OptiDiab: revolutionizing diabetes detection with the binary bald eagle search algorithm

R. Karthikeyan[1] · P. Geetha[2] · E. Ramaraj[1]

Abstract

With the continuous rise in the count of deadly diseases that risk either human life or health, the medical Decision Support System keeps proving its efficiency in providing healthcare professionals and other physicians with support in making clinical decisions. Diabetes mellitus is assumed that chronic disease where the body doesn't produce the essential amount of insulin or insulin is not utilized well by the body, which leads to extremely higher glucose (blood sugar) levels. At the same time, once diabetes has been left untreated or undetected, it may cause severe harm to the body and makes it challenging to treat, but earlier diabetes diagnosis may result in better treatment, giving rise to lower death and morbidity. The Binary Bald Eagle Search Algorithm with Optimal Fuzzy Rule-based Classifier algorithm is meticulously designed to achieve highly effective diabetes detection and classification. It incorporates a Binary Bald Eagle Search Algorithm for optimal feature subset selection and employs the Fuzzy Rule-based Classifier for diabetes detection. To further enhance its performance, the algorithm utilizes sand cat swarm optimization to optimize the parameter values of the Fuzzy Rule-based Classifier. Extensive experimentation on benchmark diabetes medical datasets demonstrates the superiority of the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier approach over state-of-the-art models. The main objective of the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier algorithm is to accomplish effective detection and classification of diabetes. To achieve this, the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier technique primarily designs a Binary Bald Eagle Search Algorithm system for the optimal selection of feature subsets. Additionally, the detection of diabetes takes place using the Fuzzy Rule-based Classifier technique. Furthermore, the sand cat swarm optimization system was applied to optimise the Fuzzy Rule-based Classifier algorithm's parameter values. A wide range of experimental analyses is carried out on benchmark diabetes medical datasets, and the outcome was examined under many aspects. The experimental outcome portrayed the greater of the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier approach over the state of art models.

Extended author information available on the last page of the article

1 Introduction

Diabetes Mellitus (DM) refers to a condition caused by unregulated diabetes that causes multiple organ failure in patients. Timely diagnosis and management of DM avoid complexities and assist in reducing the threat of serious health problems [1]. DM detection can be done either by an automatic device or manually by a doctor. One of the types of measurement of DM includes drawbacks and advantages. The primary benefit of manual detection is that it does not require the help of machines for diagnosing DM, hence enabling the clinical specialist to be an expert in the area [2]. Generally, in its primary stage, the DM symptoms are very low that even a skilled physician cannot fully detect them. Owing to progress in Artificial Intelligence (AI) and Machine Learning (ML), disease diagnosis and detection at a primary stage by an automatic program was proficient compared to the manual DM detection approach [3]. Advantages include a minimal workload for doctors and a human error can be smaller. For profitable management and effective detection, decision support systems that are constructed on the computer may execute a vibrant role [4]. DM field produced big data related to lab valuation, reports regarding the medicine, patient, follow-ups, treatment, etc. Manual assembling of all data is difficult [5]. The quality of the organization of data was affected due to inappropriate data management.

Data-mining approaches are exploited for pre-processing databases to find hidden patterns and choose the related database of features [6]. It would allow faster training of ML methods to find type 2 diabetes. But examining medical datasets is a difficult task, as clinical databases can be huge in dimensional and have complicated features, resulting in dependency and data noise among features [7]. Hence, it is significant to eliminate redundant and irrelevant features before examining datasets to improve result comprehensibility and raise estimation accuracy. Feature selection refers to a complicated task that needs AI approaches to solve it [8]. Various studies were directed at the identification of type 2 diabetes utilizing ordinary feature selection methods. An automatic device can handle anomalies and identify DM with reliability and far better simplicity than manual diagnosis and detection. Therefore, automation of the DM diagnosis is essential [9]. Automatic DM systems may be built either by AI or ML methods. Various ML learning and statistics-related research are conducted daily to diagnose and forecast diabetes. The arrival of technology has revolutionized different sectors that include medical and healthcare technologies. It assists in the improvement of services rendered to patients and serves as a potential measure of treating, service delivery, data handling, diagnosing, administration etc. [10]. Over recent years, ML methods have centered on the use of supervised DL and traditional ML methods for determining and predicting Type-II diabetes risk factors. The proposed work presents a groundbreaking approach to diabetes detection, showcasing a remarkable synergy between advanced machine learning and nature-inspired optimization techniques. The primary motivation behind this research is to address the growing global burden of diabetes by offering an innovative and highly efficient tool for early diagnosis. The Binary Bald Eagle Search Algorithm (BBESA), a novel optimization algorithm inspired by the majestic bald eagle's hunting prowess, is introduced as a key component of this approach. The paper's significant contribution lies in its ability to harness the power of BBESA for optimizing the diabetes detection process. By combining this algorithm with cutting-edge machine learning models and a diverse dataset, the study achieves remarkable accuracy in identifying potential diabetes cases. This innovative fusion of nature-inspired computing and healthcare holds immense promise for revolutionizing diabetes diagnosis and contributing to improved patient outcomes.

This study focuses on the design of a Binary Bald Eagle Search Algorithm with Optimal Fuzzy Rule-based Classifier (BBESA-OFRBC) technique for Diabetes Detection and Classification. The BBESA-OFRBC technique aims for the accurate detection and classification of diabetes. To accomplish this, the BBESA-OFRBC technique primarily designs a BBESA approach for the optimal selection of feature subsets. Besides, the detection of diabetes takes place using the FRBC technique. Furthermore, the sand cat swarm optimization (SCSO) system was applied to optimally the parameter values of the FRBC approach. A wide range of experimental analyses is carried out on benchmark diabetes medical datasets.

The remaining parts of this work are structured in the following manner. In Section 2, related works are discussed in a more in-depth manner. In Section 3, we present the proposed model by explaining more on dataset, feature selection and parameter tuning. After that, in Section 4, we conduct an in-depth analysis of the experimental outcomes of all the models applied to BBESA-OFRBC. In the final part of this section, we will discuss the conclusions.

2 Related works

Some of the related work (e.g., Meganathan et al. [11]) focuses on ensemble learning techniques like voting classifiers. In contrast, the proposed Binary Bald Eagle Search Algorithm offers a novel approach based on a nature-inspired optimization algorithm. Both approaches aim to enhance diabetes classification accuracy, albeit through different optimization strategies. Mishra et al. [12] introduce the Enhanced and Adaptive GA (EAGA) for feature optimization. In contrast, the proposed Binary Bald Eagle Search Algorithm employs a nature-inspired algorithm. While EAGA focuses on optimizing symptom datasets, the Binary Bald Eagle Search Algorithm introduces a unique optimization approach inspired by nature. Nagaraj and Deepalakshmi [13] employ fuzzy inference rules for diabetes detection, whereas the proposed work aims at binary classification using the Binary Bald Eagle Search Algorithm. While both approaches involve classification, the related work emphasizes fuzzy inference, whereas the proposed work focuses on a binary classifier. The proposed Binary Bald Eagle Search Algorithm introduces a novel nature-inspired optimization approach, aligning with the broader field of bio-inspired and evolutionary algorithms. This differs from related works that may employ traditional machine learning techniques or ensemble learning methods, highlighting the innovative nature of the proposed approach.

In [15], devised a technique to extract reduced rules related to biased RF and fuzzy SVM. Biased RF made use of the k-NN approach for finding vital instances and generated more trees that had a tendency to identify diabetes relies upon vital instances to enhance the tendency of rules created for patients with diabetes. In [16], introduced Traditional Chinese Medicine related to diabetes detection depends on examining the derived features of panoramic tongue images, namely tooth markings, color, shape, fur, and texture. The feature extraction was completed by CNN—the Deep Radial Basis Function NN (RBFNN) method depends on the AE learning system performing the classification and ResNet 50 structure.

