PATIENT FACING SYSTEMS

Enhanced Monarchy Butterfly Optimization Technique for effective breast cancer diagnosis

S. Punitha[1] · A. Amuthan[2] · K. Suresh Joseph[1]

Abstract

Breast cancer is the biggest curse for the women society in the world since the survival factor of the infected patients is ensured only when it is detected at the early localized stage. The majority of the intelligent schemes proposed for detecting the breast cancer relies on the human skill that helps in trustworthy determination of essential pattern that confirms the existence of the infected cancer cells for deciding upon the course of treatment. Further, most of the research works contributed in the literature for detecting breast cancer necessitates huge time and laborinvolved that increases the time of diagnosis. This Intelligent Artificial Bee Colony and Enhanced Monarchy Butterfly Optimization Technique (IABC-EMBOT) is proposed for effective breast cancer diagnosis. The core idea behind the formulation of IABC-EMBOT relies on two significant ameliorations that, i) focuses on the modification of Monarchy Butterfly Optimization that enhances the exploration degree based on the rate of exploitation of the searching space and ii) concentrates on the elimination in the limitations of the ABC scheme by enhancing the possibility of search diversification process through phenomenal update facilitated through the dynamic and adaptive butterfly operator that improves the search globally. The proposed IABC-EMBOT scheme investigated using the Wisconsin data set is proven to facilitate an improved average classification accuracy of 97.53%.

Keywords Monarchy butterfly optimization · Artificial bee colony · Adaptive butterfly operator · Wisconsin data set

Introduction

Breast cancer is considered as the most dangerous kind of cancer most developed in the region of the women breast that has emerged as the most crucial disease in the recent decades. The mortality rate of the women population against breast cancer needs to be greatly minimized to a maximum level through the screening process in regular interval of time [1].

✉ S. Punitha
  punitharesearch@gmail.com

  A. Amuthan
  amuthan@pec.edu

  K. Suresh Joseph
  ksjoseph.csc@gmail.com

[1] Department of Computer Science, Pondicherry University, Pondicherry, India

[2] Department of Computer Science Engineering, Pondicherry Engineering College, Pondicherry, India

The breast cancer needs to be predominantly diagnosed since this kind of cancer has the major threat that increases the infection probability of various neighboring vital organs of the body [2]. Thus, this aggressive nature of breast cancer need to be detected for reducing the mortality rate of women population of the globe.Generally, the pathologist is responsible for optimal detection of breast cancer [3]. This screening process of breast cancer requires human skill such that the early detection process is accurate and reliable in nature. In contrast, a number of recent computer assisted cancer detection approaches wide opens the probability of rapid detection with improved accuracy and reliability. Thus intelligent and automation diagnosis of breast cancer become essential for enhancing the prevention rate of mortality among the women population [4].

A number of intelligent and automation assisted breast cancer diagnosis approaches are contributed in the literature with the aim to reduce the degree of human intervention [5]. Intelligent breast cancer diagnosis approaches are replaced over the manual diagnostic approaches in order to reduce the time of initial screening in the localization stage [6]. This minimization of diagnosis time involved in the detection of

breast cancer at the premature localized stage aids in appropriate and reliable diagnosis [7, 31]. The ANN-based breast cancer diagnostic approaches are determined to be mainly beneficial in exploring multiple dimensions of dependent parameters that influence the process of decision-making [8]. These ANN-based breast cancer diagnostic approaches are also capable of ensuring accurate approximation since it is considered as the essential constraint in the process of resolving nonlinear and complex factors that influence the reliable training process involved in the detection [9]. Furthermore, ANN-aided breast cancer diagnostic approaches are determined to be effective and efficient in detection, when five influencing parameters such as, i) number of hidden layers, ii) number of hidden layers in each individual layer, iii) scheme of feature selection, iv) employed training algorithm and v) assignment of significant weights. The scheme of feature selection is highly influential during the design of any successful breast cancer diagnosis [10]. The selection of feature subset is also considered to be the vital in the design of any significant breast cancer diagnosis approach [21]. It is also inferred that the classification accuracy of a successful breast cancer diagnostic will be maximized only when the input feature subset and design factors are determined effectively. In addition, the employment of meta-heuristic algorithms like Artificial Bee Colony, Ant colony etc. during the process of optimizing the input feature subset and design factors are confirmed to improve the reliability in the process of breast cancer detection [22, 23, 32].

The motivation behind the formulation of the proposed IABC-EMBOT scheme lies in the need for hybridizing the two meta-heuristic approaches for effective feature selection process. This motivation concentrates on the optimization of hidden node size and initial weight employed during the process of feature optimization that attribute towards improved performance of the multilayer perceptron network. Further, the process of hybridizing two meta-heuristic approaches are considered to be potential in order to determine the existence of specific pattern derived from the database. This process of hybridizing two meta-heuristic approaches eliminates the degree of time consumption, degree of human error and intensification in labor. This process of hybridizing two meta-heuristic approaches also focus on the process of parameter optimization and simultaneous feature selection that attribute towards maximum detection of breast cancer.

In this paper, An Intelligent Artificial Bee Colony and Enhanced Monarchy Butterfly Optimization Technique (IABC-EMBOT) is contributed to facilitate effective breast cancer diagnosis. This proposed IABC-EMBOT concentrates on the enhancement of the Monarchy Butterfly Optimization through the improvement of exploration degree and exploitation level imposed over the searching space. This proposed IABC-EMBOT focuses on the prevention of drawbacks that are greatly involved during the process of optimizing feature

subset and design parameters through the incorporation of the traditional Artificial Bee Colony (ABC) optimization mechanism. This scheme significantly improves the rate of diversification by utilizing a global and adaptive butterfly operator with the view to improve the process of global search.

The main contributions of the proposed scheme as listed as follows:

a) The proposed IABC-EMBOT scheme is developed as an automatic ABC and MBO-based hidden node size optimization approach of the incorporated MLP in the process of the breast cancer diagnosis.

b) The proposed IABC-EMBOT scheme embedded the process of initial weight optimization for reducing the process of minimizing the possibility of being getting trapped into a local minimum.

c) The proposed scheme focuses on the evaluation of the impacts of the feature selection in the size of the hidden nodes of MLP with respect to complexity and accuracy.

The section "Previous breast cancer related work" deals with the related works that are existing in the literature. Section "Proposed-intelligent artificial bee colony and enhanced monarchy butterfly optimization technique (IABC-EMBOT)" explains the Proposed-Intelligent Artificial Bee Colony and Enhanced Monarchy Butterfly Optimization Technique (IABC-EMBOT). Section "Results and discussions" deals about the results and discussions. Section "Conclusion" concludes the paper.

