suppose $\mathcal{R}_1, \mathcal{R}_2$ are two tolerance relations on $\Lambda$ defined as

$$\mathcal{R}_1 = \Delta \cup \{(\beta_1, \beta_3), (\beta_3, \beta_1), (\beta_2, \beta_1), (\beta_1, \beta_2), (\beta_3, \beta_2), (\beta_2, \beta_3), (\beta_1, \beta_3), (\beta_3, \beta_1), (\beta_1, \beta_1), (\beta_2, \beta_2), (\beta_3, \beta_3)\}$$

and

$$\mathcal{R}_2 = \Delta \cup \{(\beta_1, \beta_2), (\beta_2, \beta_1), (\beta_1, \beta_3), (\beta_3, \beta_1), (\beta_2, \beta_3), (\beta_3, \beta_2), (\beta_1, \beta_2), (\beta_2, \beta_1), (\beta_3, \beta_3), (\beta_1, \beta_1), (\beta_2, \beta_2), (\beta_3, \beta_3)\}$$

Remark 5. If every $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$, is a transitive relation on $\Lambda$, then $\cap_{\ell=1}^n \mathcal{R}_\ell$ is a transitive relation on $\Lambda$.

Theorem 1. Let $\Lambda$ be a universe and $z \in \Lambda$, then the coming statements hold:

1. If every $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$, is a reflexive relation on $\Lambda$, then the topology $\cap_{\ell=1}^n \mathcal{R}_\ell$ is finer than the topology $\cap_{\ell=1}^n \mathcal{R}_\ell \cap \mathcal{R}_z$.

2. If every $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$, is a transitive relation on $\Lambda$, then the topology $\cap_{\ell=1}^n \mathcal{R}_\ell \cap \mathcal{R}_z$ is finer than the topology $\cap_{\ell=1}^n \mathcal{R}_\ell$.

3. If every $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$, is a dominance relation on $\Lambda$, then the topology $\cap_{\ell=1}^n \mathcal{R}_\ell \cap \mathcal{R}_z$ and the topology $\cap_{\ell=1}^n \mathcal{R}_\ell \cap \mathcal{R}_z$ coincide.

Example 5. If $\Delta$ is an identity relation defined on $\Lambda = \{\beta_1, \beta_2, \beta_3, \beta_4\}$, then we discuss the following cases:

Suppose $\mathcal{R}_1, \mathcal{R}_2$ are two reflexive relations on $\Lambda$ defined as $\mathcal{R}_1 = \{(\beta_1, \beta_1), (\beta_2, \beta_2), (\beta_3, \beta_3), (\beta_4, \beta_4)\}$ and $\mathcal{R}_2 = \{(\beta_1, \beta_2), (\beta_2, \beta_3), (\beta_3, \beta_1), (\beta_1, \beta_3), (\beta_3, \beta_2), (\beta_2, \beta_1), (\beta_2, \beta_2), (\beta_3, \beta_3), (\beta_1, \beta_1), (\beta_2, \beta_2), (\beta_3, \beta_3)\}$

Remark 6. According to Example 5,

1. The converse of 1 and 2 of Theorem 1 may not be true in general, i.e.,
   (i) If each $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$, is reflexive, $\cap_{\ell=1}^n \mathcal{R}_\ell$ is reflexive.
   (ii) If each $\mathcal{R}_\ell$, $\ell \in \{1, 2, \ldots, n\}$, is transitive, then $\cap_{\ell=1}^n \mathcal{R}_\ell$ is transitive.

2. $\cap_{\ell=1}^n \mathcal{R}_\ell$ are not comparable.

4. Proposed Feature Selection Method

In this section, the proposed FS method is based on HBO and CRSn. However, the basic steps of the HBO are introduced in Section 4.1. Then, we discuss the developed method.

4.1. Honey Badger Optimization Algorithm. In this section, the mathematical notation of honey badger optimization algorithm is introduced. In general, HBO emulates the behaviour of honey badger to catch its prey. This process is performed through a set of stages named Digging and Honey. In the Digging stage, the prey is determined based on the smelling of honey badger, whereas in the Honey stage, the honey badger follows the honey bee for determining the beehive.

The steps of HBO begin by setting the initial of value of agents using the following:

$$Z_i = LB_i + r_i \ast (UB_i - LB_i), \quad i = 1, 2, \ldots, N$$

where $LB$ is the lower boundary and $UB$ is the upper boundary of the search space. $r_i \in [0, 1]$ refers to a random
number. Followed by [57], the exploration (Digging) and exploitation (Honey) are balanced using the density factor (α) that is defined as

$$\alpha = C \times \exp\left(-\frac{t}{T}\right),$$

where $C > 1$ stands for a constant value, $T$ represents the total number of iterations, and $t$ indicates the current iteration.