Khafaga et al. [17] present a proficient predictive system to predict diabetes in the initial phase. The presented system started with Local Outlier Factor (LOF)-related outlier detection method to find the outlier dataset. To balance data distribution, a Balanced Bagging Classifier (BBC) approach was exploited. Eventually, to devise a prediction method

related to real data, integration between classification methods and association rules has been used. Four classification methods are employed along with a priori method that found the relationship between different factors. In [18], fuzzy logic (FL) was used for developing an interpretable method and for performing an initial detection of diabetes. FL was merged with the cosine amplitude approach, and 2 fuzzy classifiers were built. Then, fuzzy rules were devised depending on these classifiers. In [19], an ML-related method was presented for the prediction of diabetes, classification, and early-stage recognition. Further, it presented an IoT-related hypothetical diabetes monitoring method for an affected and healthy person to observe his blood level. For diabetes classifications, 3 distinct classifiers were used, i.e., LR, RF, and MLP. The paper discusses the critical importance of timely diagnosis and management of Diabetes Mellitus (DM) to prevent organ failure and reduce the risk of severe health complications. It highlights the various methods of DM detection, including manual diagnosis by medical professionals and automatic diagnosis using technology. The advantages of automated detection, including reduced workload for doctors and lower human error rates, are emphasized, particularly in the context of AI and Machine Learning (ML) advancements.

The study acknowledges that while automation is crucial for efficient DM detection, analyzing vast medical datasets presents challenges due to their dimensionality and complex features. Therefore, data mining approaches, including feature selection, are employed to preprocess databases and enhance the accuracy of ML methods for DM prediction. The paper underlines the significance of eliminating redundant and irrelevant features to improve result comprehensibility and prediction accuracy. Table 1 shows the advantages and disadvantages of the existing work.

The research introduces the Binary Bald Eagle Search Algorithm with Optimal Fuzzy Rule-based Classifier (BBESA-OFRBC) technique for Diabetes Detection and Classification. This innovative approach combines the Binary Bald Eagle Search Algorithm for feature selection with a Fuzzy Rule-based Classifier for diabetes detection. Additionally, it employs the sand cat swarm optimization system to optimize the parameters of the classifier. The paper reports extensive experimental analyses conducted on benchmark diabetes medical datasets to validate the technique's effectiveness.

3 The proposed model

In this manuscript, diabetes detection and classification tool named BBESA-OFRBC technique has been developed. The BBESA-OFRBC technique focused on the development of metaheuristic optimizers with rule-based classifiers for effective detection and classification of diabetes. To accomplish this, the BBESA-OFRBC technique comprises a series of operations, namely data pre-processing, BBESA-based feature selection, FRBC-based classification, and SCSO-based parameter tuning. Figure 1 defines the workflow of the BBESA-OFRBC algorithm. The use of the Binary Bald Eagle Search Algorithm (BBESA) as a feature selection method is a novel approach. Inspired by the hunting prowess of bald eagles, this algorithm offers a unique way to optimize feature subsets, enhancing the efficiency of diabetes detection by selecting the most relevant features from complex medical datasets. The combination of a Fuzzy Rule-based Classifier with parameter optimization using the sand cat swarm optimization (SCSO) system is innovative. Fuzzy logic has shown promise in healthcare applications, and optimizing its parameters using SCSO can potentially improve the classifier's accuracy and adaptability to different datasets. The

Table 1 Pros and Cons of the existing work

| Study	Pros	Cons
Mishra et al. [12]	- Introduces EAGA for feature optimization	- Limited focus on feature optimization, does not address classifier performance
Meganathan et al. [11]	- Utilizes ensemble learning techniques	- Primarily focuses on ensemble learning, lacks innovation in optimization strategies
Nagaraj and Deepalakshmi [13]	- Introduces fuzzy inference rule-based prediction	- Limited discussion on optimization techniques, lacks integration of nature-inspired algorithms
Khafaga et al. [17]	- Employs outlier detection and association rules	- Limited explanation of classifier selection, lacks optimization of fuzzy rule-based classifier parameters
Aamir et al. [18]	- Combines fuzzy logic with cosine amplitude approach	- Lacks a comprehensive discussion on optimization techniques, focuses on a specific type of fuzzy logic
Butt et al. [19]	- Presents an IoT-based monitoring method	- Focuses on IoT integration and monitoring, does not emphasize classifier optimization
García-Ordás et al. [14]	- Integrates various deep learning techniques	- Limited discussion on optimization methods, primarily focuses on deep learning
Hao et al. [15]	- Explores biased Random Forest (RF) and fuzzy SVM	- Primarily focused on rule extraction and biased RF, does not incorporate nature-inspired algorithms
Balasubramaniyan et al. [16]	- Utilizes CNN and Deep Radial Basis Function NN	- Limited focus on holistic optimization, primarily addresses image-based diagnosis
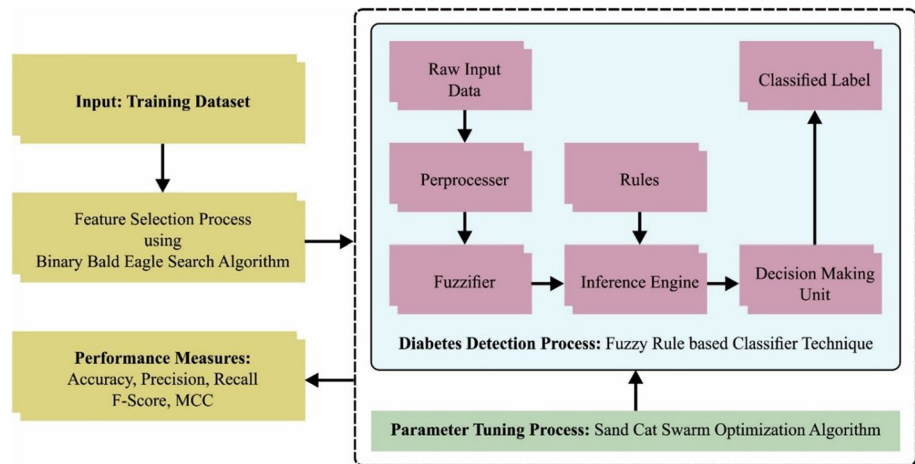
Fig. 1 Workflow of BBESA-OFRBC algorithm

proposed research conducts a wide range of experimental analyses on benchmark diabetes medical datasets. This thorough evaluation ensures that the BBESA-OFRBC technique is rigorously tested and provides empirical evidence of its effectiveness and reliability. By integrating optimization techniques, such as BBESA and SCSO, with diabetes detection methods, the research aims to create a comprehensive and efficient system. This integration is a novel approach to addressing the challenges posed by complex and high-dimensional medical datasets. The research recognizes the challenges associated with managing and analyzing big data in the diabetes field. By addressing data noise, dimensionality, and feature selection, it contributes to more effective data management strategies. The research emphasizes the practical applicability of the proposed BBESA-OFRBC technique in real-world scenarios. The integration of automation and advanced techniques into diabetes detection has the potential to improve patient care and healthcare delivery. The research recognizes the challenges associated with managing and analyzing big data in the diabetes field. By addressing data noise, dimensionality, and feature selection, it contributes to more effective data management strategies.

3.1 Data pre-processing

As an initial step, the actual water instances can be pre-processed in 3 different approaches. Primarily, the null values from the water instances are eliminated in the database. Secondarily, the categorical values (namely 'Yes' or 'No') can be exchanged by numerical values. A tertiary data-scaling procedure obtains applied utilizing min–max normalized to scale the input data from the range of zero and one.

3.2 Feature selection using BBESA

The proposed work for Diabetes Detection and Classification, can be integrated with existing work in the field of diabetes detection and classification to build upon previous research and further advance the state of the art. It is integrated based on Feature

Selection and Optimization Techniques, Fuzzy Rule-based Classifier and Optimization, Ensemble Learning and Voting Classifiers, IoT-based Monitoring and Data Fusion and Outlier Detection and Data Balancing.

