HEIGHTENED BIASED OPTIMIZED DUCK TRAVELER COLLABORATIVE FEATURE SELECTION AND CLASSIFICATION FOR BREAST CANCER DATASET PROBLEM

A.KRISHNAVENI[a]
Ph.D Research Scholar
Department of Computer Science
Chikkanna Govt. Arts College
Tirupur-2
krishnaveni120691@gmail.com

R.SHANKAR[b] & S.DURAISAMY[c]
Assistant Professor
Department of Computer Science
Chikkanna Govt. Arts College
Tirupur-2
shankarcgac2020@gmail.com
sdsamy.s@gmail.com

T.LATHA MAHESWARI[d]
Assistant Professor
Department of Computer Science & Engineering
Sri Krishna College of Engineering & Technology
Coimbatore
lathasarvesh@gmail.com

***Abstract*— Breast cancer is one of the world's leading cancer related diseases. The early prediction of breast cancer is critical as it can dramatically decrease mortality rates. Molecular biology presents an attractive method in bio inspired computing to use swarm intelligence for prediction of breast cancer. But the breast cancer database has high dimensionality problem which is solved by using Duck Traveler Optimization Algorithm (DTOA). It is inspired from the natural behavior of ducks in which they are flying randomly based on the model developed by using the Tariff Flying Appliance (TFA). TFA has some drawbacks such as swarming of the exploration area and intermission of arbitrary flights due to its huge searching steps. So, an Improved Duck Traveler Optimization Algorithm (IDTOA) was proposed for dimensionality reduction of breast cancer dataset problem. IDTOA used a Brownian gesture to solve the issues of TFA. The particle best and global best concept of basic Particle Swarm Optimization (PSO) was used in IDTOA to avoid the premature convergence. The selected features by IDTOA were given as input to Arbitrary Subspace (AS), Synthetic Neural Grid (SNG) and Chronological Trifling Optimization (CTO) to predict the breast cancer. In this paper, a Heightened Biased Optimized Duck Traveler Collaborative Feature Selection and Classification (HBODTCFSC-BC) algorithm is introduced to further refine the prediction accuracy and reduce the prediction error. At first in HBODTCFSC-BC algorithm, biases are initialized to the selected features and then a biased sum is calculated. Additionally, a conditional checking is made to check the biased sum with the optimal bias. If the biased sum is greater than optimal weight, the weighted sum value with the initialized various weight is attained. After that, a neural Duck Traveler classifier with minimum biased miscalculation (pathetic classifier) is determined and combines the pathetic classifier with the optimal bias to predict the breast cancer. Thus, the accuracy for breast cancer prediction is improved with minimum miscalculation rate. From the analysis of IDTOA-CTO and IDTOA-HBODTCFSC-BC, it is proved that the accuracy, precision, recall and F-measure of IDTOA-HBODTCFSC-BC is better than IDTOA-CTO method respectively for breast cancer extrapolation.**

Keywords — *Breast cancer extrapolation, Feature selection, Improved Duck Traveler Optimization Algorithm, Chronological Trifling Optimization, Arbitrary Subspace, Synthetic Neural Grid, HBODTCFSC-BC.*

I. INTRODUCTION

Biologically inspired computing is able to synchronously define huge number of protein sequence in a distinct investigation. In several areas of biological science, inheritable factor countenance analysis is necessary to obtain the essential awareness. When interval permits it has become more and more difficult and challenging to analysis scrutinize and delight the malady in general and breast cancer in specific. In medical image processing breast cancer prediction plays a vital role. Forecast of breast cancer is a giant anxiety and precise forecast will have a countless value to afford patients with amended precaution. Bio inspired algorithms are significant essentials and added frequently used technique for the achievement of important role for breast cancer classification.

Feature selection and extraction are the important methodology to give the significance of most prominent features in breast for identified objects in images particularly in breast. Irrelevant features are ignored by the Synthetic Neural Grid using Arbitrary Subspace. Dimensionality reduction is the technique to select the relevant data and extract the essential information from observed data in normalization. DTOA was used to reduce the dimensionality of breast expression data and LDA was used for breast cancer prediction. MIAS datasets were used for classification by Chronological Trifling Optimization (CTO).