The next step is to update the solutions using the operators of Digging stage. This is performed based on the cardioid movements formulated as

$$Z_{\text{new}} = Z_0 + F \times B \times Z_0 + F \times \alpha \times d^T \times r_3 \times \cos(2\pi r_4) \times [1 - \cos(2\pi r_5)].$$

(7)

In equation (7), $Z_{\text{new}}$ stands for the new value of $Z_i$, $Z_0$ represents the best solution found so far, and $r_3, r_4, r_5$, and $r_6$ are random numbers. ($B$) is a constant number. ($F$) is a parameter used to control the search direction and it is the value determined using the following equation:

$$F = \begin{cases} 1, & \text{if} \ (r_6 \leq 0.5), \\ -1, & \text{else}. \end{cases}$$

(8)

$I$ stands for the smell intensity of the prey ($x_b$) and it is used to represent the distance between the $x_b$ and $x_i$. It is formulated as

$$I_i = r_2 \times \frac{S}{4\pi d_i^2},$$

(9)

$$S = (Z_i - Z_i^{(t+1)})^2, d_i = Z_0 - Z_i.$$  

(10)

Meanwhile, the solutions can be updated using operators of Honey stage. This process is achieved using the following formula:

$$Z_{\text{new}} = Z_0 + F \times \gamma \times \alpha \times d_i,$$

(11)

where $r_\gamma$ is a random number. The steps of HBO are given in Algorithm 1.

4.2. Proposed HBOCRSA Framework. We present the steps of the developed feature selection based on modified HBO and combined it with CRSA approximations in this section. Figure 1 depicts the enhanced method structure. The main aim of the binary HBO (BHBO) algorithm based on CRSA (HBOCRSA) is identifying the subset of features that are more relevant regarding the dependencies degree among the features with the target features. As well as to deal with the discrete FS problem, binary HBO is applied.

The initialization of HBOCRSA is presented in identifying the features number ($F_k$, $k = 1, 2, \ldots, d$) of the dataset corresponding to each solution dimension. Then, such dataset is split into training and testing sets. As well as, through the interval [0, 1], the generation of the initial value for $N$ agents $Z$ is performed. Furthermore, the computation of the value of the objective function of $Z_i$, $i = 1, 2, \ldots, N$ with determining the finest among them is performed. Through utilizing the HBO operators, modernize the agents. The modernization process is repeated until the stopping condition is met. Each phase is discussed in detail as follows.

4.2.1. First Stage: Generating Solutions. The dataset is split into training and testing sets, by the percentages 80% and 20%, respectively, through the enhanced FS method. The $N$ agents $Z$ are initialized by the following equation:

$$Z_{i,j} = LB_j + r_1 \times (UB_j - LB_j), \quad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, D.$$  

(12)

In equation (12), $UB_j$ and $LB_j$ exemplify the dimension $j$ upper and lower boundaries.

4.2.2. Second Stage: Updating Solutions. The proposed technique is initialized through computing, for each agent $Z_i$, the objective function value for two iterations.

The main goal of the first iteration is transferring $Z_i$ into binary, as shown in

$$BZ_{i,j} = \begin{cases} 1, & \text{if} \ Z_{i,j} > 0.5, \\ 0, & \text{otherwise}. \end{cases}$$

(13)

Such process is used for transferring the real-valued agents to discrete ones which can deal with making them the feature selection problem.

The second step in the developed method contains choosing the features that are more relevant regarding ones in $BZ$ and at the same time deleting the irrelevant features that correspond to the zeros. Furthermore, the quality first features are evaluated based on the function (Fit$_i$) given as in equation (14). This is considered as the optimization problem which aims to minimize the error classification and minimize the number of features.

$$\text{Fit}(Z_i) = \eta \times \frac{|Z_i|}{d} + \mathcal{R} \times (1 - \gamma_c(D)), \quad \mathcal{R} + \eta = 1,$$

(14)

where $|Z_i|$ exemplifies the features number, as in equation (14), which is chosen by utilizing the $Z_i$ current value. In addition, $d$ exemplifies the number of features inside the dataset. $\mathcal{R}$ and $\eta$ are the coefficients that have the ability for balancing the number of selected features with the dependency degree $\gamma_c(D)$ given by

$$\gamma_c(D) = \frac{|\text{POS}_c(D)|}{|U|}.$$  

(15)

In equation (15), $\text{POS}_c(D)$ represents the positive region defined as

$$\text{POS}_c(D) = \bigcup B(Z), Z \in U / D,$$

(16)

where $B(Z)$ exemplifies the features approximating power.
Finally, the third step aims for identifying the finest agent \( Z_b \), after that the agents are modernized by utilizing the HBO operators given in Section 4.1.