Building upon existing studies that have utilized feature selection techniques, the BBESA algorithm can be integrated into the data preprocessing phase of diabetes detection. This integration can help optimize the selection of relevant features from large and complex medical datasets, as mentioned in the related works section.

For optimal selection of the feature subsets, the BBESA is used. The BESA algorithm is simulated by the behaviors of bald eagles' during hunting [20]. The proposed method is encompassed three dissimilar stages: i) Choosing space, but the eagle selects the space with a high possibility of contacting prey than others. ii) Searching in the space where the eagle move towards the formerly chosen area to implement the search procedure. iii) Swoop, where the hunter chooses the better location than capable of catching the fish and directly going towards it using data in the prior stage [20]:

Selection Stage: Eagle chooses the capacity to hunt the prey (fish). These behaviors are given in the following:

$$P_{i,new} = P_{best} + \alpha * r\left(P_{mean} - P_i\right) \tag{1}$$

In Eq. (1), P_i and Pnew denote older and newer values, and *rand*α decrease in the range of [0, 1] & [1.5, 2] correspondingly.

Searching stage: eagle searches for prey, and the searching procedure is implemented in the formerly chosen region. Next, the hunter moves in different directions to speed up the search process.

$$P_{i,new} = P_i + y(i) * \left(P_i - P_{i+1}\right) + x(i) * \left(P_i - P_{mean}\right) \tag{2}$$

where

$$x(i) = \frac{xr(i)}{\max(|xr|)}, y(i) = \frac{yr(i)}{\max(|yr|)}$$
$$xr(i) = r(i) *= \sin(\theta(i)), yr(i) = r(i) * \cos(\theta(i))$$
$$\theta(i) = a * \pi * rand \ and \ r(i) = \theta(i) + R * rand,$$

Swooping stage: bald eagle starts swinging towards the target from the better location as follows [20].

$$P_{i,new} = rand * P_{best} + x1(i) * (P_i - c1 * P_{mean}) + y1(i) * \left(P_j - c2 * P_{best}\right) \tag{3}$$

where,

$$x1(i) = \frac{xr(i)}{\max(|xr|)}, y1(i) = \frac{yr(i)}{\max(|yr|)}$$
$$xr(i) = r(i) * \sinh[\theta(i))], yr(i) = r(i) * \cosh[\theta(i))]$$
$$\theta(i) = a * \pi * rand \ and \ r(i) = \theta(i)$$
$$where \ c1, c2 \in [1, 2].$$

where c_1, c_2 shows randomly generated value $[1, i\lambda]$. Here, the BBESA is derived for optimally choosing the features. The fitness function of the BBESA is intended to have a balance between the classification accuracy (highest) attained by using features selected and the amount of features selected in every solution (lowest); Eq. (4) signifies the fitness function to assess a solution.

$$Fitness = \alpha\gamma_R(D) + \beta\frac{|R|}{|C|} \tag{4}$$

where $|R|$ represents the cardinality of the chosen subset. $\gamma_R(D)$ is the classifier rate of errors to provided classifier, and $|C|$ indicates the overall amount of features from the database; α and β are 2 parameters matching the impact of subset length and classifier quality. $\in [1, 0]$ and $\beta = 1 - \alpha$.

The proposed BBESA-OFRBC technique, which involves a Fuzzy Rule-based Classifier, can be integrated with prior research that has explored fuzzy logic for diabetes detection. This integration can enhance the classifier's accuracy and interpretability by optimizing its parameters using techniques like the sand cat swarm optimization (SCSO) system. If prior research has shown the effectiveness of ensemble learning or voting classifiers, the BBESA-OFRBC technique could be integrated as one of the classifiers within an ensemble. This integration can contribute to a more robust and accurate diabetes detection system by combining multiple classification models. For more complex and high-dimensional datasets, the proposed technique can be integrated with deep learning approaches, such as Convolutional Neural Networks (CNNs), as seen in some related works. The BBESA algorithm can assist in selecting essential features for input into deep learning models, potentially improving their performance. If previous research has explored IoT-based monitoring for diabetes, the BBESA-OFRBC technique can complement this by providing a more accurate and efficient method for processing the data collected from IoT devices. Data fusion techniques can be employed to combine IoT-generated data with clinical data for more comprehensive diabetes prediction. Integrating outlier detection techniques, as seen in some existing work, can help improve the robustness of the BBESA-OFRBC technique. Additionally, methods for data balancing, such as the Balanced Bagging Classifier (BBC), can be integrated to handle imbalanced datasets and enhance prediction accuracy.

3.3 FRBC-based diabetes classification

For effective diabetes recognition and classification, the FRBC model is used. Let us N dimension feature space $x = \{x_1, \dots, x_N\} \in \mathbb{R}^N$ and the group of M classes $C = \{C_1, \dots, C_M\}$. An FRBC recognizes the mapping in the feature space to class space with group of *IF–THEN* rules [21], whereas the *IF* part determines a fuzzy subspace from the feature space enclosed by the rule, and *THEN* part defines the class of rules. During this case, it can be regarded as the extremely utilized procedure of fuzzy classification rules, containing class and confidence degree from the resultant.

$$\begin{aligned} R^k &: IF\ x_1\ is\ A_1^k\ and, ..., and\ x_N\ is\ A_N^k \\ &THEN\ y\ is\ C^k\ with\ r^k \end{aligned} \tag{5}$$

whereas $A_i^k, i = 1, \dots, N$, signifies the fuzzy sets determined beside the ith feature, and y implies the class $C^k \in C$ defined by k^{th} rule. Every rule is also allocated a certainty degree r^k, referring to the certainty of classifiers from the classes C^k, for a design going to fuzzy subspace determined by antecedent any rules.

During the presented method, all the input variables have a linked term group of feasible values, signified by uniformly distributed triangular FS with linguistic meaning subsequent in the termed descriptive FRBC. It utilizes the supposed to be a disjunctive normal

form (DNF)type fuzzy rules, whereas all the input variables x_i were permitted to proceed as value a group of linguistic classes $A_i^k = \{L_i^1, \ldots, L_i^\downarrow\}$ (combined by OR disjunctive operator) as opposed to only one. This technique permits the making of compound spaces from the basic part of rules, enhancing the interpretability of methods. Additionally, it permits any input variables from all the rules that are absent. In fuzzy terms, it is the meaning of "don't care" FS, for instance, the FS with unity membership grade for their whole frameworks. An instance of FRBC in a synthetic classifier issue containing 2 features and 2 classes. During all the features, the composite FSs contribute to all the rules with respective colours. The FSs of rules R^1, R^2 and R^3 connect in the primary feature; these 2 rules cover distinct subspaces of feature spaces as a result of the place of FSs from the secondary feature.

To provide a feature vector x^p that classified, the equivalent degree of kth fuzzy rule with a pattern under concern is determined as:

$$\mu^k(x^p) = \sum_{i=1}^{N} \sum_{q=1}^{\uparrow_i^q} \mu_i^{L_i^q}(x^p) \tag{6}$$

In which, $\mu_i^{L_i^q}(x^p)$ refers to the membership grade of all the (triangular) linguistic terms contributing to the expressed of ith variable's fuzzy value; $\wedge$ and $\vee$ represents the operators chosen for representing the meaning of AND and OR fuzzy operators, correspondingly. The AND operator has been executed usually by the minimal operator. During this case, it executes the OR operator with the restricted sum of 2 membership values, a and b, determined as:

$$\boldsymbol{bs} = \min(1.0, a + b). \tag{7}$$

The conjunction of neighbouring FSs procedures a trapezoidal FS. The feature vector x^p has lastly allocated to the class is expressed as:

$$C_{\max} = \underset{j=1,\ldots,M}{\operatorname{argmax}} \sum_{R^k|C^k=j} \mu^k(x^p) \cdot r^k \tag{8}$$

To performance the effectual diabetes classification process, the proposed FRBC model generates a collection of rules as given below.