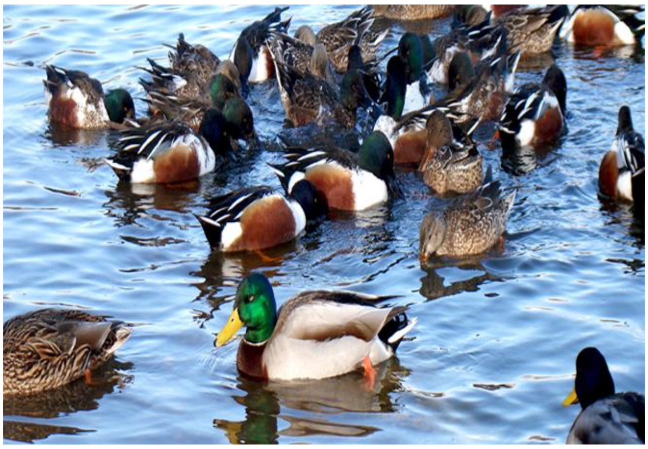

Fig. 1. Duck Flock Searching food

In DTOA, the ducks in duck flock were flying randomly based on TFA. But, it has some drawbacks like swarming of exploration zone and intermission of arbitrary flights. So, an Improved Duck Traveler Optimization Algorithm (IDTOA) [7] was introduced to solve the problems in DTOA. In IDTOA, a Brownian gesture was used to model the hunt process to get an optimal solution of ducks in duck flock. Additionally, speck best (pbest) and global best (gbest) concept of basic Particle Swarm Optimization (PSO) was used to further improve the hunt space and it evades the pre-mature coming together. The designated topographies were fed into AS, SNG and CTO classifier for prediction of breast cancer. Though, the forecast presentation of CTO classifier is better than the AS and SNG, the estimate miscalculation and accurateness are not abundant actual.

In this research, HBODTCFSC-BC algorithm is proposed to further decrease the forecast miscalculation and recover the forecast accurateness. Initially, the most relevant features are selected using IDTOA and the selected features are given as input to HBODTCFSC-BC classifier. In HBODTCFSC-BC, optimized biases corresponding to every collaborative classifier are progressively calculated with respect to the collaborative classifier outcome and connection between the outcomes of all collaborative classifiers. Primarily, a pathetic classifier with minimum biased miscalculation is calculated and a new component is then achieved on the basis of miscalculation function. The new component is processed in final collaborative classifier for better breast cancer prediction.

This research paper is organized as follows: In Section 2 various classification work by various researchers are explained. The proposed HBODTCFSC-BC classifier for breast cancer dataset feature selection and classification described in Section 3. Experimental results of HBODTCFSC-BC classifier and its performance comparison is done in Section 4. Finally Conclude the research work in Section 5 is demonstrated.

II. LITERATURE SURVEY

Sadaf Jamal Gilani et al., (2021) developed Nanostructured lipid carriers (NLC) drug for breast cancer [1]. Krishnaveni et al., (2020) introduced Improved Duck and Traveler (IDTO) algorithm using multilevel thresholding for mammogram image segmentation [2]. Krishna Gopal Dhal et al., (2019) provide a systematic review of nature inspired algorithms used for image processing especially in image segmentation [3]. B. Bhushan (2018) given the detail about bio mimicry and its significance [4]. Shen'ao Yan et al., (2017) proposed duck pack algorithm for route finding quickly than standard differential algorithm [5]. Alireza Askarzadeh (2016) proposed crow search algorithm for solving engineering design problem using intelligence of crows [6]. Nahid et al., (2018) used deep belief neural network for classification of breast cancer dataset [7]. Masoud et al., (2016) introduced chaotic PSO algorithm for wrapper selection [8]. Pratiwi et al., (2015) explained the significance of grey level co-occurrence matrix for image classification especially breast images [9]. Krishnaveni et al (2021) proposed duck travel algorithm for mammogram image segmentation using CASRG algorithm [10], and duck cluster algorithm with K Means Clustering [11]. Feng-Ping et al., (2019) proposed feed forward neural network using CNN method for breast cancer problem [12]. Alok Kumar Shukla (2020) reduced over fitting issue and dimensionality reduction new method maximum relevance minimum redundancy is introduced for accurate cancer prediction [13]. Mohamed Abd Elaziz et al., (2020) analyzed COVID-19 problem and introduced entropy using fuzzy method and improved marine predator's algorithm for CT image Segmentation [14]. David Gonzalez-Patino et al., (2020) introduced competitive algorithm for breast cancer segmentation and classification called artificial immune system association classification [15].

