Computer Systems Science & Engineering
DOI: 10.32604/csse.2023.027129
Article

Tech Science Press

Cat and Mouse Optimizer with Artificial Intelligence Enabled Biomedical Data Classification

B. Kalpana[1], S. Dhanasekaran[2], T. Abirami[3], Ashit Kumar Dutta[4], Marwa Obayya[5], Jaber S. Alzahrani[6] and Manar Ahmed Hamza[7,*]

[1]Department of Information Technology, RMD Engineering College, Chennai, 601206, India
[2]Department of Information Technology, Kalasalingam Academy of Research and Education, Srivilliputtur, 626126, India
[3]Department of Information Technology, Kongu Engineering College, Erode, 638060, India
[4]Department of Computer Science and Information Systems, College of Applied Sciences, AlMaarefa University, Ad Diriyah, Riyadh, 13713, Kingdom of Saudi Arabia
[5]Department of Biomedical Engineering, College of Engineering, Princess Nourah bint Abdulrahman University, Riyadh, 11671, Saudi Arabia
[6]Department of Industrial Engineering, College of Engineering at Alqunfudah, Umm Al-Qura University, Saudi Arabia
[7]Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, AlKharj, Saudi Arabia
*Corresponding Author: Manar Ahmed Hamza. Email: ma.hamza@psau.edu.sa
Received: 11 January 2022; Accepted: 11 March 2022

Abstract: Biomedical data classification has become a hot research topic in recent years, thanks to the latest technological advancements made in healthcare. Biomedical data is usually examined by physicians for decision making process in patient treatment. Since manual diagnosis is a tedious and time consuming task, numerous automated models, using Artificial Intelligence (AI) techniques, have been presented so far. With this motivation, the current research work presents a novel Biomedical Data Classification using Cat and Mouse Based Optimizer with AI (BDC-CMBOAI) technique. The aim of the proposed BDC-CMBOAI technique is to determine the occurrence of diseases using biomedical data. Besides, the proposed BDC-CMBOAI technique involves the design of Cat and Mouse Optimizer-based Feature Selection (CMBO-FS) technique to derive a useful subset of features. In addition, Ridge Regression (RR) model is also utilized as a classifier to identify the existence of disease. The novelty of the current work is its designing of CMBO-FS model for data classification. Moreover, CMBO-FS technique is used to get rid of unwanted features and boosts the classification accuracy. The results of the experimental analysis accomplished by BDC-CMBOAI technique on benchmark medical dataset established the supremacy of the proposed technique under different evaluation measures.

Keywords: Artificial intelligence; biomedical data; feature selection; cat and mouse optimizer; ridge regression

This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Information and Communication Technology (ICT) has evolved tremendously in the recent years which made it possible to save huge volumes of information from different fields of engineering and medical applications. This information should be provided mandatorily, in terms of objects (patterns) and massive number of features so that all the aspects of the domain get characterized [1]. However, there is a complexity exits i.e., it could often result in the inclusion of several irrelevant or redundant characteristics. This may lead to poor results when utilizing Data Mining (DM) or Machine Learning (ML) methods for knowledge discovery. Dimensionality reduction [2] is a type of reduction technique in knowledge discovery procedure when handling massive set of information. There are two major methods used in this study towards reduction problems such as selection- and transformation-based methods. Selection-based method is otherwise called as feature extraction which encodes or transforms the fundamental meaning of the feature [3]. Transformation-based method reduces the original feature space without conversion owing to which the original feature is maintained. Here, cogent interpretation is feasible, and is commonly called as Feature Selection (FS) [4]. When conducting a survey, FS models can be categorized as two types in which the former FS method returns a set of features whereas in latter, the FS method returns the ranking order of each feature (according to the importance of the feature) [5]. Likewise, based on the relationships of FS method with learning models (classification), FS method is generally categorized in addition to other two distinct methods such as classification-dependent (embedded and wrapper), and classification-independent (filter). The last few years have experienced the predominant application of numerous FS models in the fields of E-commerce, medical, and healthcare which are designed by different mainstream study communities on different metrics such as probability, data theory, correlation, and so on [6]. Fig. 1 illustrates the processes involved in biomedical data classification.

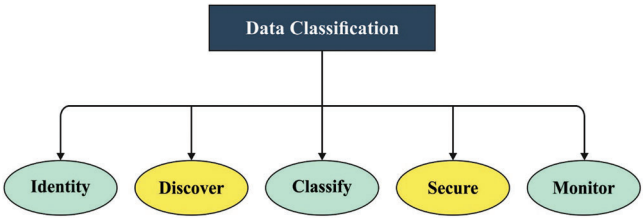

Figure 1: Biomedical data classification process

Traditionally, FS method includes four fundamental stages such as subset evaluation, subset generation, result validation, and stopping criterion [7]. Amongst the four phases, subset evaluation and subset generation are the two leading factors in the creation of FS algorithm. Subset generation is assumed to be a searching issue that focuses on identifying the optimal subset from each feasible feature subset. Various searching approaches like heuristic search, greedy search, and exhaustive search are mainly utilized and explored by the authors to recognize the optimal or sub-optimal feature subset [8]. On the other hand, exhaustive approach devises every feasible combination of the feature and its 'combinatory explosion' results in computation load that exponentially gets improved with several features. Heuristic search strategy utilizes metaheuristic approaches like Grey Wolf Optimizer (GWO), Simulated Annealing (SA), Genetic Algorithm (GA), and many other optimization methods to resolve FS issues [9]. In spite of the developments achieved in the abovementioned search methods, it remains unfeasible since the methods are computationally impractical or the methods achieved a solution apart from the optimal one [10]. The advanced methods report that wrapper or filter FS methods can be applied, when the consequence of hybrid FS is not considered. Due to the established fact that no single method can assure optimum

outcomes regarding prediction efficiency, the efficacy of the 'hybrid' method using the advantages of filter and wrapper models is examined mainly in the study.