Previous breast cancer related work

An integrated breast cancer detection scheme using Particle Swarm Optimization and Finite Differences (PSO-FD-BCD) was proposed for reconstructing the dimensions of the breast cancer cell with the view to estimate its position [11]. This PSO-FD-BCD scheme utilized two and three dimensional breast models over which the concept of finite differences is employed for estimating potential features that attribute towards breast cancer detection. The classification accuracy, precision and recall value of the PSO-FD-BCD schemes were determined to be the maximum up to a level of 99.06% and 98.42% respectively. Then an integrated Particle Swarm Optimization and Support Vector Machine (PSO-SVM-BCD) was proposed using the machine learning method for classifying and investigating breast cancer data [12]. This PSO-SVM-BCD scheme utilized the merits of SVM for reducing the degree of error in generalizations to a predominant level. This PSO-SVM-BCD scheme used PSO as the optimal technique for automatic estimation of algorithmic factors that aids in potent detection of breast cancer. This proposed PSO-SVM-BCD scheme was estimated to resolve the issues of

recognition for enhancing the rate of classification accuracy. A Sequenced Genetic Algorithm using SVM (SGA-SVM-BCD) was proposed for breast cancer diagnosis scheme in order to enhance the rate of classification accuracy and precision [13]. This approach prevented the risk of premature convergence of the optimal optimization process into its local optimal point that reduces the quality of the solutions. This SGA-SVM-BCD scheme reduced the training time, the percentage decrease in classification accuracy and sensitivity to a predominant level of 75.77%, 0.42% and 1.65% compared to the PSO-FD-BCD and PSO-SVM-BCD approaches. Further, a Genetic Algorithm-based automated breast cancer detection scheme (GA-ABCDS) was proposed using the merits of parallel parameter optimization and feature selection [14]. The GA-ABCDS scheme was implemented with three diversified back-propagation dimensions of resilient, gradient descent and Levenberg-Marquardt in order to fine tune the priority weights of ANN for optimal performance. This GA-ABCDS scheme also enhanced the mean and best classification accuracy to a maximum threshold of 98.21% and 99.24% under the investigation with the Wisconsin breast cancer dataset. A Feed Forward Neural Network-based Breast Cancer Detection (FFNN-BCD) scheme was contributed using the Multi-Layer Perceptron (MLP) for achieving an optimal cancer diagnosis rate [15]. The MLP process used in this FFNN-BCD scheme aided in assigning influencing weights related to the GA routine for improving the precision and recall value. The mean processing time incurred in this FFNN-BCD scheme was determined to be highly minimized to a maximum level of 3.87 s.

Furthermore, Deep Belief Network-based Breast Cancer Detection (DBN-BCD) scheme was proposed based on Lievenberg-Marquardt back-propagation [16]. The weight in this DBN-BCD scheme is initialized through the assignment of a deep belief network path that concentrates on maximum optimization of the selected features. The classification accuracy of approximately 99.68% is ensured in this DBN-BCD scheme since it is capable of examining the diversified factors of cancer detection together. Then, an automated assisted Deep Neural Networks-based Breast Cancer Detection (DNN-BCD) scheme was proposed for resolving the issues that focus on the elimination of recursive and classifier characteristics involved in the process of feature selection [17]. The hyper spectral method used in the DNN-BCD scheme facilitated higher learning rate by elucidating maximum learning deep features from the cancer cells. This DNN-BCD scheme confirmed a classification accuracy of 98.62% compared to the FFNN-BCD and GA-ABCDS approaches. An Artificial Bee Colony Optimization-based Breast Cancer Detection (ABCO-BCD) scheme was propounded for exploring and exploiting the feature subset used for diagnosis [18]. This ABCO-BCD

approach explored all the probabilities of combining the feature subset of cancer data, such that maximum precision, recall and mean processing time is sustained. This ABCO-BCD approach improved the classification accuracy to a superior level of 6.72% compared to the DBN-BCD and DNN-BCD schemes of the literature. A Invariant Hu Moment and Feed Forward Neural Network (IHM-FFNN) scheme was proposed for effective detection of breast cancer [19]. This IHM-FFNN scheme was confirmed to improve the sample accuracy during the enforcement of k-fold-based cross validating testing process. The classification accuracy of this IHM-FFNN scheme was estimated to be nearly 97.32% since the merits of the embedded Hu Moment aided in classifying normal cells form the cancer infected cells. Finally, a Particle Swarm Optimization with Recurrence Model (PSO-RM) was contributed for improving the rapid rate of classifying cancer infected cells [20]. This PSO-RM scheme utilized three classifiers such as a fast decision tree, naïve Bayes and k-nearest neighbor classifier for discriminating normal cells with cancer infected cells. This PSO-RM scheme ensured a classification accuracy rate of 98.13% over the compared IHM-FFNN detection approach. An integrated static classifier and random space-based computer aided breast cancer detection scheme was proposed for effective classification of malignant tumor with benign tumor [27]. This integrated static classifier and random space-based computer aided breast cancer detection scheme utilized the method of diversity for inheriting the process of constructing feature collection and its feature selection from the extracted classifier pool of features derived from the data set. This static classifier used the benefits of classifier ensemble for determining accurate diversity between the existing features in the dataset. Then, a Semi-Supervised Learning-based breast cancer detection scheme was proposed with diversified kernel functions and semi supervised support vector machine [28]. This Semi-Supervised Learning-based breast cancer detection scheme was incorporated for classifying the labeled data through the process of training. This Semi-Supervised Learning-based breast cancer detection scheme was contributed for validating the dataset that comprised of digital database used for screening mammography images. Then, a back propagation neural network-based training method for automated breast cancer detection was contributed for exploring the benign and malignant tumors prevalent in the WBCD dataset was proposed [29]. This back propagation neural network-based training method inherited nine significant features for classifying the benign and malignant tumors with an excellent accuracy rate of 99.27%. This back propagation neural network-based training method proved a sensitivity of 98.21%, specificity of 99.11% and a negative predictor

rate of 98.43% under an increasing number of features considered for detection. A S3VM-based breast cancer detection scheme was also proposed for validating the dataset that comprised of digital database used for screening mammography images [30]. This S3VM-based breast cancer detection schemewas inherited from the process of classifying the labeled data in the training process. This S3VM-based breast cancer detection proved a sensitivity of 98.12%, specificity of 99.02% and a negative predictor rate of 98.21% under an increasing number of features considered for detection.

The need for the formulation of the proposed IABC-EMBOT

The primitive ABC algorithm possesses the potential in exploring the search space in an effective manner with increased significance in determining the local optima by enforcing employee and onlooker bee phases. Thus, ABC algorithm is highly suitable for the selection of predominant solutions that enhance the rate of local search. However, the global searching process incorporated by the scout bee phase of ABC is responsible for reduced convergence speed during its implementation. Similarly, MBO algorithm is potentially significant to ABC in exploring the search space effectively. But, they are not capable enough in the exploitation of search space resulted due to the contextual utilization of Levy lights that update operators leading to random moves or steps. In this context, it is clear that ABC possesses the limitation of reduced convergence speed and MBO embeds the issue of poor exploitation in the search phase. However, the core objective of meta-heuristic optimization schemes need to good balance between the degree of exploration and exploitation is the search space of solutions. Hence, a hybrid ABC and MBO-based meta-heuristic scheme was proposed with two modifications. The first modification of the ABC is the inheriting process of modifying the butterfly operator of MBO into its employee phase. Likewise, the butterfly adjusting operator of the MBO scheme is modified for enhancing the degree of exploitation with the exploration degree with the view to increase the diversity of reaching and preventing the limitations of the ABC during its enforcement in the global searching process.

Further, the hybridization of ABC and MBO aided in the superior enforcement of eliminating the trade-off between the exploitation and exploration, which is essential in improving the rate of solving high and low dimensional problems. Hence, this hybrid ABC and MBO method of feature optimizations plays a vital role in accurate detection of breast cancer cells. In addition, the exploitation process of hybrid ABC and MBO method is capable of seeking superior solutions by utilizing the benefits of existing knowledge with exploration

phase for complete search of an optimal solution for the problem space.

Proposed-intelligent artificial bee colony and enhanced monarchy butterfly optimization technique (IABC-EMBOT)

The proposed Intelligent Artificial Bee Colony and Enhanced Monarchy Butterfly Optimization Technique (IABC-EMBOT) adopts the wrapper method of implementation for facilitating simplicity and effectiveness during the process of optimizing the number of the hidden nodes in the MLP network. This wrapper approach of implementation is mainly utilized in IABC-EMBOT for eliminating the overheads that incur during the process of initial statistical processing performed over the data set considered for optimization. Moreover, the diagnosis time enabled by IABC-EMBOT is also predominantly reduced since it eliminates the time involved in the process of preventing inconsistent data. This is contrasted with the recent existing methods of the literature. Hence, this proposed IABC-EMBOT is considered to the best approach for superior optimization of the parameters pertaining to the MLP network as it derives the merits of ABC and EMBO. In this proposed IABC-EMBOT, the drawbacks of the traditional process of ABC optimization in the global search process are improved by enabling multi-perspective search diversification strategy through the incorporation of the dynamic adjustment of the flexible butterfly operator. This proposed hybrid approach also improves the rate of exploration through the adaptation of exploitation rate by utilizing locally. The comprehensive architecture of proposed IABC-EMBOT approach is portrayed in Fig. 1.

