A Hybrid Classification Approach Based on Improved Differential Evolution Algorithm for Breast Cancer Diagnosis

Fatemeh Lanjanian[1], Farzan Rashidi[1], Shahram Golzari[1]
1- Faculty of Engineering, University of Hormozgan, Bandar-Abbas, Iran.
Email: rashidi@hormozgan.ac.ir (Corresponding author)

ABSTRACT:
Breast cancer is one of the most common malignant tumors and the main cause of cancer death among women worldwide. The diagnosis of this type of cancer is a challenging problem in cancer diagnosis researches. Several research before have proved that ensemble based machine learning classifiers are able to detect breast cancer spot more accurate. However, the success of an ensemble classifier highly depends on the choice of method to combine the outputs of the classifiers into a single one. This paper proposes a novel ensemble method that uses modified differential evolution (DE) algorithm generated weights to create ensemble of classifiers for improving the accuracy of breast cancer diagnosis. This paper proposes an ensemble-based classifier to improve the accuracy of breast cancer diagnosis. As the performance of DE algorithm is strongly influenced by selection of its control parameters, local unimodal sampling (LUS) technique is used to find these parameters. The two most popular classifiers support vector machine (SVM) and K-nearest neighbor (KNN) classifiers are used in the ensemble. The classification is then carried out using the majority vote of the ensemble. The accuracy of the presented model is compared to other approaches from literature using standard dataset. The experimental results based on breast cancer dataset show that the proposed model outperforms other classifiers in breast cancer abnormalities classification with 99.46% accuracy.

KEYWORDS: Breast cancer diagnosis, local unimodal sampling, differential evolution algorithm, majority vote, Accuracy.

1. INTRODUCTION

Breast cancer is one of the highly prevalent concerns in female and the second overall cause of death among women exceeded only by lung cancer [1]. Early detection and diagnosis of breast cancer can improve the cure rate and guarantee the patients' life quality. However, this should have an accurate and reliable diagnosis in order to differentiate benign and malignant tumor.

There are various studies on medical aided diagnosis of breast cancer. Ref. [2] proposed a classification technique based on decision tree algorithm for breast cancer diagnosis and reached 94.74% classification accuracy using 10-fold cross validation. In Ref. [3] a neural network based fuzzy technique was used and the accuracy is obtained 95.06%. Ref. [4] suggested a RIAC technique and the classification accuracy is achieved 94.99%. Authors in Ref. [5] proposed a linear discreet analysis method and got the accuracy of 96.8%. In Ref. [6] an accuracy of 95.57% was calculated with the application of supervised fuzzy clustering technique. A fuzzy based GA method was introduced in [7] and a classification accuracy of 97.36% was obtained. Ref. [8] compared four different classifiers and the highest accuracy of 97.36% was achieved by SVM. In Ref. [9] three different techniques including LVQ, big LVQ and AIRS were used and the obtained accuracies were 96.7%, 96.8%, and 97.2% respectively. Also in [10], two different methods including Bayesian classifiers and neural networks were applied and the obtained accuracies were 92.80% and 97.90% respectively. In [11] a hybrid method based on fuzzy system and neural network was suggested and the accuracy was 95.60%. Ref. [12] evaluated the effectiveness of breast tissue type integration on calcification detection using an unsupervised machine learning technique and achieved an accuracy of 94% [13]. The comparison of three classifiers LDA, KNN and Logistic Regression were presented in [14] and best accuracy reached 96.30% by KNN. Ref. [15] stated that the accuracy can be increased by combining the intelligent optimization algorithm with machine learning based classifiers. Authors in the above paper reported the classification accuracy rate has been increased to 91% from 78% in their study. Ref. [16] stated that ensembling of classifiers can increase the accuracy of classification. The results presented in this study showed that accuracy rate was improved to 98.6% from 92.8%. Authors in [17] applied various classification methods such as Bayesian network, radial base function, decision tree, and NN algorithm for the

تاریخ دریافت: ۱۳۹۷/۰۵/۰۷ تاریخ پذیرش: ۱۳۹۹/۰۲/۰۵

UCI data set. The WEKA software provides 75% of the data for training and the rest for testing. The highest accuracy belongs to the Bayesian with 89.71%. In [18], the breast cancer data have been classified in several ways, including Bayesian method, a support vector machine with radial base kernel, radial base neural network, and decision tree. The highest accuracy was achieved for the support vector machine with a radial kernel with 98.06%. In [19], dimension reduction has been done on breast cancer data. Subsequently, these data have been classified by KNN, ANN, RBF, and SVM. The best accuracy obtained 97.33% by ANN.