The current research work presents a novel Biomedical Data Classification using Cat and Mouse Based Optimizer with AI (BDC-CMBOAI) technique. The proposed BDC-CMBOAI technique involves the design of Cat and Mouse Optimizer-Based Feature Selection (CMBO-FS) technique to derive a useful subset of features. In addition, Ridge Regression (RR) model is utilized as a classifier to diagnose the disease. Moreover, the utilization of CMBO-FS technique helps in removing the unwanted features and boosts the classification accuracy. The experimental results of the analysis accomplished by BDC-CMBOAI technique on benchmark medical dataset was investigated under different dimensions and the technique's supremacy is established.

2 Related Works

Khademi et al. [11] established a novel scheme to diagnose diabetes. In this study, an ensemble classifier was utilized in applying SVM, KNN, and WOA. WOA is responsible for the creation of weight for all the classifiers to improve the accuracy of diabetes classifiers. The diabetes dataset was collected from Medical centers located in Iran and the empirical analysis was conducted. In literature [12], the authors proposed a novel FS technique to improve the classification accuracy. Differential Evolution optimization algorithm was utilized to find the optimum subset attained by Filter-based FS technique. The performance of the presented FS was evaluated using classifiers as RF, Gradient Boosting Tree, ANN, and SVM.

Abdar et al. [13] presented a novel, easy, and effectual fusion technique with uncertainty-aware component to medicinal image classifier which is named as Binary Residual Feature fusion (BARF). Monte Carlo (MC) dropout was executed to manage the uncertainties and to obtain the mean and Standard Deviation (SD) of forecasts. The presented technique utilized four distinct medical image dataset and tested the same using two important approaches such as direct and cross-validation. The authors in the literature [14] proposed an effectual method to diagnose diabetes using a hybrid-optimized SVM. The presented hybrid optimized approach is a combination of Crow Search algorithm (CSA) and Binary Grey Wolf Optimizer (BGWO). This approach is used to exploit the complete potential of SVM to diagnose the disease. In [15], a novel technique named FR–KDE was proposed. This technique is a combination of FR and Kernel Density Estimation (KDE) from Dempster–Shafer theory. The evidence model was presented to manage the classification issue. When the outcomes of both FR and KDE were fused together using Dempster combination rule, it decreased the uncertainty of FR and attained the optimum accuracy.

3 The Proposed Model

In this study, a novel BDC-CMBOAI technique is presented to determine the occurrence of diseases using biomedical data. The proposed BDC-CMBOAI technique involves different processes namely, pre-processing, feature subset selection using CMBO-FS technique, and RR-based classification. Moreover, the utilization of CMBO-FS technique helps in getting rid of unwanted features and boosts the classification accuracy.

3.1 Algorithmic Design of CMBO-FS Technique

CMBO is a population-based technique simulated by the natural phenomena in which the cat attacks the mouse while mouse gets away from haven. The search agent, from the presented technique, is separated into two sets of cats and mice which scan the problem search spaces in an arbitrary movement. The presented technique upgrades the members of the population through two stages. In primary stage, the progress of cats near mice is demonstrated while in secondary stage, the mice running away to haven so as to save

their life is modeled. From a mathematical viewpoint, all the members of the population denote the presented solutions to the problem. In general, the member of the populations contribute certain values to the problem variables based on their place from search spaces. Therefore, all the members of the population have vector whose value defines the variable of the problem. The population of this technique is defined with the help of a matrix name called population matrix as shown in Eq. (1).

$$X = \begin{bmatrix} X_1 \\ X_i \\ X_N \end{bmatrix}_{N\times m} = \begin{bmatrix} X_{1,1} & \cdots & X_{1,d} & \cdots & X_{1,m} \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ X_{i,1} & \cdots & X_{i,d} & \cdots & X_{i,m} \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ X_{N,1} & \cdots & X_{N,d} & \cdots & X_{N,m} \end{bmatrix}_{N\times m} \tag{1}$$

where X refers to the population matrix of CMBO, X_i implies the i^{th} search agent, $x_{i,d}$ signifies the value of d^{th} problem variable attained by i^{th} search agent, N signifies the number of population members, and m stands for amount of problem variables. As noted, all the members of the population define the presented values to problem variables [16]. So, the value is identified to the main function for all the population members. The values attained to the main function, are represented through a vector given in Eq. (2).

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N\times 1} \tag{2}$$

where F denotes the vector of main purpose value and F_i indicates the main purpose value to the i^{th} search agent. According to the values attained for the main purpose, the member of populations is classified as the optimum member with a minimal value of main function to worse member of populations with maximum value of main function. The arranged population matrix and arranged main function are defined in Eqs. (3) and (4).

$$X^S = \begin{bmatrix} X_1^S \\ \vdots \\ X_i^S \\ \vdots \\ X_N^S \end{bmatrix} = \begin{bmatrix} X_{1,1}^S & \cdots & X_{1,d}^S & \cdots & X_{1,m}^S \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ X_{i,1}^S & \cdots & X_{i,d}^S & \cdots & X_{i,m}^S \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ X_{N,1}^S & \cdots & X_{N,d}^S & \cdots & X_{N,m}^S \end{bmatrix}_{N\times m} \tag{3}$$

$$F^S = \begin{bmatrix} F_1^S & \min(F) \\ \vdots & \vdots \\ F_N^S & \max(F) \end{bmatrix}_{N\times 1} \tag{4}$$