4.2.3. Third Stage: Stop Conditions. Meeting the stopping condition for the proposed technique can be tested and returning the finest solution as the solutions accepted with repeating the strides existed in the second stage. Moreover, the testing set is used to assess the relevant features existed in the finest solution \( Z_b \) and perform classification for the reduced set by utilizing the KNN classifier for the features quality assessing.

5. Experimental Results and Discussion Using Real-World Datasets

The performance of the developed method can be justified through the modified CRSA which is considered as RS extension with concluding some of the experiments. Through our work and experiments, we depended on using ten datasets, in addition to comparing the results with other FS approaches. In this experiment, the proposed BHBA is used to determine the relevant feature from the given data. To conduct this, the CRSA is used as a part of the fitness function that is defined in equation (14).

5.1. Datasets Description. The quality of HBOCRSA is validated by utilizing ten datasets with varying dimensionality. The datasets, gathered from various fields, are taken online from UCI [58], and those ones that we depended in our work are shown in Table 1 where they have various instance numbers, feature numbers, and classes.

The efficiency of HBOCRSA can be validated and justified, where the dataset is split into 80% and 20% training and testing sets, respectively, where these percentages form the whole number. Through 30 independent times, each
algorithm has been conducted for guaranteeing the comparison quality. The comparisons are performed through using the algorithms of the salp swarm algorithm (SSA), self-adaptive differential evolution (SaDE), grey-wolf optimization (GWO), genetic algorithm (GA), and teaching-learning-based optimization (TLBO) as competing ones for feature selection. With taking the note that each algorithm parameter is set regarding its implementation, the iteration number and the population size are set to 15 and 20 which are considered the common parameters. Hereafter, each solution in the population has dimension which equals to the features number for each dataset.

5.2. Performance Measures. More metrics for the proposed technique, HBOCRSA, assessing can be utilized. Some of such measures or metrics are accuracy and the selected features number; such measures can be defined as

(i) Average accuracy \( \text{AVGAcc} \) exemplifies the accuracies average; overall, the runs number \( (N_r = 30) \) is given as

\[
\text{AVGAcc} = \frac{1}{N_r} \sum_{k=1}^{N_r} \frac{\text{Acc}_{\text{Best}}^k}{N_r} \tag{17}
\]

\[
\text{Acc}_{\text{Best}}^k = \frac{TP + TN}{TP + FN + FP + TN} \tag{18}
\]

(ii) Average number of the selected features \( \text{AVG}_{|B_{\text{t}}|_{\text{Best}}} \) calculates the features average chosen by each algorithm through all runs and it is given as

\[
\text{AVG}_{|B_{\text{t}}|_{\text{Best}}} = \frac{1}{N_r} \sum_{k=1}^{N_r} |B_{\text{t}}|_{\text{Best}}^k \tag{19}
\]

In equation (15), \( |B_{\text{t}}|_{\text{Best}}^k \) exemplifies the finest solution cardinality at \( k^{th} \) run.

5.3. Experimental Series 1: Results and Discussion Using UCI Datasets. The comparison results between the developed method and other competitive methods are given in Tables 2–5 and Figures 2 and 3. From these results, it can be noticed that the HBO using CRSAt (i.e., HBOCRSA) has high accuracy at four datasets (i.e., D2-D4 and D8). However, for the other datasets, its results are competitive with other methods, for example, at D1, D7, D9, HBOCRSA allocates the second rank. Moreover, the average of each method among the tested datasets is given in Figure 2 and it can be noticed the difference between HBOCRSA and LSHADE, TLBO, SSA, SGA, SaDE, and bGWO is 3.527%, 2.711%, 3.799%, 4.754%, 2.638%, 5.258%, respectively. In case the traditional RS is used as part of fitness value (as in Table 3), HBORS provides the better accuracy at D2, D4, D7, and D9. In addition, its average overall, the tested ten datasets are better than other methods with difference 3.780%, 3.7918%, 3.009%, 1.6491%, 3.218%, and 8.0456% when compared with LSHADE, TLBO, SSA, SGA, SaDE, and bGWO, respectively (see Figure 2).

Besides the accuracy obtained by each FS approach using either RS or CRSAt, it can be noticed that the accuracy of each method is increased using CRSAt. This indicates that CRSAt has ability to determine the relevant features better than traditional RS.