The subsequent section of the proposed IABC-EMBOT approach highlights the five steps involved in its implementation such as i) Representation of solutions and fitness function estimation, ii) Improved Monarchy Butterfly Optimization-based employee bee phase, iii) Onlooker bee phase,iv) scout bee phase and v) MLP network optimization process.

Representation of solutions and fitness function estimation

In the proposed IABC-EMBOT approach, each and every solution in the process of optimization pertains with the input features that estimate the input node count of the MultiLayerPerceptron (MLP) network. In this context, the initial solution relates to the possible collection of feature subset that has the potential in influencing the performance of the MLP network is generated with input node features and hidden node count. The initial solutions (collection of feature subset) may or may not necessarily facilitate the process of optimization that results in impacting the performance of the
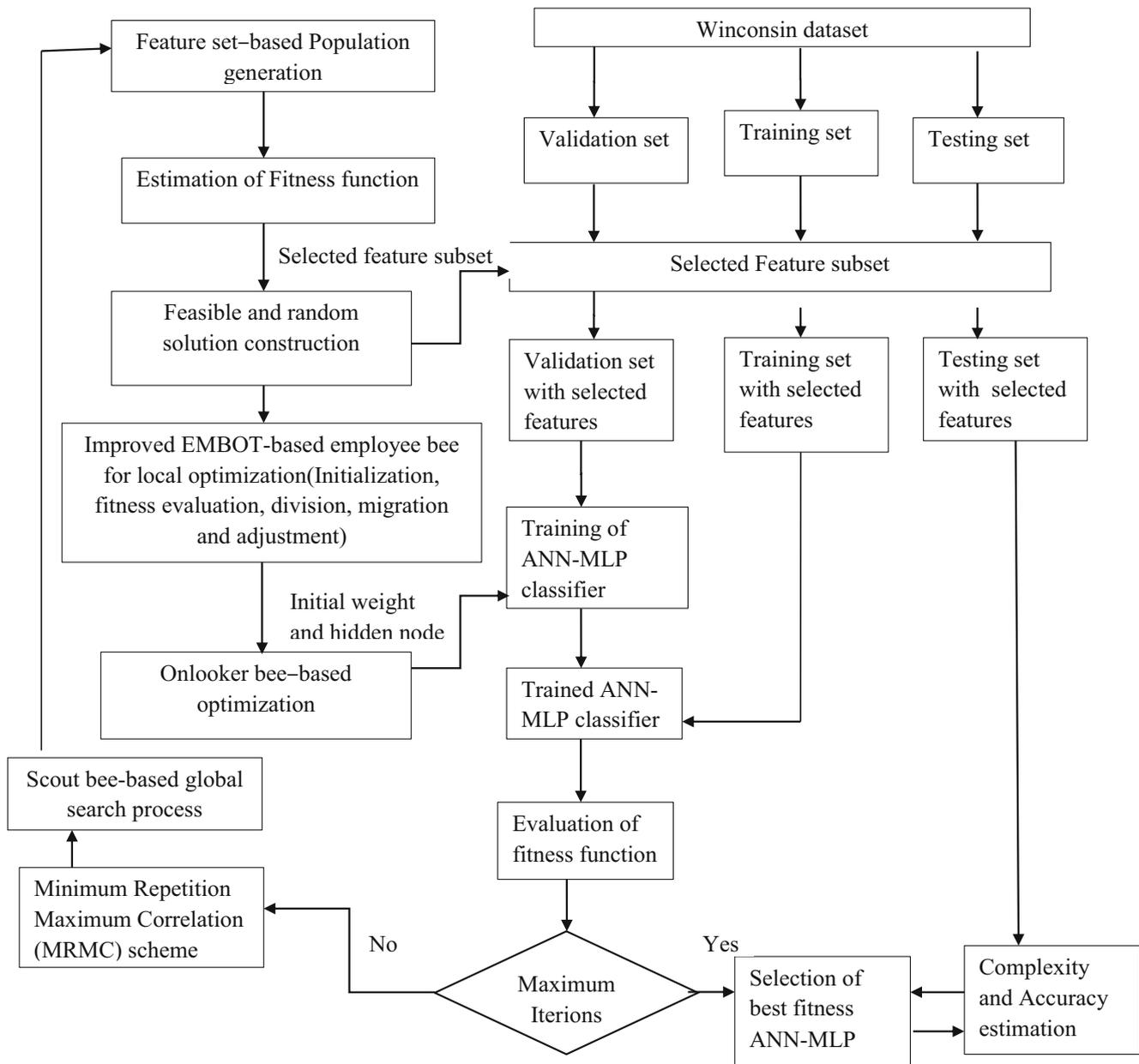
Fig. 1 Architecture of the proposed IABC-EMBOT Scheme

MLP network in the training and testing process of the utilized data set. Hence, the initial solutions need to be checked for verifying its predominant potential in optimizing the hidden node count of the MLP network by enforcing capable constraints that need to be satisfied under optimization. Thus, the function of fitness ($F(fit_{OPT})$) that verifies the potential of the initial solutions towards optimization based on imposed constraints are determined using Eq. (1).

$$F(fit_{OPT}) = \gamma \text{Cos}t(n) + \beta P_{H-I} + \delta P_{I-W} \qquad (1)$$

Where γ, β and δ relates to the adaptability coefficients utilized in the process of the ABC algorithm with $\text{Cos}t(n)$ as

the cost incurred during the process of optimizing MLP network under the influence of the hidden node count (P_{H-I}) and utilized initial weight(P_{I-W}).

Further, the adaptability coefficient 'γ' determined through the impact of average number of hidden nodes (P_{H-I}) that satisfies or dissatisfies the derived initial solutions and identifies them into feasible optimal solution is computed and updated using Eq. (2)

$$\gamma_{UP} = \frac{\gamma_{INITIAL}}{(1 + \beta)} \qquad (2)$$

Similarly, the other adaptability coefficient 'δ' estimated based on the influence of the assigned mean optimal initial

Fig. 2 Proposed IABC-EMBOT evaluated based on classification Accuracy
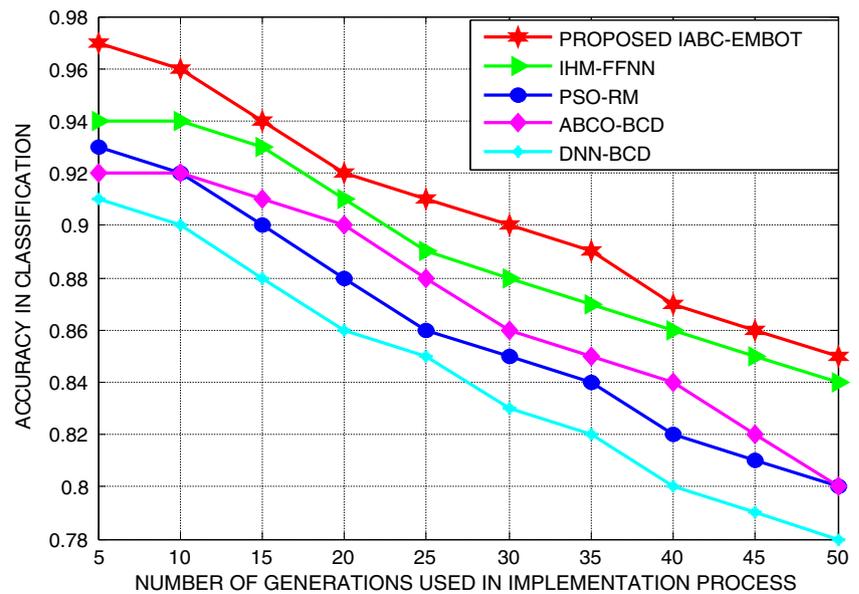
priority (P_{I-W}) that concludes a particular solution into possible optimal solution is identified and updated using Eq. (3)

$$\delta_{UP} = \frac{\delta_{INITIAL}}{(1 + \beta)} \qquad (3)$$

Hence, the estimated function of fitness is improved phenomenally in order to optimize the search solution in a higher order search domain in order to achieve a reduced number of hidden node count and enhance the classification accuracy in a precise way.