III. PROPOSED METHODOLOGY

The proposed HBODTCFSC-BC is explicated in feature for extrapolation of breast malignancy. Initially, a DNA segment countenance figure is collected and then the IDTOA is realistic to select the greatest important structures in the DNA segment countenance figures. The nominated structures are assumed as contribution to planned HBODTCFSC-BC classifier for extrapolation of breast tumor.

A. *Improved Duck Traveler Optimization Algorithm (IDTOA) based feature selection (FS)*

IDTOA is an improved version of DTOA that selects the most optimal features from the DNA segment countenance figures. In IDTOA, binary significant structures are encompassed in the predictable DTOA to extemporize the concert of DTOA. One of the structures is an interior retention which save observers the conceivable clarifications i.e., greatest significant structures that has a possible to meet to universal ideals. Added object is repetition glassy hybridization with Particle Swarm Optimization (PSO) which outfits on this assortment of protected explanations. IDTOA uses the shared and intellectual performance of PSO for earlier merging and to regulate the worldwide greatest explanation. Moreover, IDTOA uses the investigation aptitude of DTOA to travel the hunt interplanetary efficiently. The consequence of IDTOA is assumed as contribution to HBODTCFSC-BC to precisely categorize the DNA segment countenance figures.

B. *Heightened Biased Optimized Duck Traveler Collaborative Feature Selection and Classification (HBODTCFSC-BC)*

In a DTO, an artificial duck has td synapses associated to the selected features $(sf_1, sf_2, \dots sf_{td})$ and each feature has a corresponding bias b_n. In this case, the input at signal n is accumulated with the bias b_n, after that the addition of biased inputs and its linear combination are achieved. In addition to this, a prejudice p is summated with the linear combination and a biased sum b_s is calculated as,

$$\boldsymbol{b_s = p + b_1 sf_1 + b_2 sf_2 + \cdots b_{td} sf_{td}} \quad \textbf{(3.1)}$$

Following to this, a nonlinear activation function NAF is executed with b_s as shown in Eq. (3.2) that returns a result r as:

$$r = NAF(wb_s) \quad (3.2)$$

After that, the DTO classifier with minimum biased miscalculation is chosen which is called as pathetic classifier and is devised as,

$$\varepsilon_n = be = L_{C_i}[r + b_1 sf_1 + b_2 sf_2 + \cdots b_{td} sf_{td}] \quad (3.3)$$

$$= L_{C_i}[NAF(b_s)] \quad (3.4)$$

From Eq. (3.3) and (3.4), the minimum biased error ε_n is calculated according to the likelihood of circulation function L_{C_i} for a linear combination of biased inputs (i.e., selected features) $NAF(w_s)$. At last, a new component Tk_n according to the miscalculation function is computed as follows:

$$Tk_n = \frac{1}{td}\sum_{n=1}^{td}(Actual\ miscalculation - predicted\ miscalculation)^2 \quad (3.5)$$

In Eq. (3.5), td is the number of features which are selected by IDTOA.

Once the heightening iterations are completed, final collaborative classifier which has biased miscalculation that is better than chance, is computed by integrating all pathetic classifiers with an optimal heft H_h which is actual

$$A(H_s) = SIGN\left(\sum_{h=1}^{P} b_h A(s)\right) \quad (3.6)$$

In Eq. (3.6), $SIGN(.)$ is the sign function, P denotes the number of pathetic classifiers. The optimal weight is calculated using optimal weight learning ducks. The creation of DTO optimal weight can be stated as the following optimization problem:

$$Minimum\left\{\frac{\gamma_1}{2}\|Tk_n\|^2 + \frac{o_1}{2}\|r\|^2\right\} \quad (3.7)$$

Subject to

$$Tk = y_{ACE} - y_{OBE} = y_{ACE} - BSF \quad (3.8)$$

In Eq. (3.7), γ_1 and o_1 are the optimistic real regularization parameters, $\|r\|^2$ represents the regularize that, by the suitable selection of the regularization parameters, the cost function is smoothen at the singular point of the correlation matrix of the feature vectors to improve the robustness of the neural network with respect to the noisy environment (i.e., missing data in the gene expression data). y_{ACE} is the actual miscalculation for breast cancer prediction, y_{OBE} is the observed miscalculation by the breast cancer prediction method, B is the $\{b_1, b_2, \dots b_m\}$ and SF is the $\{sf_1, sf_2, \dots sf_m\}$.