To further analysis the behaviour of HBO using RS and CRSAt, its ability to reduce the number of features is computed, as given in Table 4. From these results, it can be seen that HBO based on traditional RS provides smallest number of features at 70% from the tested datasets (i.e., all datasets, except D2, D3, and D8), followed by GA which has the smallest number at three datasets. In terms of the average of selected features overall the tested sets, it can be noticed that the HBO and GA allocate the first and second ranks, respectively; however, the difference between them is not significant. Followed by LSHADE and TLBO, whereas, the worst algorithm is SSA which has the largest number of features.

In case of using CRSAt, the number of selected features using HBO is still the smallest number nearly 16 features. This is followed by bGWO and LSHADE that allocate the second and third ranks, respectively. By comparing the performance of each algorithm in terms of number of selected features using traditional RS and CRSAt, it can be observed that RS provides number of features larger than those CRSAt. For example, HBO, LSHADE, TLBO SSA, SGA, SaDE, and bGWO based on RS provide nearly 6.07, 4.18, 12.86, 3.71, −3.344, 1.22, and 18.14 features, respectively, greater than using CRSAt.

### Table 1: Description of datasets.

| Datasets          | No. of instances | No. of features | No. of classes | Data category |
|-------------------|------------------|----------------|----------------|---------------|
| Breastcancer (D1) | 699              | 9              | 2              | Biology       |
| BreastEW (D2)     | 569              | 30             | 2              | Biology       |
| CongressEW (D3)   | 435              | 16             | 2              | Politics      |
| Exactly2 (D4)     | 1000             | 13             | 2              | Biology       |
| HeartEW (D5)      | 270              | 13             | 2              | Biology       |
| IonosphereEW (D6) | 351              | 34             | 2              | Biology       |
| Lymphography (D7) | 148              | 18             | 2              | Electromagnetic |
| PenglungEW (D8)   | 73               | 325            | 2              | Biology       |
| SonarEW (D9)      | 208              | 60             | 2              | Biology       |
| SpectEW (D10)     | 267              | 22             | 2              | Biology       |

Bold values represent the best value.
Figures 4 and 5 show an example of the average of convergence curve for FS algorithms using RS and CRSA, respectively, as fitness function. From these figures, it can be concluded that BHBA has high ability to converge faster than other methods in both cases RS and CRSA. In addition, by comparing the behaviour of the algorithm(s) using the CRSA, it can be noticed that they have ability better than using RS to minimize the fitness value.

To make a decision if the difference between the developed HBO using CRSA and other methods is significant or not, a nonparametric Friedman test is applied. The statistical values obtained using FR test are given in Table 5. From these values, it can be observed that the developed BHBO has the largest mean rank in terms of accuracy using both the traditional RS and new CRSA. In addition, the smallest mean rank in terms of number of selected features is

| Table 2: Accuracy of each algorithm using CRSA. |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| BHBO | LSHADE | TLBO | SSA | SGA | SaDE | bGWO |
| D1 | 0.9705 | 0.9429 | 0.9571 | 0.9643 | **0.9714** | 0.9571 | 0.8929 |
| D2 | **0.9802** | 0.9298 | 0.9561 | 0.9474 | 0.9386 | 0.9649 | 0.9211 |
| D3 | **0.9942** | 0.9655 | 0.9333 | 0.9195 | 0.9425 | 0.9195 | 0.9655 |
| D4 | **0.9041** | 0.6750 | 0.7500 | 0.6400 | 0.6850 | 0.6250 | 0.7500 |
| D5 | 0.7465 | 0.7593 | 0.7778 | 0.7963 | 0.7963 | **0.8333** | 0.7778 |
| D6 | 0.9294 | 0.8592 | 0.8592 | 0.8592 | 0.8592 | 0.8451 | 0.8592 |
| D7 | 0.7803 | 0.7333 | **0.8333** | 0.7667 | 0.7333 | 0.8000 | 0.9467 |
| D8 | **1.0000** | 0.9333 | 0.8000 | 0.9333 | 0.8667 | 0.9467 | 0.9333 |
| D9 | 0.8786 | **0.8810** | 0.8571 | **0.8810** | 0.8333 | **0.8810** | 0.7619 |
| D10 | 0.8376 | 0.8889 | **0.9259** | 0.8333 | 0.8333 | 0.8704 | 0.7963 |

Bold values represent the best value.