Despite a great deal of public scientific research, breast cancer continues to be the most common and the second leading reason of death for women. Hence, it is needed to develop alternative identification methods to recognize breast cancer. As indicated in literature works, ensemble of classifiers is shown to possess better accuracy than any single classifier. However, the success of ensembles depends on the diversity in the outputs of each classifier, as well as on the choice of approach to combine these outputs into a single one [19]. Many heuristic and metaheuristic approaches such as PSO and GA were proposed for combining classifiers. Although literature review shows that heuristic and metaheuristic approaches, especially GA and PSO algorithm provide quite good results in the field of combining classifiers, in order to select a suitable combination method and also increase the variety of available tools, other alternatives of optimization techniques shall be taken into consideration as well. This paper proposes a novel ensemble construction method that uses DE generated weights to create ensemble of classifiers with better accuracy for breast cancer diagnosis. To achieve better performance of ensemble classifiers, LUS technique is used to find better parameters for DE algorithm.

In summary, the main contribution of this study is to introduce an expert as a collection of two classifiers that together generate a binary vector of responses. These six k-NN and six SVM are trained on the breast cancer dataset. After that, three new ensembles are created using the two intelligent optimization algorithms called DE and meta-optimized DE and the weighted majority algorithm (WMA) approach. The ensembles all combine the opinions of the twelve experts to reach the final decision. These experts are combined in three ways. In this first way, we generate weights using DE that is constructed with manually selected its own control parameters. These weights are then used with the WMA to combine the expert opinions. The second way called meta-optimized DE is similar to the first approach, except that the control parameters of DE are optimized using modified LUS (MLUS) method. The third way is to combine the opinions using WMA. Finally, the three approaches are compared. The experimental results based on breast cancer dataset showed that the new method gives better accuracy than original DE and WMA.

2. METODOLOGY

In this section, first, a brief explanation is given to some terminologies of machine learning, and then the proposed method is presented.

2.1. Support Vector Machine

Support vector machine is supervised Machine Learning technique that is based on the concepts of decision planes that define decision boundaries between the data points of classes in high dimensional space [20]. A decision plane is one that separates between a set of objects having different class memberships. SVM support both regression and classification tasks and handle multiple continuous and categorical variables [21]. SVM provides flexibility in selecting a similarity function. It gives sparseness of solution when dealing with large data sets and having ability to handle large feature spaces. In SVM, the training involves the minimization of the error function [22]:

$$\frac{1}{2}w^{T}w + C\sum_{i=1}^{N}\xi_i$$
$$\textit{Subject to the constraints:} \quad y_i\left(w^{T}\phi(x_i)+b\right) \geq 1-\xi_i, \qquad \xi_i \geq 0, i = 1,\dots,N \qquad (1)$$

Where C is the capacity constant, w is the vector of coefficients, b is a constant, and $\xi_i$represents parameters for handling inputs .The index i labels the N training cases. y_i represents the class labels and x_i represents the independent variables. The kernel ϕ is used to transform data from the input to the feature space [22].

2.2. The nearest neighbor classifier

The nearest neighbor classifier called KNN is a simple technique that stores all available instances and classifies based on a similarity measure**.** It has been used in statistical estimation. For each row of the test set, the K nearest training set vectors are found and the classification is decided by majority vote, with ties broken at random. If there are ties for the kth nearest vector, all candidates are included in the vote [23]. In order to run the KNN algorithm, a distance metric is required to measure the distance between the testing and training data.

The Euclidean distance is the most common distance metric used in KNN algorithm. This metric is the straight-line distance between two points in any number of dimensional spaces. This distance is calculated by taking the square root of the sum of the squares of differences between respective coordinates of each point [24].

$$d^2(x_i, x_j) = \left\|x_i - x_j\right\|^2 = \sum_{k=1}^{d}\left(x_{ik} - x_{jk}\right)^2 \qquad (2)$$

Where $\left(x_i, x_j\right) \in R^d$, $x_i = (x_{i1}, x_{i2}, \ldots, x_{id})$. Parameter k in KNN is the number of neighbors in the training data set. The main drawback of KNN algorithm is expensive testing of each instance [23].

2.3. Differential Evolution Algorithm

Differential evolution algorithm is an optimization technique which is very efficient and simple for global optimization over continuous spaces. It uses three operators namely selection, crossover and mutation to process its populations from one generation to another.