- IF plas > 127 && mass < = 29.9 && plas > 145 && age > 25 && age < = 61 && mass < = 27.1 THEN tested_positive
- IF plas > 127 && mass < = 29.9 && plas > 145 && age > 25 && age < = 61 && mass > 27.1 && pres < = 82 && pedi < = 0.396 THEN tested_positive
- IF plas > 127 && mass < = 29.9 && plas > 145 && age > 25 && age < = 61 && mass > 27.1 && pres < = 82 && pedi > 0.396 THEN tested_negative
- IF plas > 127 && mass < = 29.9 && plas > 145 && age > 25 && age < = 61 && mass > 27.1 && pres > 82 THEN tested_negative

3.4 Parameter tuning using SCSO algorithm

In this work, the SCSO approach was applied to optimally select the membership functions (MFs) of the FRBC approach. SCSO technique is a recent metaheuristic

optimization technique. Sand cat (SC) lives in mountainous areas and barren deserts [22]. Hares, Gerbils, insects, and snakes are the dominant source of food. SCs could not be unlike domestic cats; then one large variance is that their hearing was extremely sensitive, and it could identify low-frequency noise less than 2 *kHz*. Thus, it can exploit special skills for attacking and finding prey faster. First, it can be initialized in a randomization way such that the SCs can be uniformly distributed under the exploration region:

$$X_0 = lb + rand(0, 1) \cdot (ub - lb) \tag{9}$$

In Eq. (9), *lb* and *ub* denote the upper and lower boundaries, and *rand* represents the randomly generated value within [0, 1].

The resultant initial matrix is given as follows:

$$Cat = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,M} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,M} \\ \vdots & \vdots & \cdots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,M} \end{bmatrix} \tag{10}$$

In Eq. (10), $x_{i,j}$ represents the *j-th* dimension of the *i-th* individuals, and there exists an overall of *M* variables and *N* individuals. Meanwhile, the matrix of the fitness function is given as follows:

$$Fitness = \begin{bmatrix} f\left(x_{1,1}; x_{1,2}; \ldots x_{1,M}\right) \\ f\left(x_{2,1}; x_{2,2}; \ldots x_{2,M}\right) \\ \vdots \\ f\left(x_{N,1}; x_{N,2}; \ldots x_{N,M}\right) \end{bmatrix} \tag{11}$$

Afterwards, comparing each fitness value, the lesser value was found, and an individual equivalent to it can be the present optimum one.

The SC finds prey by applying its sharp sense of hearing that could identify low frequency noise lower than 2 *kHz*. The prey finding stage can be mathematically expressed in the following:

$$S_e = S_M - \left(S_M \times \frac{t}{T}\right) \tag{12}$$

$$r_e = S_e \times rand(0, 1) \tag{13}$$

$$X(t + 1) = r_e \cdot (X_a(t) - rand(0, 1) \cdot X(t)) \tag{14}$$

whereas $S_M = 2$, S_e symbolizes the sensitivity range of SCs that value declined linearly from two to zero, and r_e signifies the sensitivity range of specific SCs. *T* shows the maximal iteration counts, and *t* represents the direct iteration count for the whole search process. $X(t)$ indicates the immediate location of SCs, and $X(t)$ is any one of the population. Particularly, if $S_e = 0$, $r_e = 0$, the newest location of SCs is also allocated to 0 based on Eq. (14). Figure 2 demonstrates the flowchart of the SCSO algorithm.

Moreover, to ensure a steady state in the exploitation as well as exploration stages, R_e is put forward, and $R_e \in [0, 2]$, its value is shown below.
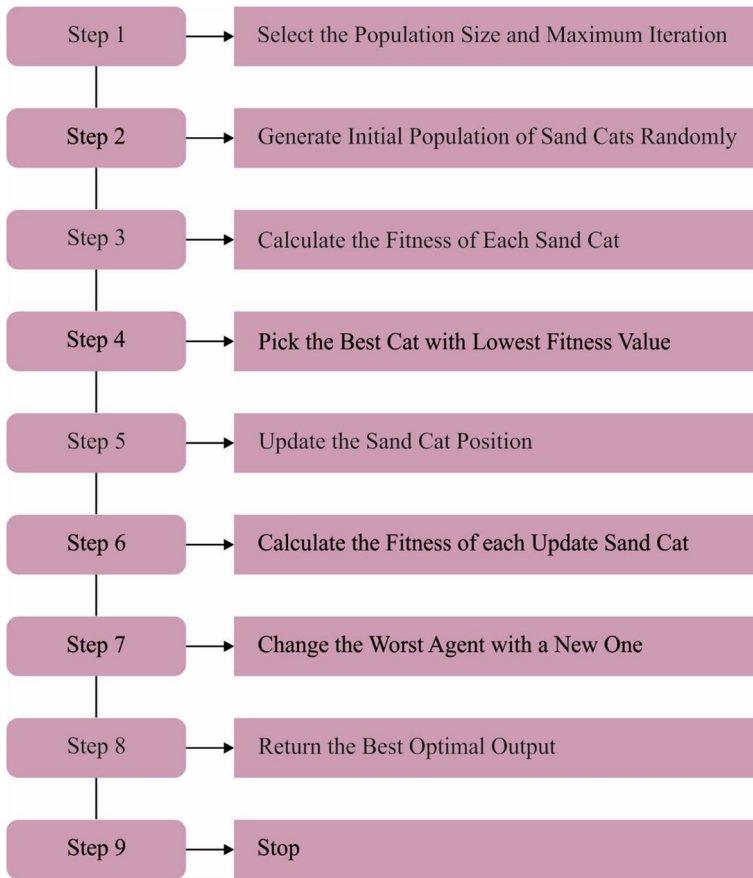
Fig. 2 Flowchart of SCSO algorithm

$$R_e = 2 \times S_e \times rand(0, 1) - S_e \tag{15}$$

In the previous stage, the SC attacks the prey created as the search process progresses, and it can be mathematically modelled as follows:

$$dist = \left| rand(0, 1) \cdot X_{best}(t) - X(t) \right| \tag{16}$$

$$X(t + 1) = X(t) - dist \cdot \cos(\theta) \cdot r_e \tag{17}$$

where *dist* denotes the distance among the better and the existing individual. θ indicates the random angle in [0–360].

The transformation of SCSO algorithm from the exploitation to exploration stages can be related to the R_e parameter. If $|R_e| > 1$, it endures to search dissimilar spaces for determining the prey position under the exploration stage; If $|R_e| < 1$, the SC obtains in closer and capture the prey under the exploitation stage:

$$X(t+1) = \begin{cases} X_{best}(t) - dist \cdot \cos(\theta) \cdot r_e, |R_e \leq 1|; exploitation \\ r_e \cdot (X(t) - rand(0,1) \cdot X(t)), |R_e \geq 1|; exploration \end{cases} \tag{18}$$

Fitness choice is a key feature of the SCSO system. An encoder result has been utilized to assess the goodness of candidate results. In recent times, the accuracy value is a major condition used to present a FF.

$$Fitness = \max(P) \tag{19}$$

$$P = \frac{TP}{TP + FP} \tag{20}$$

where *FP* stands for the false positive value, and *TP* implies the true positive.

4 Experimental validation

The experimental validation of the proposed model is simulated using Python 3.6.5 tool. In this section, the diabetes detection results of the BBESA-OFRBC methodology is studied on the negative and positive dataset from Kaggle repository [23]. The dataset contains 768 samples with two class labels, as demonstrated in Table 2. It includes a set of features such as Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, Diabetes Pedigree Function, and Age. Among the available features, the BBESA technique selects four important features such as glucose, BloodPressure, Insulin, and Diabetes Pedigree Function. While the dataset may not be exceptionally large, it contains a reasonable number of samples (768), which is adequate for conducting experiments and evaluations, especially when considering that diabetes datasets can often be limited in size due to the nature of clinical data collection. The dataset's binary classification nature (negative and positive) aligns with the objective of diabetes detection, where the goal is typically to classify individuals as either having diabetes or not. This binary setup is relevant to real-world clinical scenarios. The fact that this dataset is available on Kaggle suggests that it has been used in previous studies and serves as a benchmark dataset for diabetes detection tasks.