| Table 3: Accuracy of each algorithm using rough set. |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| BHBO | LSHADE | TLBO | SSA | SGA | SaDE | bGWO |
| D1 | 0.9558 | 0.9386 | 0.9214 | 0.9600 | **0.9600** | 0.9429 | 0.9514 |
| D2 | **0.9420** | 0.9211 | 0.9158 | 0.9158 | 0.9211 | 0.9263 | 0.9471 |
| D3 | 0.9538 | 0.9494 | 0.8966 | 0.9586 | **0.9586** | 0.9264 | 0.9111 |
| D4 | **0.7897** | 0.7180 | 0.7790 | 0.8648 | 0.8873 | 0.9211 | 0.8761 |
| D5 | 0.7198 | 0.6889 | **0.7741** | 0.6889 | 0.7593 | 0.7111 | 0.6741 |
| D6 | 0.8474 | 0.8563 | 0.8648 | 0.8873 | 0.9211 | **0.9000** | 0.7619 |
| D7 | 0.8364 | 0.7933 | 0.7791 | 0.7163 | 0.7133 | 0.7667 | 0.6554 |
| D8 | 0.8702 | 0.8200 | 0.7067 | 0.8667 | 0.8986 | 0.8761 | 0.8761 |
| D9 | **0.9060** | 0.8190 | 0.8476 | 0.8619 | 0.8952 | 0.7952 | 0.7429 |
| D10 | 0.7579 | 0.6963 | 0.7148 | 0.7185 | **0.7778** | 0.7222 | 0.7333 |

Bold values represent the best value.

| Table 4: Number of selected features using RS and CRSA. |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| BHBO | LSHADE | TLBO | SSA | GSA | SaDE | bGWO |
| CRS A | RS | CRS A | RS | CRS A | RS | CRS A | RS | CRS A | RS |
| D1 | 1.8 | 2.4 | 1.6 | 2.8 | 2.1 | 3 | 3.5 | 3.8 | **2.1** | 3 | 2.8 | 3.6 | 3.4 | 2.8 |
| D2 | **6.5** | 15.2 | 11.8 | 15.2 | 14.2 | 16 | 18.2 | 20.2 | 16.3 | **8.4** | 17.9 | 19.2 | 11.5 | 13 |
| D3 | 2.3 | 7.3 | 6.2 | 7.2 | 5.9 | 8.2 | 8.2 | 10 | 7 | **5.2** | 6.9 | 8.8 | 3.9 | 6.4 |
| D4 | 2.3 | 5.5 | 5.5 | 6.4 | 3.8 | 7 | 6.2 | 7.8 | 5 | 6 | 7.1 | 7 | 4.4 | 6.6 |
| D5 | 7 | 14 | 15.6 | 16.8 | 17.4 | 18.4 | 20.2 | 23 | 19.3 | 16.8 | 20.9 | 21 | 10.5 | 17.6 |
| D6 | 3.5 | 5.7 | 6.6 | 8 | 6.9 | 9.6 | 8.1 | 11 | 7.9 | 9.8 | 8.9 | 10.4 | 4.8 | 6.2 |
| D7 | **112.4** | 212 | 209 | 226 | 120 | 225.6 | 241 | 258.4 | 150.4 | **117.6** | 255.7 | 259 | 130.5 | 267 |
| D8 | **20.4** | 34.2 | 32.2 | 36.4 | 34.2 | 36.6 | 41.5 | 44.4 | 40.8 | 42 | 42.4 | 42.4 | 21.4 | 41.2 |
| D9 | 4.7 | **10.4** | 8.4 | 11.2 | 9.4 | 12 | 12.9 | 13.8 | 10.24 | 12.4 | 13.8 | 13.6 | 6.8 | 10.8 |

Bold values represent the best value.
Figure 2: Average of accuracy of each algorithm among the tested datasets.

Figure 3: Average of selected features of each algorithm among the tested datasets.

Figure 4: Continued.
Figure 4: Convergence curve of the algorithms using RS as fitness value.

Figure 5: Convergence curve of the algorithms using CRSA as fitness value.
achieved using BHBO. This indicates the combination between the BHBO and CRSA has been led to increase in the classification and decrease in the number of features.

It can be seen that the new HBOCRSA approach is more applicable and efficient than competing algorithms based on earlier evaluations utilizing UCI datasets. However, HBOCRSA still has significant drawbacks, such as a high computational complexity, especially when used to handle high-dimensional datasets. Using parallel processing and GPU hardware, this can be remedied. Furthermore, HBOCRSA’s behaviour in balancing exploration and exploitation to find the viable region has the greatest impact on its performance. Combining it with additional operators can improve this.