A Lag Variety technique is used to solve the Eq. (3.7). The Lag Variety function is given as follows:

$$LV_1 = \frac{\gamma_1}{2}\sum_{i=1}^{td}\sum_{j=1}^{\widehat{HN}} Tk_{nij}^2 + \frac{d_1}{2}\sum_{i=1}^{td}\sum_{j=1}^{\widehat{HN}} b_{ij}^2 - \sum_{q=1}^{\widehat{HN}}\sum_{p=1}^{td}\lambda_{ef}\left(y_{ef} - \sum_{i1=1}^{td} b_{i1e} sf_{fi1} - Tk_{ef}\right) \quad (3.9)$$

In Eq. (3.9), $i = 1,2,\dots td$, $j = 1,2,\dots \widehat{HN}$, $\widehat{HN}$ denotes the sum of hidden nodes in the DTO, Tk_{nij} is the ij-th element of the miscalculation matrix Tk, b_{ij} is the ij-th element of the weightiness matrix $\sum_{i1=1}^{td} b_{i1e} sf_{pi1} = b_e sf_f$ and λ_{ef} is the ef Lag Variety multiplier.

Differentiating LV_1 in relation with b_{ij} is given as follows:

$$\frac{\partial LV_1}{\partial b_{ij}} = C_1 b_{ij} + \sum_{t=1}^{td}\lambda_{jt} sf_{ti} \quad (3.10)$$

Assume $\frac{\partial LV_1}{\partial td_{ij}} = 0$, will results to

$$C_1 td_{ij} = -\sum_{t=1}^{td}\lambda_{jt} sf_{ti} = -\lambda_{j1} sf_{1i} + \lambda_{j2} sf_{2i} + \cdots + \lambda_{jm} sf_{mi}$$

$$= -[\lambda_{j1} \quad \lambda_{j2} \quad \lambda_{jm}]\begin{bmatrix} sf_{1i} \\ sf_{2i} \\ \vdots \\ sf_{mi} \end{bmatrix} \quad (3.11)$$

and

$$C_1[td_{11} \quad td_{21} \quad \dots \quad td_{m1}] = -[\lambda_{11} \quad \lambda_{12} \quad \dots \quad \lambda_{1m}] \times \begin{bmatrix} sf_{11} & sf_{12} & \dots & sf_{1\widehat{HN}} \\ sf_{21} & sf_{22} & \dots & sf_{2\widehat{HN}} \\ \vdots & \vdots & \dots & \vdots \\ sf_{td1} & sf_{td1} & \dots & sf_{td\,\widehat{HN}} \end{bmatrix} \quad (3.12)$$

Hence,

$$C_1 b = -SF\lambda^T \quad (3.13)$$

Moreover, differentiating LV_1 with respect to Tk_{fij} is given as follows,

$$\frac{\partial LV_1}{\partial Tk_{fij}} = \gamma_1 Tk_{fij} + \lambda_{ij} \quad (3.14)$$

Solving $\frac{\partial LV_1}{\partial Tk_{fij}} = 0$, the following relationship is given as follows:

$$\lambda = \gamma_1 Tk_f \quad (3.15)$$

Considering the constraint in Eq. (3.8), the following Eq. (3.16) is expressed as,

$$\lambda = \gamma_1 (y_{ACE} - b^T SF) \quad \textbf{(3.16)}$$

and using (3.15) in (3.13) leads to

$$C_1 w = \gamma_1 SF (y_{ACE} - b^T SF)^T$$

$$= \gamma_1 SF\, y_{ACE}^T - \gamma_1 SF F^T b \quad \textbf{(3.17)}$$

After that, the optimal weight matrix b is derived as follows:

$$b = \left(\frac{C_1}{\gamma_1} I + SFF^T\right)^{-1} SFY_{ACE}^T \quad \textbf{(3.18)}$$

Eq. (3.18) is applied in Eq. (3.6) and the final collaborative species learning classifiers is calculated as biased majority elect of the pathetic classifiers. $A(s)$ where each classifier is allocated by biasing b_h. By collaborating pathetic classifier, the breast disease forecast correctness is better-quality and miscalculation level is concentrated. The overall movement of IDTOA-**(HBODTCFSC-BC)** based breast cancer prediction is shown in Figure 3.1.