where X^S refers to the sort population matrix which depends upon main function value, X_i^S signifies the i^{th} member of sorting population matrix, $x_{i,d}^s$ represents the value to d^{th} problem variable attained by i^{th} search agent of sorting population matrix, and F^S stands for sorted vector of the main function. In CMBO, it can be considered that half of the population members offered optimum values to the main function so as to create the population of mice whereas the other half of the population members offered lesser values to the main

function which constitutes the population of cats. According to this method, the population of mice and cats are defined through the Eqs. (5) and (6), correspondingly

$$M = \begin{bmatrix} M_1 = X_1^S \\ \vdots \\ M_i = X_i^S \\ \vdots \\ M_{N_m} = X_{N_m}^S \end{bmatrix}_{N_m \times m} = \begin{bmatrix} X_{1,1}^S & \cdots & X_{1,d}^S & \cdots & X_{1,m}^S \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ X_{i,1}^S & \cdots & X_{i,d}^S & \cdots & X_{i,m}^S \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ X_{N_m,1}^S & \cdots & X_{N_m,d}^S & \cdots & X_{N_m,m}^S \end{bmatrix}_{N \times m}, \tag{5}$$

$$C = \begin{bmatrix} C_1 = X_1^S \\ \vdots \\ C_j = X_i^S \\ \vdots \\ C_{N_c} = X_{N_m+N_c}^S \end{bmatrix}_{N_c \times m} = \begin{bmatrix} X_{N_m+1,1}^S & \cdots & X_{N_m+1,d}^S & \cdots & X_{N_m+1,m}^S \\ \vdots & \ddots & \vdots & \iddots & \vdots \\ X_{N_m+j,1}^S & \cdots & X_{N_m+j,d}^S & \cdots & X_{N_m+j,m}^S \\ \vdots & \iddots & \vdots & \ddots & \vdots \\ X_{N_m+N_c,1}^S & \cdots & X_{N_m+N_c,d}^S & \cdots & X_{N_m+N_c,m}^S \end{bmatrix}_{N_c \times m}, \tag{6}$$

where M refers to the population matrix of mice, N_m signifies the number of mice, M_i implies the j^{th} mouse, C stands for population matrix of cats, N_c defines the number of cats, and C_j determines the i^{th} cat. To update the search factor, a primary stage i.e., the the place of the cats is altered according to the natural performance of cats and progress near mice. This upgradation stage of the presented CMBO is mathematical modeled with the help of the Eqs. (7)–(9).

$$C_j^{new} : c_{j,d}^{new} = c_{j,d} + r \times \left(m_{k,d} - I \times c_{j,d}\right) \& j = 1 : N_{c'} d = 1 : m, k \in 1 : N_{m'} \tag{7}$$

$$I = round(1 + rand), \tag{8}$$

$$C_j = \begin{cases} C_j^{new}, |F_j^{c,new} < F_{j^C}, \\ C_j, |else \end{cases} \tag{9}$$

At this point, C_j^{new} denotes the novel condition of j^{th} cat, $c_{j,d}^{new}$ indicates the novel value to d^{th} problem variable attained by j^{th} cat, r implies the arbitrary number in the interval of 0 and1, $m_{k,d}$ means the d^{th} dimensional of k^{th} mouse, $F_j^{c,new}$ stands for the main function value which is dependent upon the novel condition of j^{th} cat. In secondary stage of the presented CMBO, the run-away of the mice to haven is demonstrated. The place of haven, from the search space, is arbitrarily generated based on how the places of distinct members of the technique are modelled. This upgradation stage of the place of mice is mathematically processed through the Eqs. (10)–(12).

$$H_i : h_{i,d} = x_{l,d} \& i = 1 : N_m, d = 1 : m, l \in 1 : N, \tag{10}$$

$$M_i^{new} : m_{i,d}^{new} = m_{i,d} + r \times \left(h_{i,d} - I \times m_{i,d}\right) \times sign\left(F_i^m - F_i^H\right) \& i = 1 : N_m, d = 1 : m, \tag{11}$$

$$M_i = \begin{cases} M_i^{new}, |F_i^{m,new} < F_i^m \\ M_{i'} |else \end{cases} \tag{12}$$

At this point, H_i defines the haven to i^{th} mouse and F_i^H represents their main function value. M_i^{new} signifies the novel condition of i^{th} mouse and $F_i^{m,new}$ denotes the main function value. Then, each member of this population is upgraded and this technique enters the next iteration based on the Eqs. (5)–(12). The technique is applied iteratively until the termination criteria is obtained. This start-to-end

optimization technique denote the specific amount of iterations to determine the acceptable error between the attained solutions and consecutive iteration. Fig. 2 demonstrates the flow process of CMO technique.

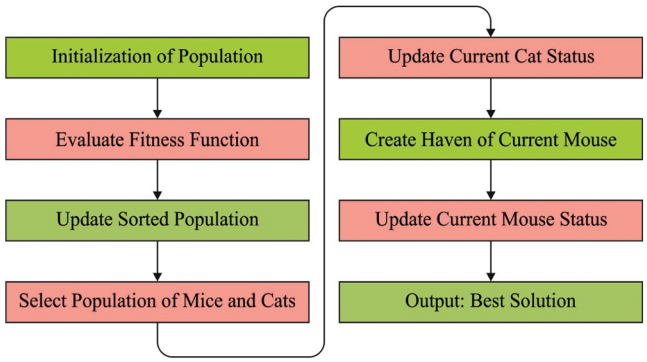

Figure 2: Process in CMO

Algorithm 1 Pseudocode of CMBO

Begin CMBO.
Input problem data: objective function, constraint, and variable.
Fixed the amount of search agents (N) and iteration (T).
Create a primary population matrix at arbitrary.
Estimate the objective function.
For $t = 1 : T$
Arrange the population matrix dependent upon main function values.
Choose the population of mice M.
Choose population of cats C.
Phase1: upgrade condition of cats.
For $j = 1 : N_c$
Upgrade status of j^{th} cat.
end
Phase2: upgrade condition of mice.
For $i = 1 : N_m$
Generate haven to i^{th} mouse.
Upgrade status of i^{th} mouse.
end
End
Outcome optimum quasi-optimal solution attained with CMBO.
End CMBO