In this proposed IABC-EMBOT approach, the method of Correlation-oriented Feature Selection is used as the preprocessing method that aids in deriving possible initial solutions of the search domain. The Correlation-

oriented Feature Selection is capable of ranking the subset of features in addition to the fitness function through the estimation of the evaluation function derived using Eq. (4) in a heuristic manner.

$$OPT_{rank} = \frac{dC_{MC(Features)}}{\sqrt{d + d(d-1)C_{MC(Features)}}} \qquad (4)$$

The data set used in the proposed IABC-EMBOT approach for detecting breast cancer is very large and hence the incorporation of the Correlation-oriented Feature Selection method can be beneficial in deriving the advantages of the individual features for forecasting the possibility of cancer cell detection. This correlation mechanism uses the method of maximal similarity among the classes and minimal similarity between the

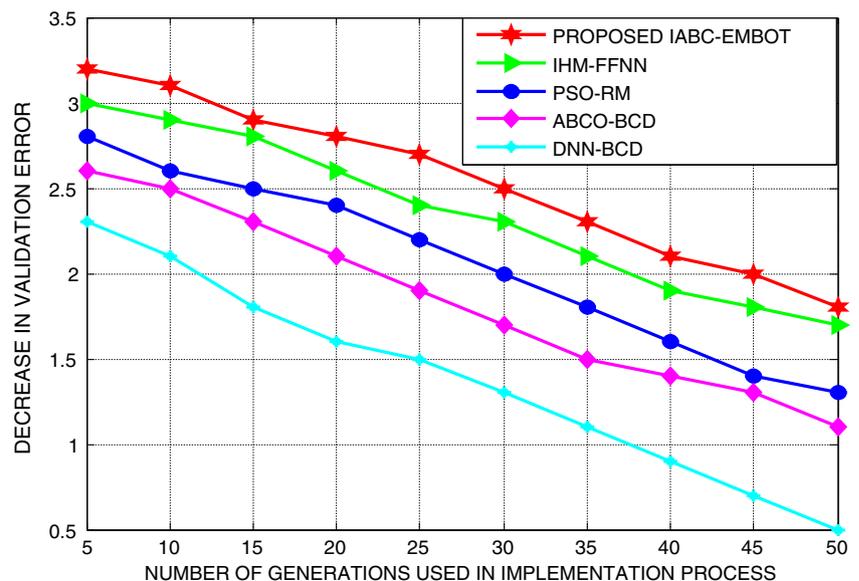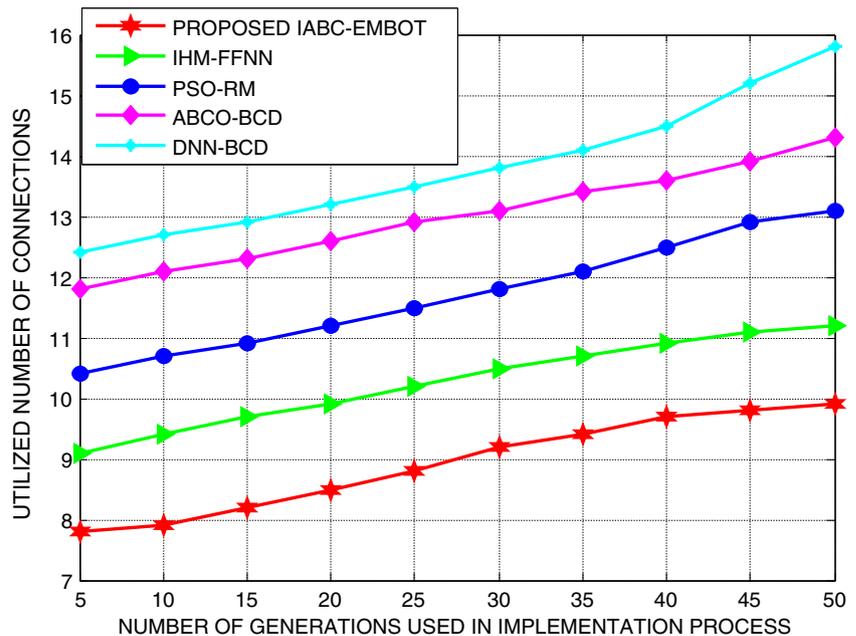**Fig. 3** Proposed IABC-EMBOT evaluated based on validation error

Fig. 4 Proposed IABC-EMBOT evaluated based on numbers of connections

features considered for breast cancer detection. This correlation mechanism also manipulates two matrices that represent feature to feature and feature to class correlation over the data used for training process. This method of correlation used for preprocessing initially starts from the feature empty set and then the feature subsets with the highest rank is selected as the initial solutions.

Improved monarchy butterfly optimization-based employee bee phase

The employee bees are responsible for exploiting the possible initial solutions at x_i with the view to estimate the best solution

index at each of the new positions p_i. The updated best solution index on each of the new position is computed by the employee bee based on Eq. (5)

$$p_{ij} = x_{ij} + gr_{ij}(x_{ij} - x_{kj}) \qquad (5)$$

Where $p_i = [p_{i1}, p_{i2}, \ldots\ldots\ldots, p_{im}]$ and $x_i = [x_1, x_2, \ldots., x_{im}]$ is the updated best solution index and the predecessor best solution index with the right random number 'k' meeting the constraints $k \neq j$. In this context, the right random number is considered to range from 1 to N_{AB} (the number of artificial bees that represent the solution to this optimization problem). Further, gr_{ij} is the randomly distributed number that ranges