Many mutation strategies are available in the literature [25]. The most common one is "DE/rand/1". In this strategy, a mutant vector is generated by multiplying F by the difference between two random vectors and the result is added to a third random vector as follows:

$$\begin{gathered} V_i = X_{r1} + F \times (X_{r2} - X_{r3}) \\ X_{r1} \neq X_{r2} \neq X_{r3} \neq X_i \end{gathered} \tag{3}$$

where r_1, r_2, r_3 are random integers within the range $[1, NP]$ and $F > 0$ is a user-defined real parameter, called mutation constant. This controls the amplification of the difference between two individuals to avoid search stagnation. Following the crossover phase, the crossover operator is applied to each pair of the target vector X_i and its corresponding mutant vector V_i. Then a trial vector $U_i = [u_{i1}, u_{i2}, \ldots, u_{id}]$ is generated by [26]:

$$u_{ij} = \begin{cases} v_{ij} & if\ (rand(0,1) \leq CR)\ or\ (j = k) \\ x_{ij} & if\ (rand(0,1) \geq CR)\ and\ (j \neq k) \end{cases} \tag{4}$$

where u_{ij}, v_{ij}, and x_{ij} are the jth dimensional components of the vectors U_i, V_i, and X_i respectively. The index $k \in \{1,2,\cdots,d\}$ is randomly chosen, $CR \in [0,1]$ is a user-defined crossover constant. In other words, the trial vector consists of some of the components of the mutant vector, and at least one of the components of a randomly selected individual of the population. Then it comes to the replacement phase. To maintain the population size, we have to compare the fitness of U_i and X_i, then choose the better

$$X_i^* = \begin{cases} U_i & \text{if } f(U_i) \leq f(X_i) \\ X_i & otherwise \end{cases} \tag{5}$$

Successful implementation of DE depends on proper selection of its control parameters. Factors which impact performance of DE are crossover rate CR, scale factor F, maximum number of generation $MaxGen$, and population size NP. There is no agreement in the literature on how to optimally select these factors. To overcome above drawbacks and also avoid the tuning of parameters by trial-and-error procedure, the LUS method described in next section is used in this paper for finding the optimal values of DE parameters.

2.4. LUS algorithm

LUS method uses a search-range through iteration methods to find best choice of tuning parameters for optimization algorithms [27]. This method does not require derivative calculations. Instead, using a sampling method to find the optimal parameter, it uses a gradual matching of the searching area in a number of iterations. The main idea is to slow down frequently the search space for each parameter i to achieve the optimal solution. The search range is defined as $d = (d_1, d_2, \ldots, d_n)$, where n is the number of parameters that need to be optimized [23]. The search space has been initially specified with an upper and lower limit determined by the user.

$$\vec{d} \leftarrow \vec{b}_{up} - \vec{b}_{low} \tag{6}$$

The optimal solution s is defined as follows:

$$s = (s_1, s_2, \ldots, s_n) \tag{7}$$

In the first step, the value of each parameter is the random number with uniform distribution:

$$\vec{s} \sim U\left(\vec{b}_{low}, \vec{b}_{up}\right) \tag{8}$$

At this stage, the objective function for the primary set s is evaluated. The goal is to find a set of weights that maximizes the accuracy for each class.

After evaluating the primary objective function, the new set of parameters is defined by the following equation.

$$\vec{s}_{new} = \vec{s} + \vec{a} \tag{9}$$

where $a = (a_1, a_2, \ldots a_n)$ is a random value vector with normal distribution in search space d.

$$\left(-\vec{d}, \vec{d}\right) \sim \vec{a} \tag{10}$$

The value of the objective function is calculated using the new set of parameters and the initial values. If the new set of parameters produces more accurate weights, this set is accepted as the current optimal solution, otherwise d decreases as the following.

$$\vec{d} = q\vec{d} \tag{11}$$

where q value is obtained from Eq. (9):

$$q = 2^{-\beta/n} \tag{12}$$

In the above equation, β is specified by the user and n is the number of parameters that need to be optimized.

Pseudocode 1: Original LUS Algorithm

```
set $q = 2^{-\beta/n}$
for dimension $j \in \{1,2,\cdots,d\}$ do the following
    set $s_j \sim U\left(b_{low,j}, b_{up,j}\right)$
    set $d_j = \left|b_{up,j} - b_{low,j}\right|$
while $Iter < MaxGen$ and $fitness > acceptFitness$ do
    set $a_j \sim U\left(-d_j, d_j\right)$
    set $s_{new,j} = s_j + a_j$
    if $fitness(s_{new,j}) < fitness(s_j)$ then
        set $s_{new,j} = s_j$
    else
        $d_j = q.d_j$
```

This process repeats until the number of iterations for the intended algorithm reaches a specified range or the accuracy for the created weight reaches to 100% [28]. The pseudo-code of the LUS algorithm is shown in Pseudocode 1 [28].