Transfer function manner refers to the chances of different place vector elements in terms of *zero* to one, conversely a further simplified ad efficiency model. When exploring the ab optimum group of features, the transfer function significantly controls the resultant of FS approach than it avoids the local optimum problems and preserves the trade-off between exploration and exploitation methods. Based on the above-mentioned scenario, FF provides a solution in this condition to obtain a balance between the purposes as follows.

$$fitness = \alpha\Delta_R(D) + \beta\frac{|Y|}{|T|} \tag{13}$$

$\Delta_R(D)$ signifies the classifier error rate. $|Y|$ implies the size of subsets that this technique selects and $|T|$ refers to the whole count of features contained in the presented dataset. α demonstrates the parameter $\in [0, 1]$ compared with the weight of error rate of classification correspondingly. However $\beta = 1 - \alpha$ stands for the importance of reducing feature.

3.2 Process Involved in RR-Based Classification

At the time of classification, RR model is applied to derive a meaningful subset of features. RR [17], a type of SLFN in which the weight between the input as well as hidden layers are selected following an arbitrary method. So, the weights between the hidden and resultant layers require the learning whereas, in RR, a least square-based learning technique is utilized to tackle binary- and multi-class classifier issues. RR is computationally-free in iterations which generate the results rapidly, by considerably decreasing the computational time required for training the SLFN. SLFN technique frequently needs a maximum amount of hidden neurons since it creates optimum solution. The resultant function of the SLFN with L hidden node is demonstrated through the formulas given below.

$$f_L(x) = \sum_{i=1}^{L} \beta_i g_i x = \sum_{i=1}^{L} \beta_i G(a_i, b_i, x), x \in R^d, \beta_i \in R^m \tag{14}$$

To additive nodes with activation function g, g is demonstrated as follows

$$g_i = G(a_i, b_i, x) = g(a_i x + b_i), a_i \in R^d, b_i \in R \tag{15}$$

$$\sum_{i=1}^{L} \beta_i G(a_i, b_i, x) = t_j, j = 1, \ldots, N \tag{16}$$

The above formula is revised as follows.

$$H\beta = T \tag{17}$$

At this point,

$$H(w_1 \cdots w_L, b_1 \cdots b_L, x_1 \cdots x_N) = \begin{pmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \cdots & g(w_L \cdot x_N + b_L) \end{pmatrix} \tag{18}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{bmatrix} \tag{19}$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix} \tag{20}$$

H implies the hidden layer resultant matrix of NN. SLFN is trained to resolve a linear system optimized problem with the help of subsequent formulas:

$$\left\|H\hat{\beta}-T\right\| = \min_{\beta}||H\beta - T|| \tag{21}$$

At this point, $\hat{\beta}$ is represented as:

$$\hat{\beta} = H^{\dagger}T = \left(H^{T}H\right)^{-1}H^{T}T \tag{22}$$

is the lowest norm least square solution of $w_i = T$ and H^T signifies the Moore Penrose generalization inverse of H.

The process of RR is outlined in the subsequent steps.

Step1 Select the input weight w_i and hidden layer bias b_i arbitrarily.

Step2 Calculate the hidden layer resultant matrix H.

Step3 Achieve the resultant weight $\hat{\beta}$ utilizing the formula $\hat{\beta} = H^{\dagger}T$

4 Experimental Validation

The performance validation of the proposed BDC-CMBOAI technique was conducted using three benchmark medical datasets [18]. The results were inspected under varying runs of execution on each dataset. The details related to the dataset are given in Tab. 1.

Table 1: Dataset description

No.	Datasets	Classes	Instances	Features	Missing value
1	Wisconsin breast cancer (Wisconsin)	2	699	9	Yes
2	Pima Indians diabetes (Pima)	2	768	8	No
3	Thyroid	3	215	5	No

Tab. 2 demonstrates the results of the analysis, accomplished by BDC-CMBOAI technique against other methods on Wisconsin breast cancer dataset, in terms of accuracy and Computation Time (CT) [19].

Table 2: Results of the analysis of BDC-CMBOAI technique with distinct Cross Validation (CV) runs under Wisconsin breast cancer dataset

No. of CV-Runs	BDC-CMBOAI	FOA-SVM	PSO-SVM	Grid-SVM
Accuracy (%)				
CV-Run 1	97.90	97.20	96.10	96.30
CV-Run 2	98.10	97.10	96.00	92.40
CV-Run 3	98.00	96.80	96.40	96.40
CV-Run 4	98.30	96.40	96.70	96.50
CV-Run 5	98.30	97.20	96.80	96.50
Computation Time (s)				
CV-Run 1	3.77	90.16	5.77	41.80
CV-Run 2	4.07	91.11	5.77	33.27
CV-Run 3	3.87	90.16	5.77	39.91
CV-Run 4	5.12	60.77	6.72	38.01
CV-Run 5	4.27	36.11	5.77	34.22

Fig. 3 provides the results for comparative accuracy analysis, achieved by BDC-CMBOAI technique against recent methods under different Cross Validation (CV) runs. The results indicate that the proposed BDC-CMBOAI technique obtained a high accuracy under all CV runs. For instance, with CV run-1, BDC-CMBOAI technique achieved an increased accuracy of 97.90%, whereas FOA-SVM, PSO-SVM, and Grid SVM techniques achieved the least accuracy values such as 97.20%, 96.10%, and 96.10% respectively. Along with that, under CV-run 5, the proposed BDC-CMBOAI technique offered a maximum accuracy of 98.30%, whereas FOA-SVM, PSO-SVM, and Grid SVM techniques reached minimum accuracy values namely, 97.20%, 96.80%, and 96.50% respectively.