Algorithm for Breast Cancer Prediction: IDTOA-HBODTCFSC

Input: DNA segment countenance figures, w, $b = (b_1, b_2, \dots b_{td})$,
Iteration $i = 1,2, \dots m$, optimal weight μ
Output: Breast Tumor Prediction
1: Begin
2: Select the most significant structures SF from the DNA segment Countenance figures using IDTOA
3: Initialize b
4: For each feature sf_i in F and iteration i
5: Calculate b_s using Eq. (3.1)
6: if $b_s \le \mu$ then
7: Calculate ε_n using Eq. (3.3) and Eq. (3.4)
8: Get Tk_n using Eq. (3.5)
9: Calculate the optimal weight of each pathetic Classifier using Eq. (3.18)
10: Obtain $A(H_S)$ using Eq. (3.6)
11: End if
12: Else
13: Go to step 5
14: End for
15: End

IV. RESULT AND DISCUSSION

The effectiveness of IDTOA-HBODTCFSC based breast cancer prediction method is implemented in MATLAB 2015a and compared with the DTOA and IDTOA-CTO based breast cancer prediction method with respect to accuracy, precision, recall and F-measure. For the experimental purpose, a breast cancer dataset is used which is available in https://www.mammoimage.org/databases/. This dataset consists of 322 mammogram i.e.161 pair of samples. All image having 1024x1024 pixel size in PGM format.

Table 1. Confusion Matrix (CM) for breast cancer

Analysis		Forecasted	
		Pessimistic	**Optimistic**
Real	**Pessimistic**	p	q
	Optimistic	r	s

Accuracy (ACC)	(p+s)/(p+q+r+s)
True Optimistic Value (TOV)	s/(r+s)
False Pessimistic Value (FPV)	r/(r+s)
False Optimistic Value (FPV)	q/(p+q)
True Pessimistic Value (TPV)	p/(p+q)

Table 2. Metrics involved in classification

Metrics	DTOA-SAF	IDTOA-CTO	IDTOA-HBODTCFSC
Accuracy	0.942	0.963	0.985
Precision	0.945	0.967	0.975
Recall	0.952	0.967	0.972
F-Measure	0.946	0.953	0.966

$$Accuracy = \frac{TOV + TPV}{TOV + TPV + FOV + FPV}$$

$$Precision = \frac{TOV}{TOV + FOV}$$

$$Recall = \frac{TOV}{TOV + FPV}$$

$$F - Measure = \frac{2 * P * R}{P + R}$$

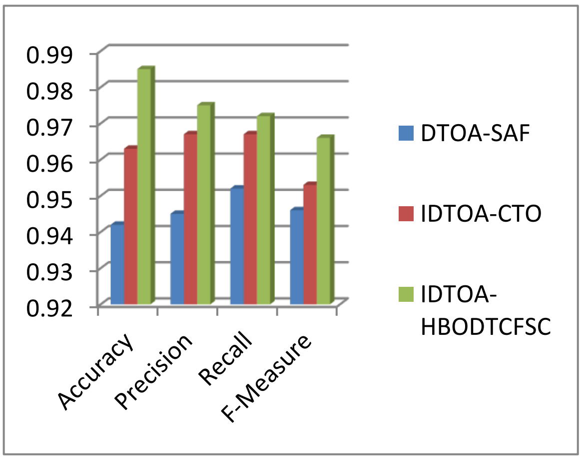

Fig. 2. Comparison of Performance metrics between proposed and existing algorithms