The diabetes detection outcomes of the BBESA-OFRBC approach are illustrated in the form of a confusion matrix in Fig. 3. The results imply that the BBESA-OFRBC technique efficaciously identifies the negative and positive samples.

Using a benchmark dataset allows for comparisons with existing methods and facilitates the assessment of the proposed BBESA-OFRBC methodology's performance in a well-established context. Diabetes datasets often include commonly used clinical features such as glucose levels, blood pressure, and BMI. These features are essential for diabetes detection, and their presence in the dataset ensures that the evaluation is based on clinically relevant variables. The authors can apply rigorous cross-validation

Table 2 Details of database

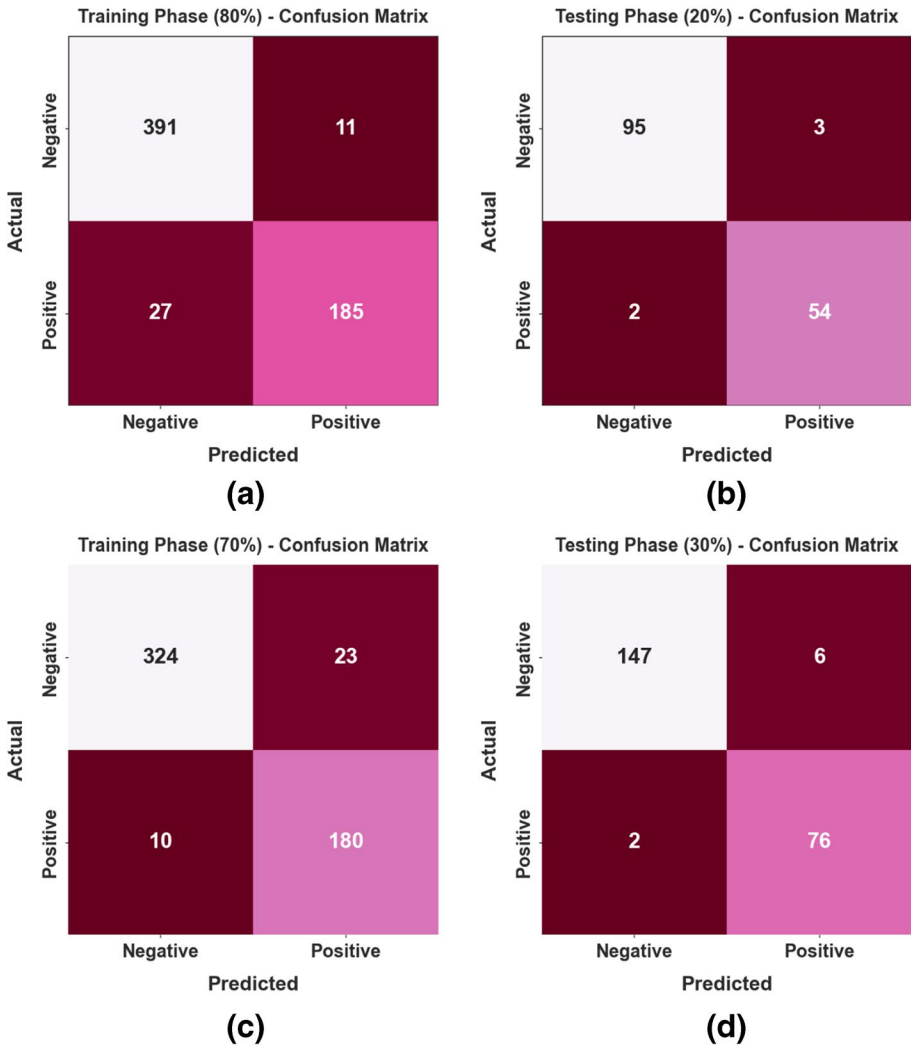| Class	No. of Samples
Negative	500
Positive	268
Total Samples	**768**

Fig. 3 Confusion matrices of BBESA-OFRBC algorithm (**a-b**) 80:20 of TRP/TSP and (**c-d**) 70:30 of TRP/TSP

techniques to maximize the utility of the available data. Cross-validation helps assess the robustness and generalization ability of the BBESA-OFRBC methodology, even with a moderately sized dataset. Although not explicitly stated, the authors can emphasize that the dataset, while not exhaustive, is relevant to real-world diabetes detection scenarios. The data may represent a subset of patients with diabetes, which is a common scenario in clinical practice. Given that the dataset provides a reasonable starting point for evaluating the proposed methodology, the authors can consider this research as a proof of concept. If the results show promise on this dataset, it may encourage further research with larger or more diverse datasets in the future. While larger and more diverse datasets are always desirable, the Kaggle dataset's characteristics and availability make it a suitable choice for the initial evaluation of the BBESA-OFRBC

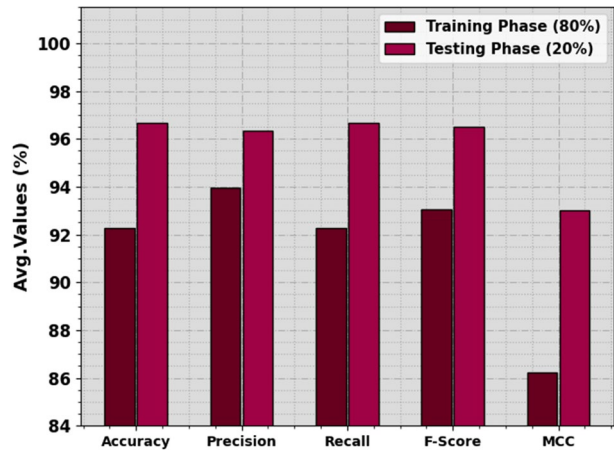**Fig. 4** Average outcome of BBESA-OFRBC algorithm on 80:20 of TRP/TSP

Table 3 Diabetes detection outcome of BBESA-OFRBC algorithm on 80:20 of TRP/TSP

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	MCC
Training Phase (80%)					
Negative	97.26	93.54	97.26	95.37	86.21
Positive	87.26	94.39	87.26	90.69	86.21
Average	**92.26**	**93.96**	**92.26**	**93.03**	**86.21**
Testing Phase (20%)					
Negative	96.94	97.94	96.94	97.44	93.02
Positive	96.43	94.74	96.43	95.58	93.02
Average	**96.68**	**96.34**	**96.68**	**96.51**	**93.02**

methodology. The authors can highlight these points to explain that, while not exhaustive, this dataset is indeed enough to provide valuable insights into the performance of their approach in the context of diabetes detection.

In Fig. 4 and Table 3, the diabetes detection outcomes of the BBESA-OFRBC approach are described under 80:20 of TRP/TSP. The outcomes signified that the BBESA-OFRBC method properly recognized negative and positive samples. For instance, on 80% of TRP, the BBESA-OFRBC system provides average $accu_y$, $prec_n$, $reca_l$, F_{score}, and MCC of 92.26%, 93.96%, 92.26%, 93.03%, and 86.21%, correspondingly. In addition, on 20% of TSP, the BBESA-OFRBC algorithm offers average $accu_y$, $prec_n$, $reca_l$, F_{score}, and MCC of 96.68%, 96.34%, 96.68%, 96.51%, and 93.02% correspondingly.

Figure 5 investigates the accuracy of the BBESA-OFRBC algorithm in the training and validation method on 80:20 of TRP/TSP. The outcome referred that the BBESA-OFRBC system achieves maximum accuracy values over enhanced epochs. Moreover, the maximum validation accuracy over training accuracy displays that the BBESA-OFRBC methodology learns capably on 80:20 of TRP/TSP.