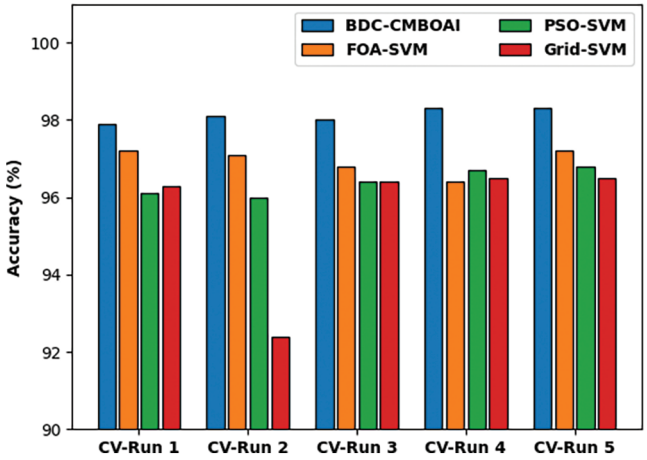

Figure 3: Accuracy analysis of BDC-CMBOAI technique under Wisconsin breast cancer dataset

A detailed CT analysis results accomplished by BDC-CMBOAI technique against existing techniques is shown in Fig. 4. The results infer that BDC-CMBOAI technique achieved better outcomes with least values of CT. For instance, with CV-run 1, BDC-CMBOAI technique required a low CT of 4.27s, whereas other techniques such as FOA-SVM, PSO-SVM, and Grid-SVM reached high CT values such as 90.16, 5.77 and 41.80s respectively. Likewise, with CV-run 5, the presented BDC-CMBOAI technique attained a low CT of 3.77s, whereas FOA-SVM, PSO-SVM, and Grid-SVM techniques demanded more CT values such as 36.11, 5.77 and 34.22s respectively.

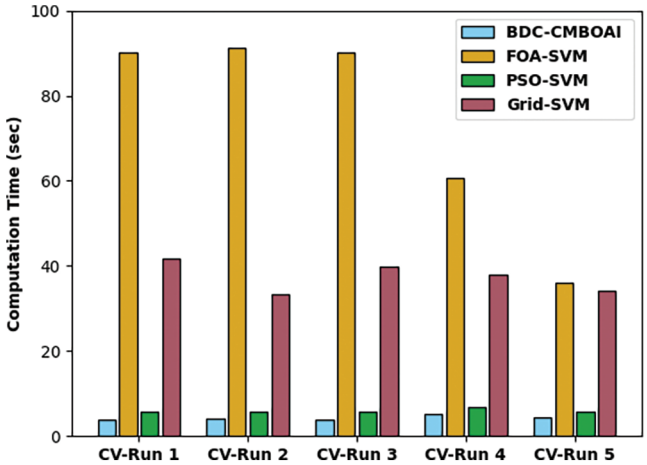

Figure 4: CT analysis of BDC-CMBOAI technique under Wisconsin breast cancer dataset

Fig. 5 demonstrates the ROC analysis results generated by IDTL-MPDC technique on test dataset. The figure exposes that IAOA-DLFD technique reached an enhanced outcome with a minimum ROC of 99.8792.

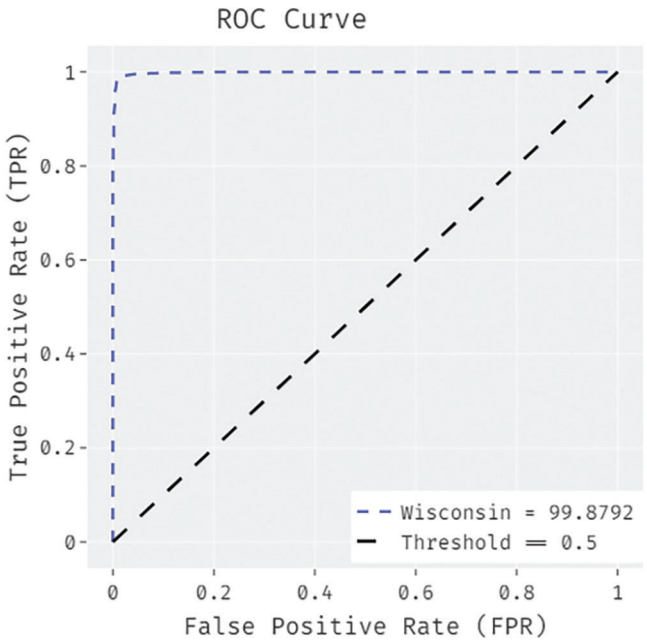

Figure 5: ROC analysis results of BDC-CMBOAI technique under Wisconsin breast cancer dataset

Tab. 3 illustrates the analytical results achieved by BDC-CMBOAI approach against other techniques on Pima Indians diabetes dataset in terms of accuracy and CT.