6. Conclusion and Future Works

In this article, we focused on creating a novel model of rough set approximations (RSA), namely, the rough set approximation models depending on containment neighborhoods (CRSA), that generalize the classical notions of (RSA) and derive a number of distinguished results. To evaluate the appropriateness of this model, it is applied to enhance the classification of contrasting dataset by utilizing it an objective function for distinct feature selection approach. This has been carried out by applying the binary version of honey badger optimization algorithm (BHBO) as feature selection method. The effects of HBOCRSA are compared with different MH techniques, which involve GWO, LSHADE, SSA, GA, TLBO, and SaDE. A class of ten datasets is employed to assess the performance of the developed method. The experiential results clarified the high performance of the developed method as FS approach which owns accuracy better than other methods. Furthermore, the number of the selected features acquired utilizing the HBOCRSA is smaller than the other methods. Additionally, the accomplishment of the model of (CRSA) is better than classical (RS) approach that is based on the factors of performance measures. Based on favourable consequences gained from the developed method, it can be applied in diverse scopes, for instance, cloud computing, image processing, and IoT applications. In addition, it can be reconstructed as multiobjective technique and used to several sets of reality multiobjective problems involving feature, selection, and engineering issues and others.

Data Availability

The data used to support the findings of this study are available from the authors upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This project was supported financially by the Academy of Scientific Research and Technology (ASRT), Egypt, Grant 6684.

References

[1] P. Ni, S. Zhao, X. Wang, H. Chen, C. Li, and E. C. Tsang, “Incremental feature selection based on fuzzy rough sets,” Information Sciences, vol. 536, 2020.
[2] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, “Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods,” Applied Soft Computing, vol. 86, Article ID 105836, 2020.
[3] Z. Zhang and K. K. Parhi, “M3su: Minimum mean minimum uncertainty feature selection for multiclass classification,” Journal of Signal Processing Systems, vol. 92, no. 1, pp. 9–22, 2020.
[4] G. Kaur, S. S. Gill, and M. Rattan, “Whale optimization algorithm for performance improvement of silicon-on-insulator finfets,” International Journal of Artificial Intelligence, vol. 18, pp. 63–81, 2020.
[5] R.-E. Precup, R.-C. David, R.-C. Roman, E. M. Petriu, and A.-I. Szedlak-Stinean, “Slime mould algorithm-based tuneding of cost-effective fuzzy controllers for servo systems,” International Journal of Computational Intelligence Systems, vol. 14, no. 1, pp. 1042–1052, 2021.
[6] N. Bacanin, “Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence metaheuristic,” Algorithms, vol. 13, 2020.
[7] G. Sun, J. Li, J. Dai, Z. Song, and F. Lang, “Feature selection for IoT based on maximal information coefficient,” Future Generation Computer Systems, vol. 89, pp. 606–616, 2018.
[8] R.-E. Precup, E.-L. Hedrea, R.-C. Roman, E. M. Petriu, A.-I. Szedlak-Stinean, and C.-A. Bojan-Dragos, “Experiment-based approach to teach optimization techniques,” IEEE Transactions on Education, vol. 64, pp. 88–94, 2020.
[9] L. Abualigah, M. Abd Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, “Reptile search algorithm (rsa): A nature-inspired meta-heuristic optimizer,” Expert Systems with Applications, vol. 191, Article ID 116158, 2021.
[10] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. A. Al-qaness, and A. H. Gandomi, “Aquila optimizer: A novel meta-heuristic optimization algorithm,” Computers & Industrial Engineering, vol. 157, Article ID 107250, 2021.
[11] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, “Salp swarm algorithm: A bio-inspired optimizer for engineering design problems,” Advances in Engineering Software, vol. 114, pp. 163–191, 2017.
[12] Z.-j. Teng, J.-l. Lv, and L.-w. Guo, “An improved hybrid grey wolf optimization algorithm,” Soft computing, vol. 23, no. 15, pp. 6617–6631, 2019.
[13] M. Moattari and M. H. Moradi, “Conflict monitoring optimization heuristic inspired by brain fear and conflict systems,” Int J Artif Intell, vol. 18, pp. 45–62, 2020.
[14] F. Merrikh-Bayat, “The runner-root algorithm: a meta-heuristic for solving unimodal and multimodal optimization problems inspired by runners and roots of plants in nature,” Applied Soft Computing, vol. 33, pp. 292–303, 2015.
[15] A. A. Ewees, M. A. A. Al-qaness, L. Abualigah et al., “Boosting arithmetic optimization algorithm with genetic algorithm operators for feature selection: Case study on cox proportional hazards model,” Mathematics, vol. 9, no. 18, Article ID 2321, 2021.
[16] R. A. Ibrahim, L. Abualigah, A. A. Ewees et al., “An electric fish-based arithmetic optimization algorithm for feature selection,” Entropy, vol. 23, no. 9, Article ID 1189, 2021.
[17] R. A. Ibrahim, D. Yousri, M. Abd Elaziz, S. Alshathri, and I. Attiya, “Fractional calculus-based slime mould algorithm
for feature selection using rough set," IEEE Access, vol. 9, pp. 131625–131636, 2021.