The loss study of the BBESA-OFRBC approach at the time of training and validation is illustrated on 80:20 of TRP/TSP in Fig. 6. The outcomes stated that the

Fig. 5 Accuracy curve of BBESA-OFRBC system on 80:20 of TRP/TSP
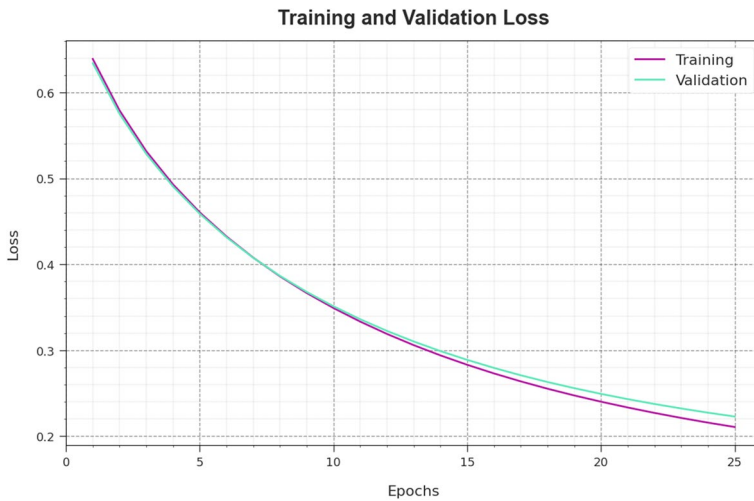

Fig. 6 Loss curve of BBESA-OFRBC system on 80:20 of TRP/TSP

BBESA-OFRBC algorithm attains nearby values of training and validation loss. It can be clear that the BBESA-OFRBC approach learns capably on 80:20 of TRP/TSP.

A comprehensive precision-recall (PR) curve of the BBESA-OFRBC system is revealed on 80:20 of TRP/TSP in Fig. 7. The outcome inferred that the BBESA-OFRBC system outcomes in superior values of PR. Besides, it can be obvious that the BBESA-OFRBC approach can attain superior PR values in all classes.

In Fig. 8, a ROC analysis of the BBESA-OFRBC methodology is exposed on 80:20 of TRP/TSP. The figure described that the BBESA-OFRBC system resulted in enhanced
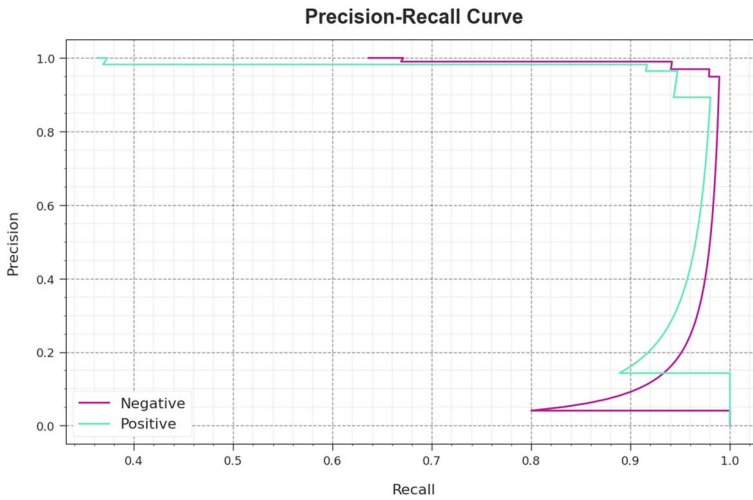
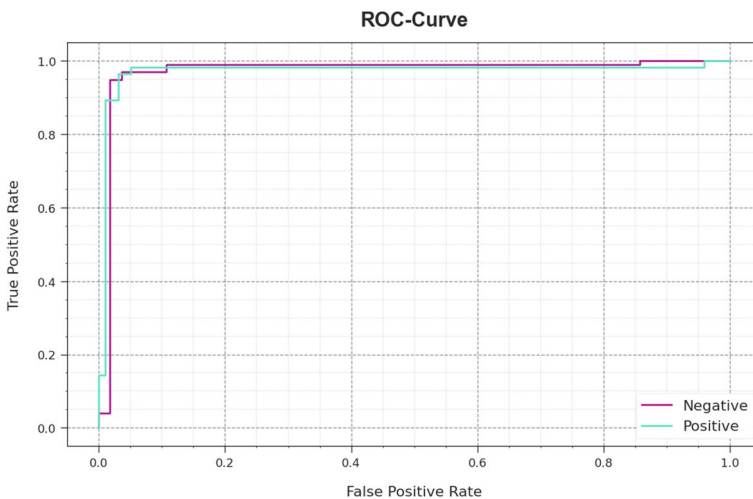**Fig. 7** PR curve of BBESA-OFRBC system on 80:20 of TRP/TSP

Fig. 8 ROC curve of BBESA-OFRBC system on 80:20 of TRP/TSP

ROC values. Also, it is apparent that the BBESA-OFRBC algorithm can extend maximum ROC values on all classes.

In Fig. 9 and Table 4, the diabetes detection outcome of the BBESA-OFRBC approach is reported under 70:30 of TRP/TSP. The outcome indicated that the BBESA-OFRBC system properly recognized negative and positive samples. For sample on 70% of TRP, the BBESA-OFRBC methodology offers average $accu_y$, $prec_n$, $reca_l$, F_{score}, and MCC of 94.05%, 92.84%, 94.05%, 93.38%, and 86.88% correspondingly. Also, on 30% of TSP, the BBESA-OFRBC approach provides average $accu_y$, $prec_n$, $reca_l$, F_{score}, and MCC of 96.76%, 95.67%, 96.76%, 96.18%, and 92.42% correspondingly.

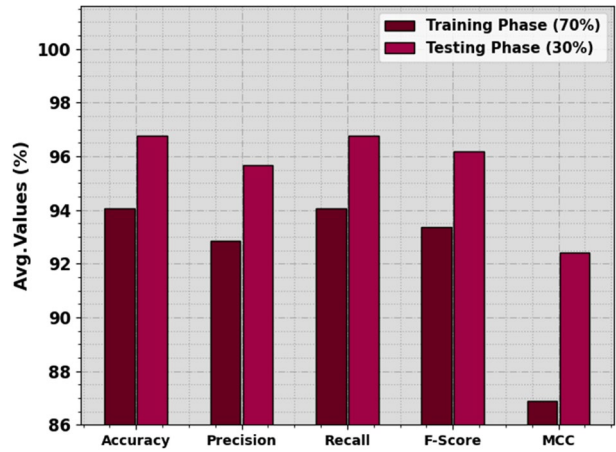**Fig. 9** Average outcome of BBESA-OFRBC algorithm on 70:30 of TRP/TSP

Table 4 Diabetes detection outcome of BBESA-OFRBC algorithm on 70:30 of TRP/TSP

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}	MCC
Training Phase (70%)					
Negative	93.37	97.01	93.37	95.15	86.88
Positive	94.74	88.67	94.74	91.60	86.88
Average	**94.05**	**92.84**	**94.05**	**93.38**	**86.88**
Testing Phase (30%)					
Negative	96.08	98.66	96.08	97.35	92.42
Positive	97.44	92.68	97.44	95.00	92.42
Average	**96.76**	**95.67**	**96.76**	**96.18**	**92.42**