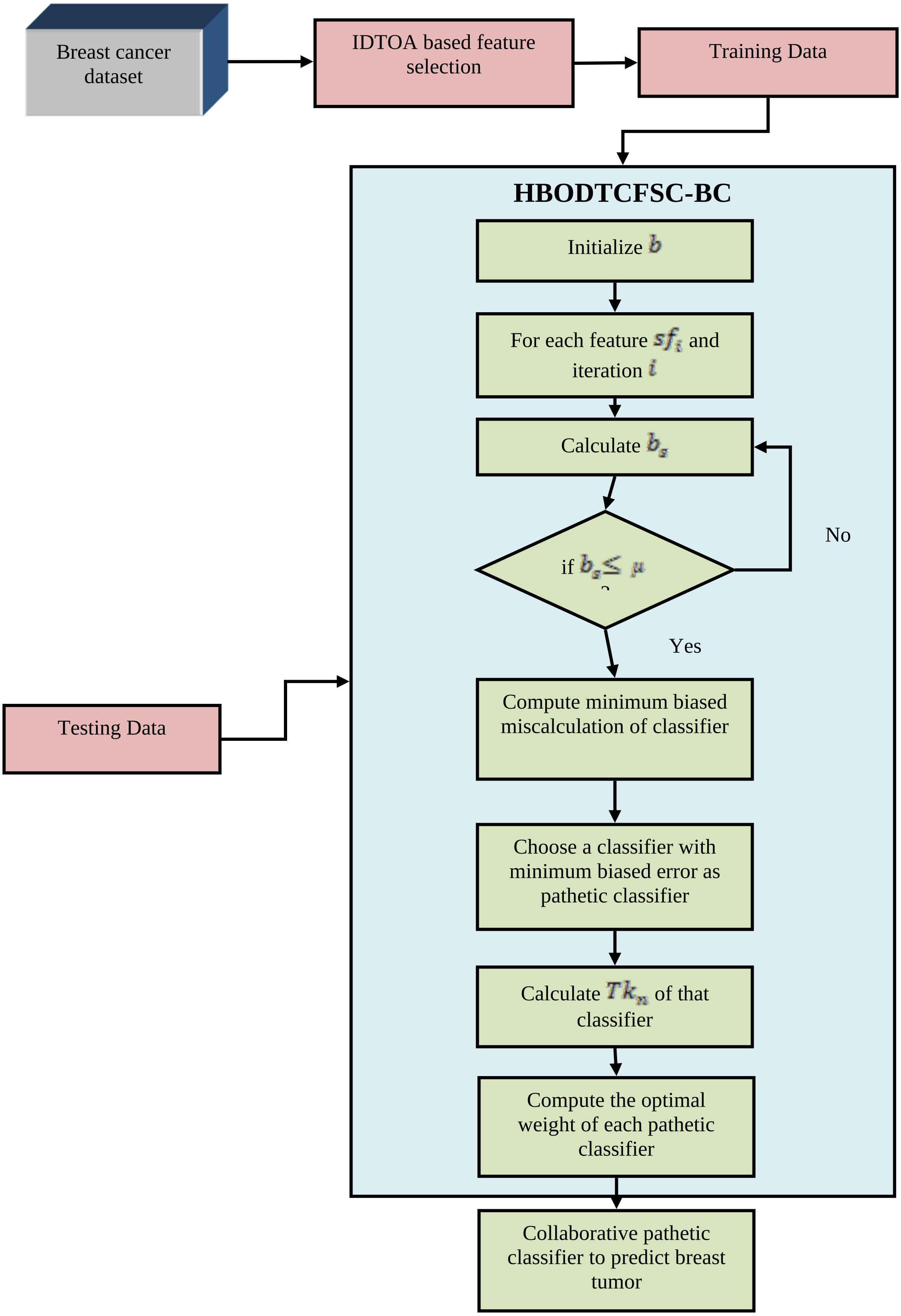

Fig. 3. System Architecture of IDTOA-HBODTCFSC-BC based breast tumor prediction

V. CONCLUSION

In this research, a HBODTCFSC to further develop the breast cancer prediction accuracy and reduce the forecast miscalculation. Here, the breast cancer prediction undergoes two processes are feature selection and collaborative classification. IDTOA choose the most significant features in feature selection process. The selected features are given as input to the collaborative classification where HBODTCFSC is applied for breast sarcoma expectation. In HBODTCFSC, optimized bias with respect to the assessment of all collaborative classifier are labeled hostilely, based on the consequence of all collaborative classifier and the construction between the consequences of all collaborative classifiers. Therefore, the breast growth estimate accuracy is enriched with low miscalculation frequency. The investigational scrutiny proves that the proposed IDTOA- HBODTCFSC based breast malignant cells prophecy method has extraordinary accuracy, precision, recall and F-measure compared to DTOA-SAF and IDTOA-CTO method. In future, HBODTCFSC is extended as hybrid IDTO classifier to achieve sparse learning with less parameter for breast cancer calculation.