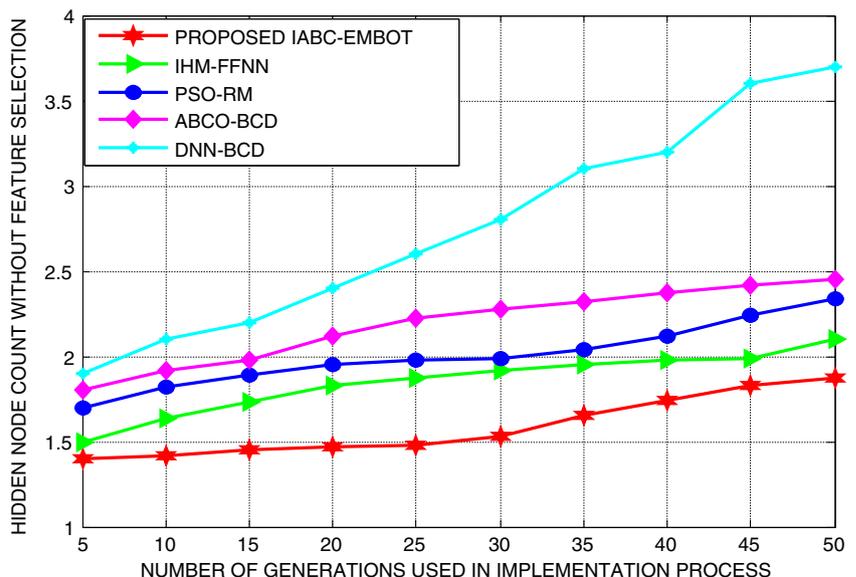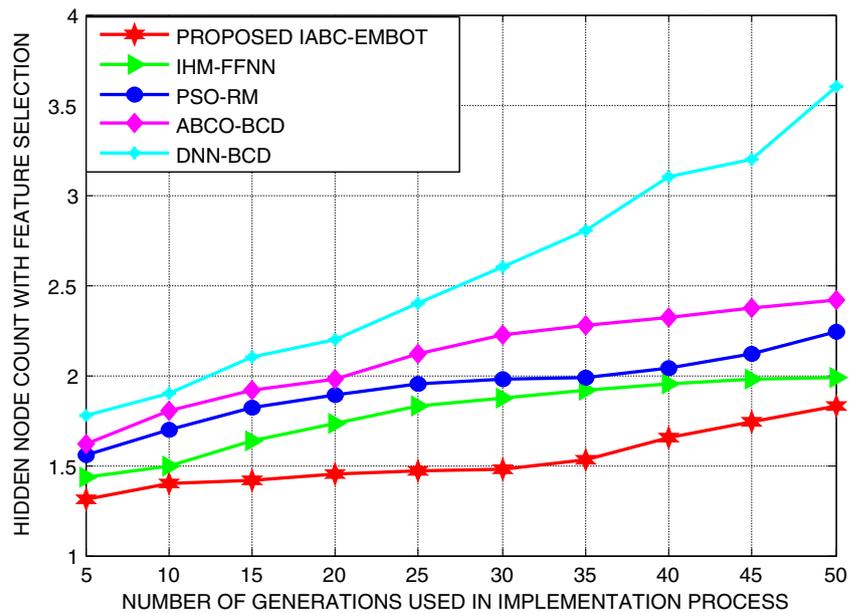**Fig. 5** Proposed IABC-EMBOT-Hidden node count with feature selection

Fig. 6 Proposed IABC-EMBOT-Hidden node count without feature selection

between −1 and 1. Furthermore, the selection in the random best solution index of the Winconsin dataset is facilitated based on Eq. (6)

$$x_{ij} = L_t + rand(0, 1)*(U_t - L_t) \qquad (6)$$

Where L_t and U_t refer to the upper and lower limit of variable x_i. In this proposed approach, the value of L_t and U_t are assigned to 0 and $(Maximum\_bestsolution\_index) - 1)$ respectively. While $rand(0, 1)$ pertains to the random numbers selected between the range 0 and 1. In addition, if the updated best solution index is estimated then the process of optimization must be initiated using the fitness function that is represented using Eq. (1). The fitness value in this proposed IABC-EMBOT approach is estimated based on the degree of classification accuracy as determined by the utilized MLP classifier. If the current estimated fitness value is greater than the preceding fitness value, then the old fitness value is discarded. But, this exploitation of the feasible solution needs to be

improved since they are prone to delays convergence and has the possibility of trapping into a local optimum point and hence Improved Monarchy Butterfly Optimization-based employee bee adjusting operator is incorporated for preventing the optimizing process to fall into a local optimal point of convergence.

This improved Monarchy Butterfly Optimization-based employee bee phase of IABC-EMBOT consists of four processes that include initialization, fitness evaluations, division, migration and adjustment. The initialization and fitness evaluations of the improved Monarchy Butterfly Optimization-based employee bee phase of IABC-EMBOT follows the same process involved in the employee bee phase of the primitive ABC optimization algorithm. In the third process of EMBOT division, the initial solutions that constitute the entire search process is partitioned into two sub populations of the search domain based on the assignment of two vital factors L_1 and L_2 respectively. This potential factor L_1 and L_2 in turn depends on the pre-assigned factor 'p'. In the fourth step, a

Table 1 Proposed IABC-EMBOT Scheme evaluated using classification accuracy and number of connections under a different number of generations

Number of threshold generation	Proposed IABC-EMBOT Scheme				IHM-FFNN Scheme				ABCO-BCD Scheme			
	Classification Accuracy		Numbers of Connections		Classification Accuracy		Numbers of Connections		Classification Accuracy		Numbers of Connection	
	Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean	Best	Mean
10	97.53	97.21	7.31	7.82	94.21	94.45	9.21	9.08	93.21	93.11	11.3	11.1
20	97.32	97.27	7.34	7.92	94.34	94.32	9.32	9.21	93.42	93.22	11.9	11.6
30	97.82	97.32	7.42	7.88	94.76	94.54	9.41	9.23	93.63	93.52	12.3	12.1

Table 2 Proposed IABC-EMBOT Scheme-Confusion Matrix under 10 generations

Different schemes used for comparison	Categories of test cases	Actual number of cases	Predicted test results	
			Benign	Malignant
Proposed IABC-EMBOT scheme	Benign	1200	1115(TN)	358(FP)
	Malignant	800	35(FN)	765(TP)
IHM-FFNN Scheme	Benign	1200	1145(TN)	55(FP)
	Malignant	800	40(FN)	760(TP)
ABCO-BCD Scheme	Benign	1200	1130(TN)	70 (FP)
	Malignant	800	30(FN)	770(TP)

migration operator is used for constructing the initial portion of the new population based on the number of butterflies generated randomly in the sub population determined through L_1 and L_2 factors respectively. Then the size of the possibly generated sub population is initialized by the size of the L_1 vector. Further, each solution in the newly generated first portion of sub population, which represents the best solution index 'i' of each individual solution 'k' is determined using Eq. (7) and (8).

$$x_{i,k}^{t+1} = x_{rn1,k}^t \quad where \quad rn \leq p \tag{7}$$

$$x_{i,k}^{t+1} = x_{rn2,k}^t \quad where \quad rn > p \tag{8}$$

Where $x_{rn1,k}^t$ and $x_{rn2,k}^t$ are two random solutions determined based on factors L_1 and L_2 on any estimated rounds 't'. In this context, rn is the random number, analogous to the gr_{ij} parameter used in Eq. (4) is estimated based on migration period ($mig\_period = 1.2$ as defined in [24])using Eq. (9)

$$rn = rand(0,1)*mig\_period \tag{9}$$

Where the value of mig_period is derived through Eq. (10) based on estimated $x_{i,j}^t$ defined in Eq. (7) and Eq. (8)

$$mig\_period = x_{i,j}^t*rand(0,1) + 0.5 \tag{10}$$

Furthermore, the second portion of the sub population is constructed using the factor L_2based on the incorporation of

an EMBOT adjustment operator. This utilized EMBOT adjustment operator is responsible for generating the second portion of the sub population based on random solutions and best solutions derived from the factor L_2. The individual solution in the newly generated second sub population $x_{i,k}^{t+1}$, which represents the best solution index 'i' of each individual solution 'k' is determined using Eq. (11) and (12)

$$x_{i,k}^{t+1} = x_{best,k}^t \quad where \quad rand(0,1) \leq p \tag{11}$$

$$x_{i,k}^{t+1} = x_{rn3,k}^t \quad where \quad rand(0,1) > p \tag{12}$$

In addition, the elements in the second portion of the sub population are further enhanced based on Eq. (13)

$$x_{i,k}^{t+1} = x_{i,k}^{t+1} + S_F(\alpha)(df_x - 0.5) \tag{13}$$

In this context, if the butterfly adjustment operator rate is greater than randomly generated number, then the adjustment operator rate is set to a value named F_{BAR} that thereafter acts as the partition constant in this process. In Eq. (12), $S_F(\alpha)$ acts as the step factor that influences the local process of exploitation enabled by Levy Flight approach derived using Eq. (14)

$$df_x = Levy\left(x_j^t\right) \tag{14}$$

Finally, the first and second portion of sub population is integrated into a new population using an adjustment and