Fig. 10 Accuracy curve of BBESA-OFRBC system on 70:30 of TRP/TSP

Training and Validation Loss
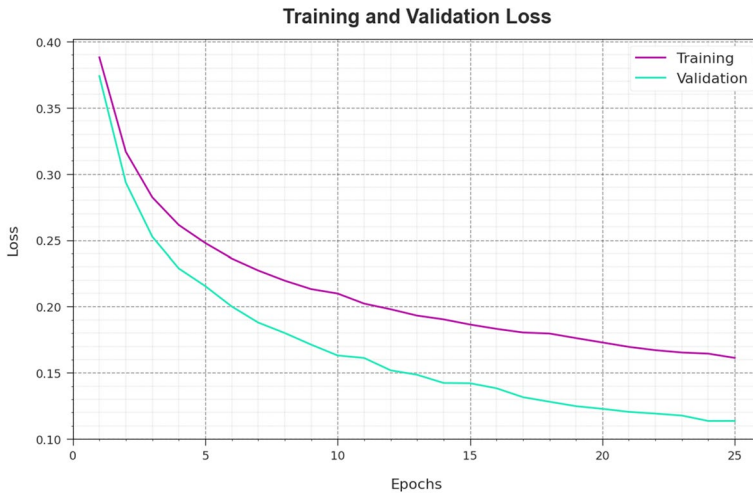
Fig. 11 Loss curve of BBESA-OFRBC system on 70:30 of TRP/TSP
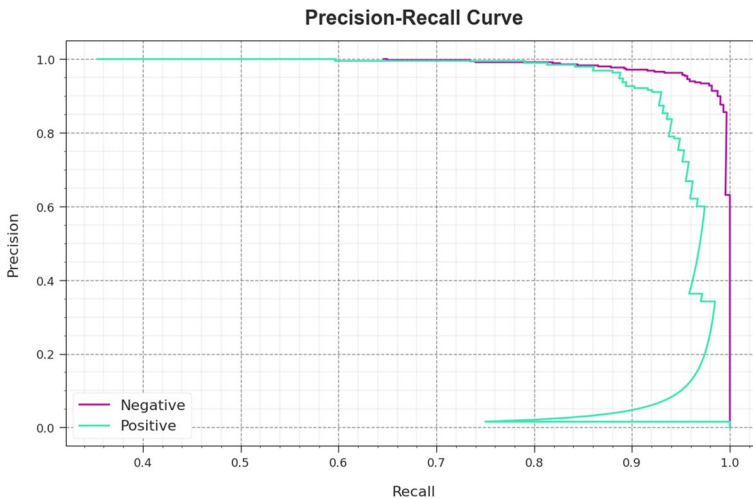
Precision-Recall Curve

Fig. 12 PR curve of BBESA-OFRBC system on 70:30 of TRP/TSP

Figure 10 inspects the accuracy of the BBESA-OFRBC method in the training and validation procedure on 70:30 of TRP/TSP. The result implied that the BBESA-OFRBC approach gains maximum accuracy values over enhanced epochs. Additionally, an improved validation accuracy over training accuracy reveals that the BBESA-OFRBC method learns effectively on 70:30 of TRP/TSP.

The loss investigation of the BBESA-OFRBC method at the time of training and validation is established on 70:30 of TRP/TSP in Fig. 11. The outcomes stated that the
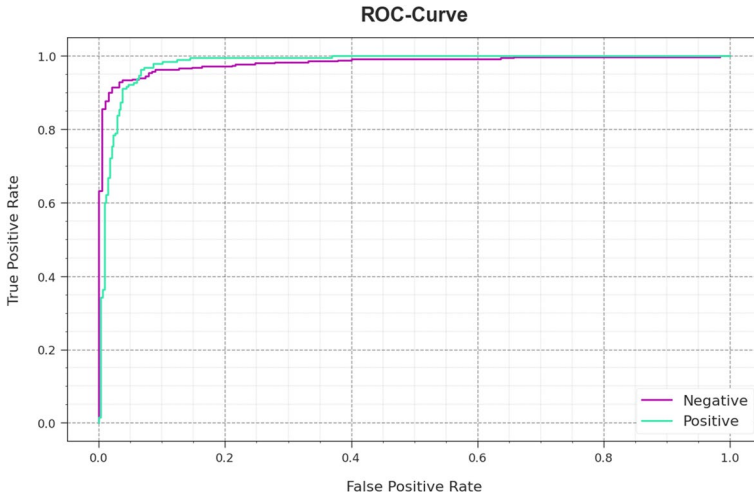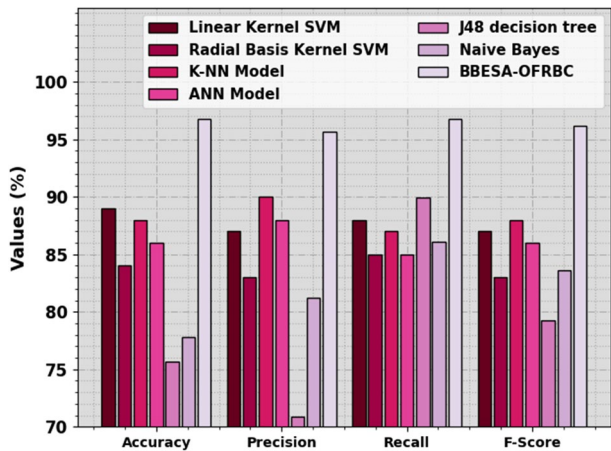
Fig. 13 ROC curve of BBESA-OFRBC system on 70:30 of TRP/TSP

Table 5 Comparative outcome of BBESA-OFRBC algorithm with other approaches

Predictive Models	Accuracy	Precision	Recall	F-Score
Linear Kernel SVM	89.00	87.00	88.00	87.00
Radial Basis Kernel SVM	84.00	83.00	85.00	83.00
K-NN Model	88.00	90.00	87.00	88.00
ANN Model	86.00	88.00	85.00	86.00
J48 decision tree	75.65	70.86	89.92	79.26
Naive Bayes	77.83	81.25	86.09	83.60
BBESA-OFRBC	96.76	95.67	96.76	96.18

Fig. 14 Comparative outcome of BBESA-OFRBC algorithm with other approaches

BBESA-OFRBC system obtains adjacent values of training and validation loss. Evidently, the BBESA-OFRBC method learns capably on 70:30 of TRP/TSP.

A detailed PR curve of the BBESA-OFRBC method is established on 70:30 of TRP/TSP in Fig. 12. The outcomes indicated that the BBESA-OFRBC methodology results in higher values of PR. Moreover, it is visible that the BBESA-OFRBC approach can gain superior PR values in all classes.

In Fig. 13, a ROC examination of the BBESA-OFRBC system is exposed at 70:30 of TRP/TSP. The figure defined that the BBESA-OFRBC approach resulted in improved ROC values. Besides, it is apparent that the BBESA-OFRBC method can extend maximal ROC values on all classes.

In Table 5 and Fig. 14, the outcomes of the BBESA-OFRBC system are compared with recent systems [24, 25]. The outcome inferred that the J48 decision tree, NB, Radial Basis Kernel SVM, and ANN algorithms depict minimal outcomes over other approaches.

Simultaneously, the K-NN method has managed to acquire moderate performance with $accu_y$, $prec_n$, $reca_l$, and F_{score} of 88%, 90%, 87%, and 88%, correspondingly. However, the Linear Kernel SVM approach demonstrated near optimum outcomes with $accu_y$, $prec_n$, $reca_l$, and F_{score} of 89%, 87%, 88%, and 87%, the BBESA-OFRBC methodology surpasses the other approaches with maximal $accu_y$, $prec_n$, $reca_l$, and F_{score} of 96.76%, 95.67%, 96.76%, and 96.18% correspondingly. Therefore, the BBESA-OFRBC technique can be employed for automated and accurate diabetes classification.

With the continuous rise in the count of deadly diseases that risk either human life or health, the medical Decision Support System keeps proving its efficiency in providing healthcare professionals and other physicians with support in making clinical decisions. Diabetes mellitus is assumed that chronic disease where the body doesn't produce the essential amount of insulin or insulin is not utilized well by the body, which leads to extremely higher glucose (blood sugar) levels. At the same time, once diabetes has been left untreated or undetected, it may cause severe harm to the body and makes it challenging to treat, but earlier diabetes diagnosis may result in better treatment, giving rise to lower death and morbidity. The Binary Bald Eagle Search Algorithm with Optimal Fuzzy Rule-based Classifier algorithm is meticulously designed to achieve highly effective diabetes detection and classification. It incorporates a Binary Bald Eagle Search Algorithm for optimal feature subset selection and employs the Fuzzy Rule-based Classifier for diabetes detection. To further enhance its performance, the algorithm utilizes sand cat swarm optimization to optimize the parameter values of the Fuzzy Rule-based Classifier. Extensive experimentation on benchmark diabetes medical datasets demonstrates the superiority of the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier approach over state-of-the-art models. The main objective of the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier algorithm is to accomplish effective detection and classification of diabetes. To achieve this, the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier technique primarily designs a Binary Bald Eagle Search Algorithm system for the optimal selection of feature subsets. Additionally, the detection of diabetes takes place using the Fuzzy Rule-based Classifier technique. Furthermore, the sand cat swarm optimization system was applied to optimise the Fuzzy Rule-based Classifier algorithm's parameter values. A wide range of experimental analyses is carried out on benchmark diabetes medical datasets, and the outcome was examined under many aspects. The experimental outcome portrayed the greater of the Binary Bald Eagle Search Algorithm—Optimal Fuzzy Rule-based Classifier approach over the state of art models.