References

[1]. Sadaf Jamal Gilani, May Bin-Jumah, Md. Rizwanullah, Syed Sarim Imam, Khalid Imtiyaz, Sultan Alshehri, and Mohd. Moshahid Alam Rizvi, "Chitosan Coated Luteolin Nanostructured Lipid Carriers: Optimization, In Vitro-Ex Vivo Assessments and Cytotoxicity Study in Breast Cancer Cells", Coatings (2021), 11, 158.

[2]. Krishnaveni Arumugam, Shankar ramasamy, Duraisamy Subramani, "Improved Duck and Traveler Optimization (IDTO) Algorithm: A Two-way Efficient Approach for Breast Tumor Segmentation using Multilevel Thresholding", European Journal of Molecular & Clinical Medicine ISSN 2515-8260 Volume 7, Issue 10, pp. 3793-3808, (2020)

[3]. Krishna Gopal Dhal, Arunita Das, Swarnajit Ray, Jorge Gálvez, Sanjoy Das, "Nature-Inspired Optimization Algorithms and Their Application in Multi-Thresholding Image Segmentation", Computational Methods in Engineering Springer 2019

[4]. B. Bhushan, Biomimetics, Biomimetics and Green Science and Technology Springer Series in Materials Science 279, Springer Nature Switzerland AG 2018

[5]. Shen'ao Yan, Weihong Wang, Sentang Wu, Ke Lu, "Duck pack algorithm—A new swarm intelligence algorithm for route planning based on imprinting behavior", 978-1-5090-4657-7/17/$31.00_c 2017 IEEE

[6]. Alireza Askarzadeh A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm Computers and Structures 169 (2016) 1–12

[7]. Nahid, Mehrabi, Kong, "Histopathological Breast Cancer Image Classification by deep neural network techniques guided by local clustering", Hindawi Biomed Research Int 20 pages AID 2362108 (2018)

[8]. Masoud, Nasibeh, "A hybrid search method of wrapper feature selection by chaos particle swarm optimization and local search", Turk J Elec Eng & Comp Sci volume: 24, pages: 3852-3861 (2016)

[9]. Pratiwi, Harefa, Nanda, "mammograms classification using grey level co-occurrence matrix and radial basis function neural network" Pro Cmp Sci pages 83-91 (2015)

[10]. Krishnaveni A, R.Shankar, Duraisamy S, "An Efficient Methodology for Breast Tumor Segmentation using Duck Traveler Optimization Algorithm", PalArch's Journal of Archaelogy of Egypt/Egyptology, SCOPUS Indexed ISSN NO: 1567-214X Volume 18, Issue 3, (2021).

[11]. A.Krishnaveni, R.Shankar, S.Duraisamy, "Duck Cluster Optimization Algorithm with K-Means Clustering for Mammogram Image Segmentation", Solid State Technology, SCOPUS Indexed ISSN NO: 0038-111X Volume 64, Issue 1, (2021).

[12]. Feng-Ping An, Zhi-Wen Liu, "Medical image segmentation algorithm based on feedback mechanism convolutional neural network", Biomedical Signal Processing and Control 53 (2019) 101589

[13]. Alok Kumar Shukla, "Feature selection inspired by human intelligence for improving classification accuracy of cancer types", Computational Intelligence. 2020 1–28.

[14]. Mohamed Abd Elaziz, Ahmed A. Ewees, Dalia Yousri, Husein S. Naji Alwerfali, Qamar A. Awad, Songfeng Lu, And Mohammed A. A. Al-Qaness, "An Improved Marine Predators Algorithm With Fuzzy Entropy for Multi-Level Thresholding: Real World Example of COVID-19 CT Image Segmentation", VOLUME 8, pp. 125306-125330 IEEE (2020)

[15]. David Gonzalez-Patino, Yenny Villuendas-Rey, Amadeo Jose Arguelles-Cruz, Oscar Camacho-Nieto, and Cornelio Yanez-Marquez, "AISAC: An Artificial Immune System for Associative Classification Applied to Breast Cancer Detection", Appl. Sci. (2020), 10, 515; doi:10.3390/app10020515.