Table 3 Proposed IABC-EMBOT Scheme-Confusion Matrix under 20 generations

Different schemes used for comparison	Categories of test cases	Actual number of cases	Predicted test results	
			Benign	Malignant
Proposed IABC-EMBOT scheme	Benign	1200	1125(TN)	75 (FP)
	Malignant	800	50(FN)	750(TP)
IHM-FFNN Scheme	Benign	1200	1120(TN)	80 (FP)
	Malignant	800	60 (FN)	740(TP)
ABCO-BCD Scheme	Benign	1200	1145 (TN)	55 (FP)
	Malignant	800	25(FN)	775(TP)

Table 4   Proposed IABC-EMBOT Scheme-Confusion Matrix under 30 generations

Different schemes used for comparison	Categories of test cases	Actual number of cases	Predicted test results	
			Benign	Malignant
Proposed IABC-EMBOT scheme	Benign	1200	1140(TN)	60 (FP)
	Malignant	800	60(FN)	740(TP)
IHM-FFNN Scheme	Benign	1200	1120(TN)	80 (FP)
	Malignant	800	65(FN)	735(TP)
ABCO-BCD Scheme	Benign	1200	1135(TN)	65 (FP)
	Malignant	800	55(FN)	745(TP)

migration operator based on Eq. (15) and (16).

$$x_{i,j}^{curr} = x_{i,j}^{prev}$$
$$+ 0.5*rand(0,1)*\left(\left(x_{worst,j}^{prev} - x_{rn2}^{prev} \right) + \left(x_{rn2}^{prev} - x_{best,j}^{prev} \right) \right)$$
$$(15)$$

$$x_{i,j}^{curr} = x_{i,j}^{prev}$$
$$+ 0.5*rand(0,1)*\left(\left(x_{best,j}^{prev} - x_{rn3}^{prev} \right) + \left(x_{rn3}^{prev} - x_{worst,j}^{prev} \right) \right)$$
$$(16)$$

This process of optimization is iterated until a satisfactory solution is identified or until a fixed number of iterations is completed.

Onlooker bee phase of IABC-EMBOT

In this phase, the information related to the best solutions of the search space is shared by the employee bee to the onlooker bee when they complete the objective of exploiting solution. Then the posed IABC-EMBOT scheme the optimal solutions based on the probability of fitness which is analogous to the roulette wheel selection phase of GA. The aforementioned probability of fitness is manipulated by the onlooker bee based on Eq. (17)

$$Fit_{prob} = \frac{F\left(fit_{OPT(i)} \right)}{\sum\limits_{j=1}^{N_{AB}} F\left(fit_{OPT(i)} \right)} \qquad (17)$$

Scout bee phase of IABC-EMBOT

In this phase, employee bee or the onlooker bee is responsible for searching the best optimal solutions for predefined number and restricted number of iterations. The bee acting as an employee bee or the onlooker bee gets transformed into a scout bee during the process of exploration when the fitness probability used for estimating primitive solutions does not exhibit any predominant enhancement. In other words, the initial possible solution that does not guarantee any possibility of improvement over a limited number of iterations is explored by the scout bee. The scout bee chooses an index of solutions from the Winconsin dataset (search space) in a random manner. Further, the proposed IABC-EMBOT algorithm is realized to suffer from some crucial limitations that impact the speed and computation efficiency. Especially, the implementation of proposed IABC-EMBOT algorithm is also crucial when the aforementioned algorithm is enforced over the higher dimensional Winconsin dataset for estimates optimal solutions from the initial possible number of solutions that are feasibly derived during the initial phase. This determined drawback formed the induction behind the essentiality for enhancing the performance of the proposed IABC-EMBOT algorithm towards its hybridization with Minimum Repetition Maximum Correlation (MRMC) scheme. This hybridization of the proposed IABC-EMBOT algorithm with MRMC scheme is proved to be successful in determining the solutions that possess reduced repetition and maximum correlation of cancer specific feature sets. This hybridization aids in

Table 5   Evaluation of the proposed IABC-EMBOT scheme with feature selection

Different schemes used for comparison	Number of Hidden Nodes Utilized	Utilized count of selected features	Utilized number of connections
Proposed IABC-EMBOT	1.32	4.3	7.2
IHM-FFNN Scheme	1.47	5.4	9.1
ABCO-BCD Scheme	1.53	6.2	10.3

Table 6 Evaluation of the proposed IABC-EMBOT scheme without feature selection

Different schemes used for comparison	Number of Hidden Nodes Utilized	Utilized count of selected features	Utilized number of connections
Proposed IABC-EMBOT	1.41	4.7	7.8
IHM-FFNN Scheme	1.52	5.8	9.6
ABCO-BCD Scheme	1.59	6.5	11.2

establishing a better balance between feature dependency and computational impact. This hybridization also helps in reducing the number of hidden nodes of the MLP network that focuses on optimization with improved classification accuracy.

MLP network optimization process using the proposed IABC-EMBOT method

The MLP network used for optimization in order to facilitate accurate detection of breast cancer consists of three layers such as the input layer, hidden layer and the output layer. The input layer consists of nodes that represent the input features that are derived from the utilized data set. The hidden layer corresponds to the number of nodes (feature subset) that determines better accuracy rate as the output during detection. Thus the number of feature subset needs to be optimized for detection which means that the number of hidden nodes must be significantly reduced to determine optimal results with higher classification accuracy. In the MLP network, IABC-EMBOT is utilized as the training algorithm. This proposed IABC-EMBOT approach is useful in exploring the biases and connection weights that reduce the prediction error in the network. Further, the implementation of the proposed IABC-EMBOT approach in the MLP network necessitates encoding process that depends on three essential steps such as, a) each possible solution may be used as the connection weight

between input layer and the hidden layer, the output layer and the biases, and the connection weight between the hidden layer and the output layer. Furthermore, the possible solution is encoded as vectors that lies in the range of real numbers [−1,1]. Therefore the number of dimensions that represent the possible solution in the optimization process is estimated using Eq. (18)

$$D_{MLP} = \left(N_{IV(d)}*N_{N(HL)}\right) + \left(2*N_{N(HL)}\right) + 1 \qquad (18)$$

Where $N_{IV(d)}$ and $N_{N(HL)}$ relates to the number of input variables that are considered from the dataset and the number of neurons that constitute the hidden layer. Hence, the MLP network trained using the proposed IABC-EMBOT approach focuses on determining the best collection of connection weights that introduces the possibility of minimizing the degree of approximation or prediction error under a reduced number of hidden nodes. Thus, the cancer cells of the breast cancer are detected in a better precise manner through the benefits of ABC and IMBO algorithms that facilitate a higher degree of feature optimization process.

Results and discussions

This proposed IABC-EMBOT is implemented using Matlab 7.5 with the aid of the neural network tool box with necessary back propagation techniques utilized during the process of analysis. In this proposed IABC-EMBOT scheme, the activation function is based on hyperbolic tangent. Further, the method of winner-take-all strategy is utilized during the process of implementing the proposed IABC-EMBOT scheme, such that the effective and efficient output classification of cancer cells from the normal cells is determined from the used Winconsin Breast Cancer Data Set (WBCD). This WBCD data set is the UCI machine learning-based freely existing dataset containing maximum number of breast cancer characteristics that aids in extensive investigation in the process of diagnosing breast cancer cells. This Wisconsin data set that comprises of malignant and benign classes of breast cancer cells. The data set samples of the data set was periodically collected by Dr. Wolberg from the Wisconsin hospital and it is publicly available for the researchers [25]. This dataset highlights the chronological clustering of the data and its

Table 7 Classification accuracy and sensitivity of the proposed IABC-EMBOT scheme