[18] S. Mirjalili, "Moth–firefly optimization algorithm: A novel nature-inspired heuristic paradigm," Knowledge-based systems, vol. 89, pp. 228–249, 2015.

[19] R. A. Ibrahim, A. A. Ewees, D. Oliva, M. Abd Elaziz, and S. Lu, "Improved salp swarm algorithm based on particle swarm optimization for feature selection," Journal of Ambient Intelligence and Humanized Computing, vol. 10, no. 8, pp. 3155–3169, 2019.

[20] B. Khotimah, Bain Khusnul, M. Miswanto, and H. Suprajitno, "Optimization of Feature Selection Using Genetic Algorithm in Naïve Bayes Classification for Incomplete Data," International Journal of Intelligent Engineering and Systems, vol. 13, no. 1, pp. 334–343, 2020.

[21] R. A. Ibrahim, M. A. Elaziz, and S. Lu, “Chaotic opposition-based grey-wolf optimization algorithm based on differential evolution and disruption operator for global optimization,” Expert Systems with Applications, vol. 108, pp. 1–27, 2018.

[22] M. Landowski and A. Landowska, “Usage of the rough set theory for generating decision rules of number of traffic vehicles,” Transportation research procedia, vol. 39, pp. 260–269, 2019.

[23] M. Zavareh and V. Maggioni, “Application of rough set theory to water quality analysis: A case study,” Data, vol. 3, no. 4, p. 50, 2018.

[24] A. K. Sinha and N. Namdev, “Feature selection and pattern recognition for different types of skin disease in human body using the rough set method,” Network Modeling Analysis in Health Informatics and Bioinformatics, vol. 9, pp. 1–11, 2020.

[25] H. A. Nabwey, “An intelligent mining model for medical diagnosis of heart disease based on rough set data analysis,” International Journal of Engineering Research and Technology, vol. 13, no. 2, pp. 355–363, 2020.

[26] R. Ceklik and A. K. Uysal, “A novel filter feature selection method using rough set for short text data,” Expert Systems with Applications, vol. 160, Article ID 113691, 2020.

[27] H. A. Nabwey, “A methodology based on rough set theory and hypergraph for the prediction of wart treatment,” International Journal of Engineering Research and Technology, vol. 13, no. 3, pp. 552–559, 2020.

[28] J. Zhao, J.-m. Liang, Z.-n. Dong, D.-y. Tang, and Z. Liu, “Accelerating information entropy-based feature selection using rough set theory with classified nested equivalence classes,” Pattern Recognition, vol. 107, Article ID 107517, 2020.

[29] M. A. Tawhid and A. M. Ibrahim, “Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm,” International Journal of Machine Learning and Cybernetics, vol. 11, no. 3, pp. 573–602, 2020.

[30] R. A. Ibrahim, M. Abd Elaziz, D. Oliva, and S. Lu, “An improved runner-root algorithm for solving feature selection problems based on rough sets and neighborhood rough sets,” Applied Soft Computing, vol. 97, Article ID 105517, 2020.

[31] D. P. Acharjya, “Knowledge inferencing using artificial bee colony and rough set for diagnosis of hepatitis disease,” International Journal of Healthcare Information Systems and Informatics (IJHISI), vol. 16, pp. 49–72, 2021.

[32] G. Jothi, “Hybrid tolerance rough set–firefly based supervised feature selection for mri brain tumor image classification,” Applied Soft Computing, vol. 46, pp. 639–651, 2016.

[33] M. A. Tawhid and A. M. Ibrahim, “Hybrid binary particle swarm optimization and flower pollination algorithm based on rough set approach for feature selection problem,” in Nature-Inspired Computation in Data Mining and Machine Learning, pp. 249–273, Springer, Berlin, Germany, 2020.

[34] S. Patra and B. Barman, “A novel dependency definition exploiting boundary samples in rough set theory for hyperspectral band selection,” Applied Soft Computing, vol. 99, p. 106944, 2021.

[35] M. Abdolrazagz-Nezhad, H. Radgohar, and S. N. Salimian, “Enhanced cultural algorithm to solve multi-objective attribute reduction based on rough set theory,” Mathematics and Computers in Simulation, vol. 170, pp. 332–350, 2020.