Table 3: Results of the analysis of BDC-CMBOAI technique with distinct Cross Validation (CV) runs under Pima Indians diabetes dataset

No. of CV-Runs	BDC-CMBOAI	FOA-SVM	PSO-SVM	Grid-SVM
Accuracy (%)				
CV-Run 1	78.12	77.40	76.73	75.92
CV-Run 2	78.22	77.74	77.12	75.68
CV-Run 3	78.03	77.26	76.88	76.35
CV-Run 4	78.46	77.50	76.78	76.40
CV-Run 5	78.75	77.60	76.49	76.11
Computation Time (s)				
CV-Run 1	66.87	79.67	313.11	183.42
CV-Run 2	110.65	131.55	390.93	183.42
CV-Run 3	88.01	105.61	390.93	261.24
CV-Run 4	145.19	157.49	364.99	235.30
CV-Run 5	159.22	183.42	364.99	209.36

Fig. 6 shows the results of comparative accuracy analysis achieved by BDC-CMBOAI methodology against recent techniques under distinct CV runs. The results infer that the proposed BDC-CMBOAI technique obtained a high accuracy under all CV runs. For instance, with CV run-1, BDC-CMBOAI technique achieved a high accuracy of 78.12%, whereas FOA-SVM, PSO-SVM, and Grid SVM methods achieved less accuracy values such as 77.40%, 76.73%, and 75.92% respectively. Besides, under CV-run 5, the BDC-CMBOAI approach offered a maximum accuracy of 78.75%, whereas FOA-SVM, PSO-SVM, and Grid SVM systems gained the least accuracy values such as 77.60%, 76.49%, and 76.11% correspondingly.

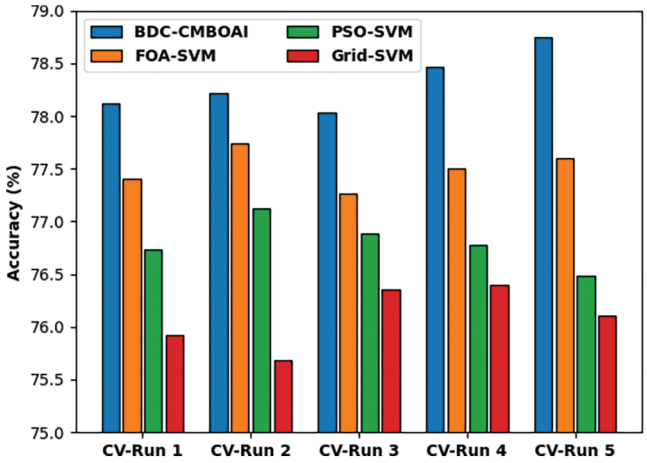

Figure 6: Accuracy analysis results of BDC-CMBOAI technique under Pima Indians diabetes dataset

A brief CT analysis was conducted between BDC-CMBOAI technique and existing techniques and the results are shown in Fig. 7. The outcomes demonstrate the supremacy of the proposed BDC-CMBOAI approach with minimal CT values. For sample, with CV-run 1, the presented BDC-CMBOAI technique decreased the CT value to 66.87s whereas FOA-SVM, PSO-SVM, and Grid-SVM systems reached high CT values such as 79.67, 313.11 and 183.42s respectively. Eventually, with CV-run 5, the proposed BDC-CMBOAI approach attained a reduced CT of 159.22s, whereas FOA-SVM, PSO-SVM, and Grid-SVM methodologies needed high CT values such as 183.42, 364.99 and 209.36s respectively.

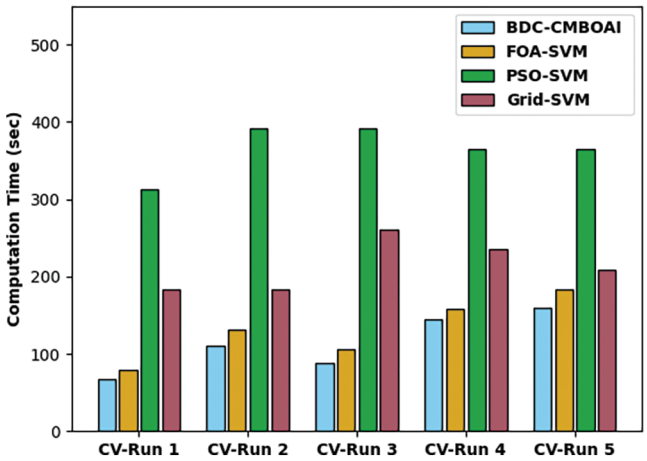

Figure 7: CT analysis of BDC-CMBOAI technique under Pima Indians diabetes dataset

Fig. 8 illustrates the ROC analysis results achieved by IDTL-MPDC method on Pima Indians diabetes dataset. The figure shows that IAOA-DLFD approach obtained superior increased outcomes with a minimal ROC of 95.6760.

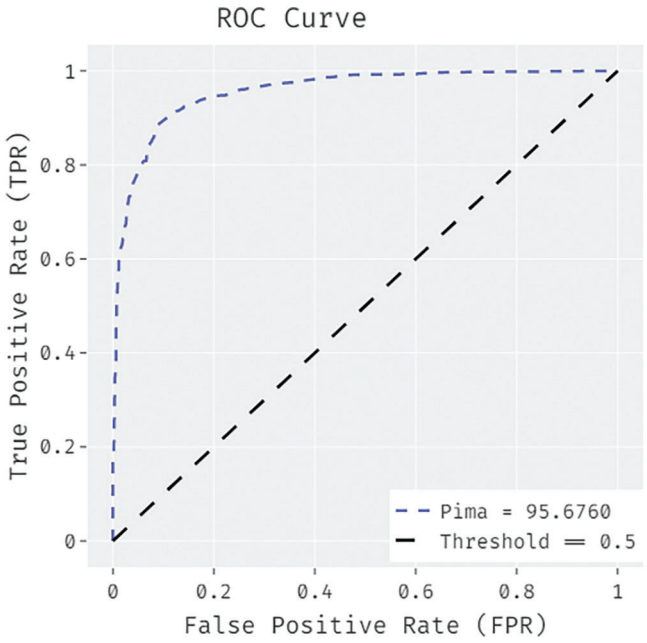

Figure 8: ROC analysis results of BDC-CMBOAI technique under Pima Indians diabetes dataset

Tab. 4 depicts the analytical results accomplished by BDC-CMBOAI method against other techniques on Thyroid dataset with respect to accuracy and CT.