First Author and Year	Classification Accuracy	Sensitivity
Zainud-Deen, 2008 [11]	93.24	91.21
Arafi, 2014 [12]	92.86	91.32
Xu, 2017 [13]	94.57	93.65
Ahmad, 2014 [14]	95.21	93.42
Belciug, 2012 [15]	95.79	92.35
Abdel-Zaher, 2016 [16]	96.32	95.42
Karthick, 2016 [17]	95.32	94.78
Nazarian, 2013 [18]	92.34	91.21
Zhang, 2018 [19]	95.97	94.18
Sakri, 2018 [20]	95.31	94.65
Proposed IABC-EMBOT scheme	97.53	96.75

Table 8   Specificity and average processing time of the proposed IABC-EMBOT scheme

First Author and Year	Specificity	Average processing time (msec)
Zainud-Deen, 2008 [11]	94.22	134.63
Arafi, 2014 [12]	93.38	156.12
Xu, 2017 [13]	94.52	132.86
Ahmad, 2014 [14]	93.28	142.11
Belciug, 2012 [15]	95.82	132.92
Abdel-Zaher, 2016 [16]	94.65	152.93
Karthick, 2016 [17]	92.83	148.91
Nazarian, 2013 [18]	93.63	141.53
Zhang, 2018 [19]	94.82	161.95
Sakri, 2018 [20]	95.24	121.34
Proposed IABC-EMBOT scheme	97.04	113.42

clustering information. The features of the WBCD dataset are determined from the breast mass through the fine needle aspirate of the digitized image. This fine needle aspirate determined in the WBCD dataset emphasize the features of the cell nuclei existing in the derived image of the breast mass [26]. This WBCD data set is significant in the accurate detection process of breast lumps in order to classify the malignant and benign types of cancer cell pattern features. This WBCD data set comprises of nine important features that determines the presence of the data related to the diagnosis of the breast cancer. The significant nine features existing in the UCI machine learning repository-based WBCD data set are mitoses, bare nucleus, clump thickness, bland chromatin, the size of each epithelial cell, cell shape uniformity, marginal adhesion, normal nucleus and deformed nucleus. The evaluation measure corresponding with each of the considered nine features from the WBCD dataset is determined to be an integer that is considered to vary between 1 and 10. This integer value assigned to each of the evaluation measuresis considered to vary from 1 to 10 depending on its closeness to the benign and anaplastic categories of breast cancer cells determined during diagnosis. The comprehensive records of the incorporated WBCD dataset of the proposed IABC-EMBOT scheme consists of 34% and 66% of benign and anaplastic categories of breast cancer cell categories. The experimental evaluation of the proposed IABC-EMBOT scheme is conducted for quantifying its predominance over the benchmarked schemes based on classification accuracy, decrease in validation error, utilized number of connections, estimation of the hidden node count used without feature selection and by feature selection in the influence of monotonically increasing iterations used in features optimization during the process of the cancer diagnosis.

First, Fig. 2 highlights the significance of the significance of the proposed IABC-EMBOT scheme quantified based on accuracy in classification with an increase in the number of generations. The accuracy in classification of the proposed IABC-EMBOT, IHM-FFNN, PSO-RM,ABCO-BCD and DNN-BCD schemes are determined to be the maximum of 97.53%, 94.32%, 93.65% and 91.25% respectively. The improvement in the mean classification accuracy is estimated to be 6.28% compared to the existing baseline breast cancer detection schemes considered for analysis. This increase in the classification accuracy rate of the proposed IABC-EMBOT scheme is mainly due to the integration of ABC and monarchy butterfly optimization approach that maintains the balance between the degree of exploitation and exploration. Then, Fig. 3 exemplars the decrease in validation error of the proposed IABC-EMBOT scheme quantified based on an increase in the number of generations utilized under implementation. The decrease in the range of validation error facilitated by the EMBOT, IHM-FFNN, PSO-RM,ABCO-BCD and DNN-BCD schemes are estimated to be 3.5-1.8, 3.1-1.6, 2.8-1.4, 2.6-1.1 and 2.3-0.5 respectively. Thus, the mean decrease in the validation error facilitated by the proposed IABC-EMBOT scheme is nearly 5.24% compared to the benchmarked breast cancer diagnosis approaches used for investigation. Furthermore, Fig. 4 depicts the performance of the proposed IABC-EMBOT and IHM-FFNN, PSO-RM,ABCO-BCD and DNN-BCD schemes evaluated using a number of connections utilized in the MLP for breast cancer diagnosis. The mean number of connections used by the proposed IABC-EMBOT and IHM-FFNN, PSO-RM,ABCO-BCD and DNN-BCD schemes are 7.8-9.9, 9.1-11.2, 10.6-13,2, 11.8-14.4 and 12.5-15.7 respectively. Mean decrease in the number of connections utilized by the proposed IABC-EMBOT scheme is 5.6% compared to the benchmarked breast cancer diagnosis approaches used for investigation. This average decrease in the number of connections of the proposed IABC-EMBOT scheme is mainly due to the utilized butterfly adjustment operator in the exploitation.

In addition, Figs. 5 and 6 quantifies the potential of the proposed IABC-EMBOT scheme quantified using utilized number of hidden nodes with and without feature selection respectively. The average number of

hidden node counts without feature selection of the proposed IABC-EMBOT and IHM-FFNN, PSO-RM,ABCO-BCD and DNN-BCD schemes are 1.4-1.8, 1.6-2.2, 1.7-2.4, 1.9-2.5 and 2.1-3,8 respectively. The mean hidden node count without feature selection facilitated by the proposed IABC-EMBOT scheme is nearly 7.83% compared to the benchmarked breast cancer diagnosis approaches used for investigation. Similarly, the average number of hidden nodes counts with feature selection of the proposed IABC-EMBOT, IHM-FFNN, PSO-RM, ABCO-BCD and DNN-BCD schemes are 1.32-1.74, 1.43-1.98, 1.61-2.33, 1.73-2.14 and 1.84-3.72 respectively. The mean number of the hidden node count with feature selection facilitated by the proposed IABC-EMBOT scheme is nearly 7.21% compared to the benchmarked breast cancer diagnosis approaches used for investigation. Furthermore, Table 1 unveils the significance of the proposed IABC-EMBOT scheme investigated using classification accuracy and utilized numbers of connections. The classification accuracy and utilized numbers of connections in the proposed IABC-EMBOT scheme under 10,20,30 generations is estimated to have improved by the mean rate of 5.62% and 5.93%, 5.02% and 5.12%, 4.12% and 4.02% respectively compared to the existing IHM-FFNN and ABCO-BCD Schemes used for benchmarking.

Further, Tables 2, 3 and 4 glorifies the confusion matrix of the proposed IABC-EMBOT scheme compared to the existing IHM-FFNN and ABCO-BCD Schemes determined under ten, twenty and thirty generations of implementation using True Positive (TP),True Negative (TN),False Positive (FP),False Negative (FN) values.

The results from Tables 5 and 6 exemplars the mean number of connections utilized in the proposed IABC-EMBOT scheme is determined to be reduced by 8.54% compared to the baseline IHM-FFNN,ABCO-BCD approaches considered for investigation. Likewise, the number of hidden node count incorporated by the proposed IABC-EMBOT scheme is determined to be reduced by 7.21% compared to the benchmarked considered for investigation. In addition, the selected number of features used in the proposed IABC-EMBOT scheme is maximized by 6.32% compared to the baseline breast cancer schemes considered for investigation.

Tables 7 and 8 highlights the significance of the proposed IABC-EMBOT scheme quantified in terms of classification accuracy, sensitivity, specificity and average processing times. The classification accuracy,sensitivity,specificity,average processing time of the proposed IABC-EMBOT scheme is confirmed to be increased by 5.21%,6.12%,5.92%,6.72%, respectively on par with the comparatively recent breast cancer detection schemes presented in the related work section.