[36] A. T. Sahlol, M. Abd Elaziz, M. A. A. Al-Qness, and S. Kim, “Handwritten arabic optical character recognition approach based on hybrid whale optimization algorithm with neighborhood rough set,” IEEE Access, vol. 8, pp. 23011–23021, 2020.

[37] G. Jothi, H. H. Inbarani, A. T. Azar, and K. R. Devi, “Rough set theory with jaya optimization for acute lymphoblastic leukemia classification,” Neural Computing and Applications, vol. 31, no. 9, pp. 5175–5194, 2019.

[38] M. M. Mafarja and S. Mirjalili, “Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection,” Soft Computing, vol. 23, no. 15, pp. 6249–6265, 2019.

[39] G. T. Reddy, M. P. K. Reddy, K. Lakshmanana, D. S. Rajput, R. Kauri, and G. Srivastava, “Hybrid genetic algorithm and a fuzzy logic classifier for heart disease diagnosis,” Evolutionary Intelligence, vol. 13, no. 2, pp. 185–196, 2020.

[40] L. Zou, H. Li, W. Jiang, and X. Yang, “An improved fish swarm algorithm for neighborhood rough set reduction and its application,” IEEE Access, vol. 7, pp. 90277–90288, 2019.

[41] Z. a. Pawlak, "Rough sets," International Journal of Computer & Information Sciences, vol. 11, no. 5, pp. 341–356, 1982.

[42] H. M. Abu-Donia, “Comparison between different kinds of approximations by using a family of binary relations,” Knowledge-Based Systems, vol. 21, no. 8, pp. 911–919, 2008.

[43] H. M. Abu-Donia, “Multi knowledge based rough approximations and applications,” Knowledge-Based Systems, vol. 26, pp. 20–29, 2012.

[44] T. M. Al-shami, “An improvement of rough sets’ accuracy measure using containment neighborhoods with a medical application,” Information Sciences, vol. 569, pp. 110–124, 2021.

[45] T. Al-shami, H. Isik, A. Nawar, and R. Hosny, “Some topological approaches for generalized rough sets via ideals,” Mathematical Problems in Engineering, vol. 2021, p. 11, 2021.

[46] A. A. Alam, M. Y. Bakeir, and E. A. Abo-Tabl, “New approach for basic rough set concepts,” in International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, pp. 64–73, Springer, Berlin, Germany, 2005.

[47] A. Alam, M. Bakeir, and E. Abo-Tabl, “New approach for closure spaces by relations,” Acta Mathematica Academiae Paedagogicae Nyiregyhaziensis, vol. 22, pp. 285–304, 2006.

[48] A. Alam, M. Bakeir, and E. Abo-Tabl, “Some methods for generating topologies by relations,” Bulletin of the Malaysian Mathematical Sciences Society, vol. 31, pp. 35–45, 2008.

[49] R. Hosny, B. Aasaad, A. Azzam, and T. Al-shami, “Various topologies generated from neighbourhoods via ideals,” Complexity, vol. 2021, p. 11, 2021.

[50] A. Kozae, S. El-Sheikh, and M. Hosny, “On generalized rough sets and closure spaces,” International Journal of Applied Mathematics, vol. 23, pp. 997–1023, 2010.

[51] X. Wang and L. Ma, “Study on covering rough sets with topological methods,” CAAI Transactions on Intelligence Technology, vol. 4, no. 3, pp. 129–134, 2019.
[52] C. Wu, Y. Yue, M. Li, and O. Adjei, “The rough set theory and applications,” Engineering Computations, vol. 21, 2004.
[53] Y. Y. Yao, “Two views of the theory of rough sets in finite universes,” International journal of approximate reasoning, vol. 15, no. 4, pp. 291–317, 1996.
[54] Y. Y. Yao, “Relational interpretations of neighborhood operators and rough set approximation operators,” Information Sciences, vol. 111, no. 1-4, pp. 239–259, 1998.
[55] B. De Baets and E. Kerre, “A revision of bandler-kohout compositions of relations,” Mathematica Pannonica, vol. 4, pp. 59–78, 1993.
[56] R. M. Wolpert and J. G. Gimpel, “Information, Recall, and Accountability: The Electorate’s Response to the Clarence Thomas Nomination,” Legislative Studies Quarterly, vol. 22, no. 4, pp. 535–550, 1997.
[57] F. A. Hashim, E. H. Houssein, K. Hussain, M. S. Mabrouk, and W. Al-Atabany, “Honey badger algorithm: New metaheuristic algorithm for solving optimization problems,” Mathematics and Computers in Simulation, vol. 192, 2021.
[58] A. Frank, “Uci machine learning repository,” 2010, https://archive.ics.uci.edu/ml.