Table 4: Results of the analysis of BDC-CMBOAI technique with distinct Cross Validation (CV) runs under Thyroid dataset

No. of CV-Runs	BDC-CMBOAI	FOA-SVM	PSO-SVM	Grid-SVM
Accuracy (%)				
CV-Run 1	97.28	96.71	94.78	93.12
CV-Run 2	97.67	95.83	93.52	95.26
CV-Run 3	97.48	97.12	96.14	93.86
CV-Run 4	97.86	95.76	95.40	94.81
CV-Run 5	97.62	97.07	95.78	95.78
Computation Time (s)				
CV-Run 1	0.64	16.01	13.12	0.78
CV-Run 2	0.41	16.25	12.03	0.66
CV-Run 3	0.64	17.75	11.97	0.78
CV-Run 4	0.78	19.86	13.54	1.02
CV-Run 5	0.95	16.43	12.82	1.08

Fig. 9 shows the comparative accuracy analysis results accomplished by BDC-CMBOAI approach with recent techniques in terms of varying CV runs. The results reveal that the proposed BDC-CMBOAI technique gained a superior accuracy under all CV runs. For instance, with CV run-1, the presented BDC-CMBOAI system achieved an enhanced accuracy of 97.28% whereas FOA-SVM, PSO-SVM, and Grid SVM techniques achieved less accuracy values namely, 96.71%, 94.78%, and 93.12% correspondingly. In addition, under CV-run 5, BDC-CMBOAI methodology offered an increased accuracy of 97.62%, whereas FOA-SVM, PSO-SVM, and Grid SVM techniques achieved less accuracy values namely, 97.07%, 95.78%, and 95.78% correspondingly.

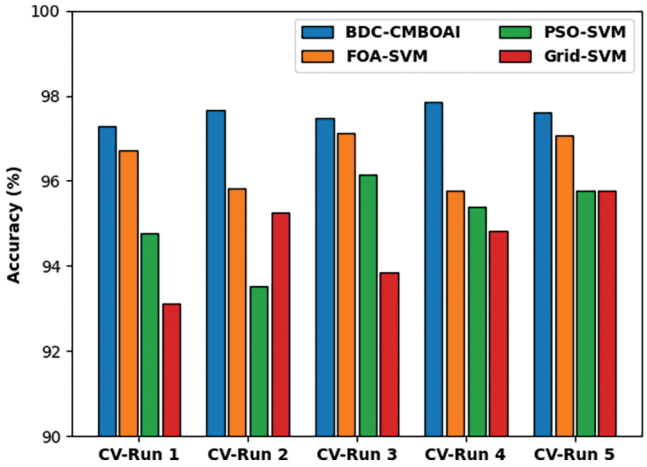

Figure 9: Accuracy analysis of BDC-CMBOAI technique under Thyroid dataset

A detailed CT analysis was conducted between BDC-CMBOAI method against existing methods and the results are shown in Fig. 10. The outcomes show that the presented BDC-CMBOAI approach achieved minimal CT values and outperformed other methods. For sample, with CV-run 1, the BDC-CMBOAI technique required a low CT of 064s, whereas other techniques such as FOA-SVM, PSO-SVM, and Grid-SVM methods gained high CT values such as 16.01, 13.12 and 0.78s correspondingly. In addition, with CV-run 5, BDC-CMBOAI system obtained a low CT of 0.95s, whereas FOA-SVM, PSO-SVM, and Grid-SVM methodologies demanded high CT values such as 16.43, 12.82 and 1.08s correspondingly.

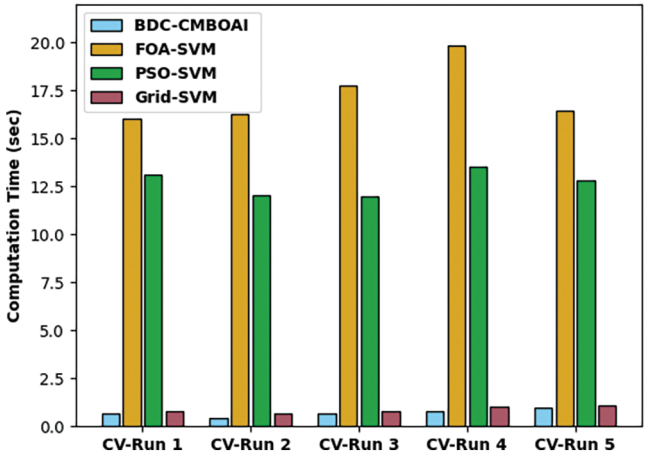

Figure 10: CT analysis of BDC-CMBOAI technique under Thyroid dataset

Fig. 11 illustrates the ROC analysis results achieved by IDTL-MPDC methodology on Thyroid dataset. The figure demonstrates that IAOA-DLFD approach achieved an increased outcome with a low ROC of 99.8176. By looking into the above mentioned tables and figures, it is obvious that the proposed BDC-CMBOAI technique has the ability to attain the maximum medical data classification accuracy.