Conclusion

The proposed IABC-EMBOT approach was presented as an attempt for achieving a better accuracy rate during the process of breast cancer detection by extracting the benefits of ABC and IMBO for sustaining the balance between the exploitation and exploration rate. The local searching ability in the employee bee phase of the proposed scheme as the training algorithm is leveraged by introducing an improved IMBO that divides the entire search space into two such that migration and adjustment may lead to optimized performance of the MLP network. The performance of the proposed IABC-EMBOT approach is confirmed to be more potential where the classification accuracy reached up to 97.53%, sensitivity up to of 96.75%, specificity up to 97.04% and average processing time is 113.42. As the future plan of work, It is planned to devise an integrated ABC- BFA (Bacterial Foraging Algorithm) that introduces swarming operator in the employee and the onlooker bee phase of ABC for enhancing the degree of predominant exploitation and exploration that results in improved classification accuracy with increased speed and precision during the process of intelligent breast cancer detection.

Complaince with ethical standards

Conflict of Interest   This paper has not communicated anywhere till this moment, now only it is communicated to your esteemed journal for the publication with the knowledge of all co-authors.

Ethical approval   This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Bick, U., Intensified Surveillance for Early Detection of Breast Cancer in High-Risk Patients. Breast Care 10(1):13–20, 2015.
2. Reinisch, M., Ataseven, B., and Kümmel, S., Neoadjuvant Dose-Dense and Dose-Intensified Chemotherapy in Breast Cancer - Review of the Literature. Breast Care 11(1):13–20, 2016.
3. Almeida, L. M., and Ludermir, T. B., A multi-objective memetic and hybrid methodology for optimizing the parameters and performance of artificial neural networks. Neurocomputing 73(7-9): 1438–1450, 2010.
4. Abdel-Zaher, A. M., and Eldeib, A. M., Breast cancer classification using deep belief networks. Expert Syst. Appl. 46(2):139–144, 2016.
5. Uthoff, J., and Sieren, J. C., Information theory optimization based feature selection in breast mammography lesion classification. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) 1(1):56–67, 2018.
6. Van der Waal, D., Verbeek, A. L., and Broeders, M. J., Breast density and breast cancer-specific survival by detection mode. BMC Cancer 18(1):67–78, 2018.
7. Agrawal, V., and Chandra, S., Feature selection using Artificial Bee Colony algorithm for medical image classification. 2015 Eighth

International Conference on Contemporary Computing (IC3) 1(2): 56–63, 2015.

8. Alshamlan, H. M., DQB: A novel dynamic quantitative classification model using an artificial bee colony algorithm with application on gene expression profiles. Saudi Journal of Biological Sciences 1(2):56–67, 2018.

9. Hasan, H., and Tahir, N. M., Feature selection of breast cancer based on Principal Component Analysis. 2010 6th International Colloquium on Signal Processing & its Applications 2(1):34–45, 2010.

10. Alshamlan, H. M., Badr, G. H., and Alohali, Y. A., ABC-SVM: Artificial Bee Colony and SVM Method for Microarray Gene Selection and Multi Class Cancer Classification. International Journal of Machine Learning and Computing 6(3):184–190, 2016.

11. Zainud-Deen, S., Hassen, W. M., Ali, E. M., Awadalla, K., and Sharshar, H., Breast cancer detection using a hybrid Finite difference frequency domain and particle swarm optimization techniques. 2008. National Radio Science Conference 2(1):88–99, 2008.

12. Arafi, A., Fajr, R., and Bouroumi, A., Breast cancer data analysis using support vector machines and particle swarm optimization. 2014 Second World Conference on Complex Systems (WCCS) 1(1):23–34, 2014.

13. Xu, H., Chen, T., Lv, J., and Guo, J., A combined parallel genetic algorithm and support vector machine model for breast cancer detection. Journal of Computational Methods in Sciences and Engineering 16(4):773–785, 2017.

14. Ahmad, F., Mat Isa, N. A., Hussain, Z., Osman, M. K., and Sulaiman, S. N., A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. Pattern. Anal. Applic. 18(4):861–870, 2014.

15. Belciug, S., and Gorunescu, F., A hybrid neural network/genetic algorithm applied to breast cancer detection and recurrence. Expert. Syst. 30(3):243–254, 2012.

16. Abdel-Zaher, A. M., and Eldeib, A. M., Breast cancer classification using deep belief networks. Expert Syst. Appl. 46(1):139–144, 2016.

17. Karthik, S., SrinivasaPerumal, R., and Chandra Mouli, P. V., Breast Cancer Classification Using Deep Neural Networks. Knowledge Computing and Its Applications 2(1):227–241, 2018.

18. Nazarian, M., Dezfouli, M. A., and Haronabadi, A., Classification of Breast Cancer Samples Through Using the Artificial Bee Colony Algorithm. International Journal of Computer Applications Technology and Research 2(5):522–525, 2013.

19. Zhang, X., Yang, J., and Nguyen, E., Breast cancer detection via Hu moment invariant and feedforward neural network. Proceedings of AIP Conference 2(1):56–65, 2018.

20. Sakri, S., Rashid, N. A., and Zain, Z. M., Particle Swarm Optimization Feature Selection for Breast Cancer Recurrence Prediction. IEEE Access 1(2):78–88, 2018.

21. Faris, H., Aljarah, I., and Mirjalili, S., Improved monarch butterfly optimization for unconstrained global search and neural network training. Appl. Intell. 48(2):445–464, 2017.

22. Arora, S., and Singh, S., An improved butterfly optimization algorithm with chaos. J. Intell. Fuzzy Syst. 32(1):1079–1088, 2017.

23. Wang, G., Deb, S., Zhao, X., and Cui, Z., A new monarch butterfly optimization with an improved crossover operator. Oper. Res. 1(1): 54–66, 2016.

24. Alshamlan, H. M., Co-ABC: Correlation artificial bee colony algorithm for biomarker gene discovery using gene expression profile. Saudi Journal of Biological Sciences 1(1):23–36, 2018.

25. Wolberg, W. H., Street, W., Heisey, D. M., and Mangasarian, O. L., Computer-derived nuclear features distinguish malignant from benign breast cytology. Hum. Pathol. 26(7):792–796, 1995.

26. Wolberg, W. H., Street, W., and Mangasarian, O., Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. Cancer Lett. 77(2-3):163–171, 1994.

27. Cheriguene, S., Azizi, N., Zemmal, N., Dey, N., Djellali, H., and Farah, N., Optimized Tumor Breast Cancer Classification Using Combining Random Subspace and Static Classifiers Selection Paradigms. Intelligent Systems Reference Library 1(1):289–307, 2015.

28. Zemmal, N., Azizi, N., Dey, N., and Sellami, M., Adaptive Semi Supervised Support Vector Machine Semi Supervised Learning with Features Cooperation for Breast Cancer Classification. Journal of Medical Imaging and Health Informatics 6(1):53–62, 2016.

29. Bhattacherjee, A., Roy, S., Paul, S., Roy, P., Kausar, N., and Dey, N., Classification Approach for Breast Cancer Detection Using Back Propagation Neural Network. Advances in Bioinformatics and Biomedical Engineering 1(1):210–221, 2016.

30. Zemmal, N., Azizi, N., Dey, N., and Sellami, M., Adaptive S3VM Semi Supervised Learning with Features Cooperation for Breast Cancer Classification. Journal of Medical Imaging and Health Informatics 6(4):957–967, 2016.

31. Kriti, V. J., Dey, N., and Kumar, V., PCA-PNN and PCA-SVM Based CAD Systems for Breast Density Classification. Intelligent Systems Reference Library 1(1):159–180, 2015.

32. Chakrabarty, S., Pal, A. K., Dey, N., Das, D., and Acharjee, S., Foliage area computation using Monarch Butterfly Algorithm. 2014 1st International Conference on Non Conventional Energy (ICONCE 2014) 1(1):78–89, 2014.