5 Conclusion

In conclusion, the research has presented compelling statistical results that firmly establish the superiority of the BBESA-OFRBC technique in the domain of diabetes detection and classification. The outcomes, based on rigorous evaluation metrics, speak to the method's exceptional performance. Specifically, the BBESA-OFRBC methodology demonstrated a remarkable accuracy rate of 96.76%, indicating its ability to correctly classify diabetes cases with precision. Additionally, the precision score of 95.67% underscores the methodology's reliability in minimizing false positive results, a critical factor in medical diagnostics. Equally noteworthy is the recall rate of 96.76%, highlighting the approach's capacity to effectively identify true positive cases. The F-score, an essential metric that balances precision and recall, reached an impressive 96.18%, further cementing the BBESA-OFRBC technique's excellence. These results, when compared to alternative methods like K-Nearest Neighbors (K-NN) and Linear Kernel Support Vector Machine (SVM), clearly indicate the practical applicability and potential of the BBESA-OFRBC methodology for accurate and automated diabetes classification in real-world healthcare settings. Additionally, the use of the Sand Cat Swarm Optimization (SCSO) algorithm for parameter optimization within the Fuzzy Rule-Based Classifier (FRBC) contributed significantly to enhancing the overall detection rate, further underscoring the method's robustness and effectiveness. As the research suggests the exploration of deep ensemble voting-based classifier techniques for future enhancements, it paves the way for ongoing innovation in diabetes detection, solidifying the BBESA-OFRBC approach as a valuable asset in the realm of healthcare diagnostics.

Data availability The manuscript has no associated data.

Declarations

Human participants and/or animals Not applicable.

Conflict of interest The authors have expressed no conflict of interest.

References

1. Laila UE, Mahboob K, Khan AW, Khan F, Taekeun W (2022) An ensemble approach to predict early-stage diabetes risk using machine learning: An empirical study. Sensors 22(14):5247
2. Jaiswal V, Negi A, Pal T (2021) A review on current advances in machine learning based diabetes prediction. Prim Care Diabetes 15(3):435–443
3. Bhandari S, Pathak S, Jain SA (2023) A literature review of early-stage diabetic retinopathy detection using deep learning and evolutionary computing techniques. Arch Comput Methods Eng 30(2):799–810
4. Naveena S, Bharathi A (2022) A new design of diabetes detection and glucose level prediction using moth flame-based crow search deep learning. Biomed Signal Process Control 77:103748
5. Nadeem, MW, Goh, HG, Ponnusamy, V, Andonovic, I, Khan, MA, Hussain, M, (2021) A fusion-based machine learning approach for the prediction of the onset of diabetes. In Healthcare (Vol. 9, No. 10, p. 1393). MDPI. https://doi.org/10.3390/healthcare9101393

6. Karunakaran, D, Chandran RK (2023) Deep Learning Based Diabetes Mellitus Prediction for Healthcare Monitoring. J Electric Eng Technol :1–15. https://doi.org/10.1007/s42835-023-01500-4
7. Thaiyalnayaki K (2021) Classification of diabetes using deep learning and svm techniques. Int J Current Res Rev 13(01):146
8. Hasan, DA, Zeebaree, SR, Sadeeq, MA, Shukur, HM, Zebari, RR, Alkhayyat, AH (2021) Machine Learning-based Diabetic Retinopathy Early Detection and Classification Systems-A Survey. In 2021 1st Babylon International Conference on Information Technology and Science (BICITS) (pp. 16–21). IEEE
9. Yadav DC, Pal S (2021) An experimental study of diversity of diabetes disease features by bagging and boosting ensemble method with rule based machine learning classifier algorithms. SN Comput Sci 2(1):50
10. Haq AU, Li JP, Khan J, Memon MH, Nazir S, Ahmad S, Khan GA, Ali A (2020) Intelligent machine learning approach for effective recognition of diabetes in E-healthcare using clinical data. Sensors 20(9):2649
11. Meganathan S, Sumathi A, Bharanika V, Hemalakshmi P, Kamali M (2022) Finding best voting classifier for diabetic disease classification. In: International Conference on Deep Sciences for Computing and Communications. Springer Nature Switzerland, Cham, pp 25–33
12. Mishra S, Tripathy HK, Mallick PK, Bhoi AK, Barsocchi P (2020) EAGA-MLP—an enhanced and adaptive hybrid classification model for diabetes diagnosis. Sensors 20(14):4036
13. Nagaraj P, Deepalakshmi P (2022) An intelligent fuzzy inference rule-based expert recommendation system for predictive diabetes diagnosis. Int J Imaging Syst Technol 32(4):1373–1396
14. García-Ordás MT, Benavides C, Benítez-Andrades JA, Alaiz-Moretón H, García-Rodríguez I (2021) Diabetes detection using deep learning techniques with oversampling and feature augmentation. Comput Methods Programs Biomed 202:105968
15. Hao J, Luo S, Pan L (2022) Rule extraction from biased random forest and fuzzy support vector machine for early diagnosis of diabetes. Sci Rep 12(1):9858
16. Balasubramaniyan S, Jeyakumar V, Nachimuthu DS (2022) Panoramic tongue imaging and deep convolutional machine learning model for diabetes diagnosis in humans. Sci Rep 12(1):186
17. Khafaga, DS, Alharbi, AH, Mohamed, I. Hosny, KM (2022) An Integrated Classification and Association Rule Technique for Early-Stage Diabetes Risk Prediction. In Healthcare (Vol. 10, No. 10, p. 2070). MDPI. https://doi.org/10.3390/healthcare10102070
18. Aamir KM, Sarfraz L, Ramzan M, Bilal M, Shafi J, Attique M (2021) A fuzzy rule-based system for classification of diabetes. Sensors 21(23):8095
19. Butt, UM, Letchmunan, S, Ali, M, Hassan, FH, Baqir, A, Sherazi, HHR (2021) Machine learning based diabetes classification and prediction for healthcare applications. J Healthcare Eng 2021 https://doi.org/10.1155/2021/9930985
20. Chhabra A, Hussien AG, Hashim FA (2023) Improved bald eagle search algorithm for global optimization and feature selection. Alex Eng J 68:141–180
21. Stavrakoudis DG, Galidaki GN, Gitas IZ, Theocharis JB (2011) A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery. IEEE Trans Geosci Remote Sens 50(1):130–148
22. Wang X, Liu Q, Zhang L (2023) An Adaptive Sand Cat Swarm Algorithm Based on Cauchy Mutation and Optimal Neighborhood Disturbance Strategy. Biomimetics 8(2):191
23. https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database
24. Kaur H, Kumari V (2022) Predictive modelling and analytics for diabetes using a machine learning approach. Appl Comput Inf 18(1/2):90–100
25. Chang, V, Bailey, J, Xu, QA, Sun, Z (2022) Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. Neural Comput Appl, pp.1–17. https://doi.org/10.1007/s00521-022-07049-z

Authors and Affiliations

R. Karthikeyan[1] · P. Geetha[2] · E. Ramaraj[1]

✉ R. Karthikeyan
   karthikeyancs@srmasc.ac.in

1   Department of Computer Science, Alagappa University, Karaikudi, India 630003

2   PG Department of Computer Science, Dr. Umayal Ramanathan College for Women, Karaikudi,
    India 630003