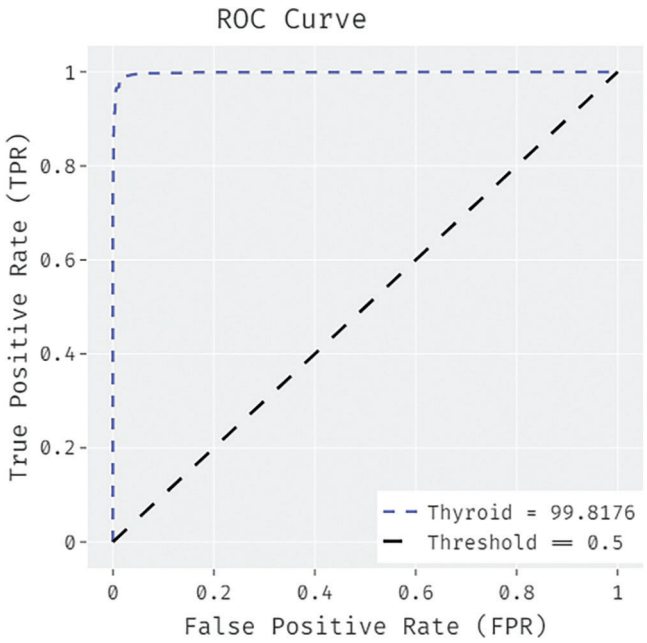

Figure 11: ROC analysis results of BDC-CMBOAI technique under Thyroid dataset

5 Conclusion

In this study, a novel BDC-CMBOAI technique is presented to determine the occurrence of diseases using biomedical data. The proposed BDC-CMBOAI technique involves different processes namely pre-processing, feature subset selection using CMBO-FS technique, and RR-based classification. Moreover, the utilization of CMBO-FS technique helps in removing unwanted features and boosts the classification accuracy. The experimental analysis results of BDC-CMBOAI technique on benchmark medical dataset were investigated under several aspects. The extensive comparative results established the enhanced outcomes of BDC-CMBOAI technique under different evaluation measures. Therefore, BDC-CMBOAI technique can be recognized as a novel approach for biomedical data classification. In future, outlier detection approaches can be utilized to design effective biomedical data classification processes.

Funding Statement: Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R203), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: 22UQU4340237DSR03.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Boyapati, S. R. Swarna, V. Dutt and N. Vyas, "Big data approach for medical data classification: A review study," in *2020 3rd Int. Conf. on Intelligent Sustainable Systems (ICISS)*, Thoothukudi, India, pp. 762–766, 2020.

[2] S. Yang, J. Z. Guo and J. W. Jin, "An improved Id3 algorithm for medical data classification," *Computers & Electrical Engineering*, vol. 65, no. 4, pp. 474–487, 2018.

[3] B. Sahmadi, D. Boughaci, R. Rahmani and N. Sissani, "A modified firefly algorithm with support vector machine for medical data classification," in *IFIP Int. Conf. on Computational Intelligence and its Applications*, Oran, Algeria, pp. 232–243, 2018.

[4] W. Xing and Y. Bei, "Medical health big data classification based on knn classification algorithm," *IEEE Access*, vol. 8, pp. 28808–28819, 2020.

[5] X. Gu, C. Zhang and T. Ni, "Feature selection and rule generation integrated learning for takagi-sugeno-kang fuzzy system and its application in medical data classification," *IEEE Access*, vol. 7, pp. 169029–169037, 2019.

[6] A. Kalantari, A. Kamsin, S. Shamshirband, A. Gani, H. A. Rokny *et al.,* "Computational intelligence approaches for classification of medical data: State-of-the-art, future challenges and research directions," *Neurocomputing*, vol. 276, no. 2, pp. 2–22, 2018.

[7] R. Tang and X. Zhang, "CART Decision tree combined with boruta feature selection for medical data classification," in *2020 5th IEEE Int.Conf. on Big Data Analytics (ICBDA)*, Xiamen, China, pp. 80–84, 2020.

[8] M. M. Kodabagi and A. Tikotikar, "Clustering-based approach for medical data classification, concurrency and computation," *Practice and Experience*, vol. 31, no. 14, 2019.

[9] V. J. Kadam and S. M. Jadhav, "Optimal weighted feature vector and deep belief network for medical data classification," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 18, no. 2, pp. 2050006, 2020.

[10] R. Bania and A. Halder, "R-Ensembler: A greedy rough set based ensemble attribute selection algorithm with kNN imputation for classification of medical data," *Computer Methods and Programs in Biomedicine*, vol. 184, no. 4, pp. 105122, 2020.

[11] F. Khademi, M. Rabbani, H. Motameni and E. Akbari, "A weighted ensemble classifier based on WOA for classification of diabetes," *Neural Computing and Applications*, vol. 25, pp. 3892–3913, 2021.

[12] C. S. Kumar and P. Thangaraju, "Optimal feature subset selection method for improving classification accuracy of medical datasets," *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 2, pp. 3892–3913, 2021.

[13] M. Abdar, M. A. Fahami, S. Chakrabarti, A. Khosravi, P. Pławiak *et al.,* "BARF: A new direct and cross-based binary residual feature fusion with uncertainty-aware module for medical image classification," *Information Sciences*, vol. 577, no. 11, pp. 353–378, 2021.

[14] C. Mallika and S. Selvamuthukumaran, "A hybrid crow search and grey wolf optimization technique for enhanced medical data classification in diabetes diagnosis system," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 157, 2021.

[15] X. Song, B. Qin and F. Xiao, "FR-KDE: A hybrid fuzzy rule-based information fusion method with its application in biomedical classification," *International Journal of Fuzzy Systems*, vol. 23, no. 2, pp. 392–404, 2021.

[16] M. Dehghani, S. Hubálovský and P. Trojovský, "Cat and mouse based optimizer: A new nature-inspired optimization algorithm," *Sensors*, vol. 21, no. 15, pp. 5214, 2021.

[17] P. Mohapatra, S. Chakravarty and P. K. Dash, "Microarray medical data classification using kernel ridge regression and modified cat swarm optimization based gene selection system," *Swarm and Evolutionary Computation*, vol. 28, no. Suppl. 8, pp. 144–160, 2016.

[18] https://archive.ics.uci.edu/ml/datasets.php.

[19] L. Shen, H. Chen, Z. Yu, W. Kang, B. Zhang *et al.,* "Evolving support vector machines using fruit fly optimization for medical data classification," *Knowledge-Based Systems*, vol. 96, no. 3, pp. 61–75, 2016.