3. PROPOSED SYSTEM ODEL

As stated earlier, the objective of this paper is to develop ensemble based classifiers that improves the accuracy. The proposed classification framework is illustrated in Fig. 1. As can be seen, the framework has the following steps.

3.1. Data preprocessing

After converting all data to a numerical value, the division is performed as follows. First, the whole data is divided into training and test sets. Normally 70% of the available data is allocated for training. The remaining 30% data are equally partitioned and referred to as validation and test data sets.
In order to select the kernel in the support vector machine, the RBF kernel is the first logical choice [28]. This kernel takes the data into a space with a higher dimension to be easily separable. The kernel formula is as follows:

$$k(x_i, x_j) = e^{\gamma\|x_i - x_j\|^2} \qquad (13)$$

where $\|x_i - x_j\|^2$ is the square of the Euclidean space of two points. The parameter γ is a parameter that can be determined by decision boundary. If γ is very small, then the decision boundary is no longer a curvature and is more like a small arc. In fact, the complexity of the model gets very low. If the gamma value is very large, the model may get overfitting.
The radius considered for the kernel function can be a determinant factor in the classification accuracy. Therefore, for comparison and increasing of the accuracy, six functions with different radii were used as follows: $RBF = [3,2,1,0.5,0.2,0.1]$. The second method used for classification is KNN. The KNN algorithm requires an integer value be set for parameter K. For the purpose of this research, the odd integers from 1-11 were selected for use. Only odd integers were used to avoid a tie amongst the votes. On the contrary, larger integer values for K were not used to avoid underfitting the data.

3.3. Crossover Operation

In this paper, crossover operation is performed by four methods:

- Majority vote: The views of each of the 12 classifiers are compared and the highest output is selected as the final one.
- Weighted majority vote in which the precision of each classifier is considered as its weight. Finally, after weighting, the majority vote of the output is determined.
- Again, it is the weighted majority vote, but the differential evolution algorithm determines the weight of each classifier.

It is the same as the previous method, with the difference that the parameters of the differential evolution algorithm are optimized using a local unimodal sampling algorithm method, then a weighting algorithm is used.

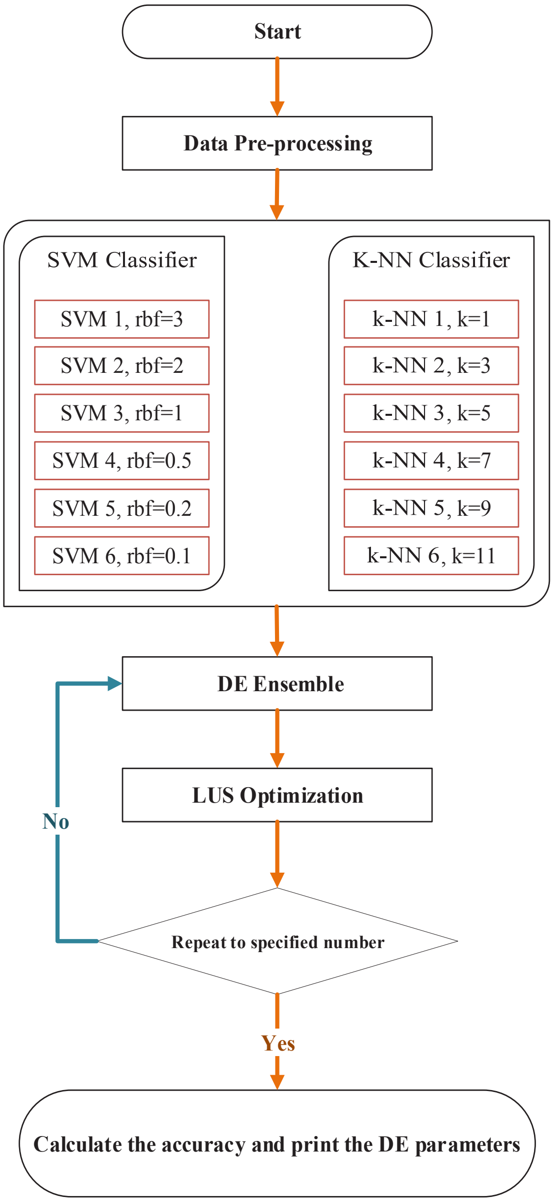

Fig. 1. The proposed system framework

3.4. Modified LUS Based DE Algorithm

3.4.1. Ensemble Classification Based on DE Algorithm

In this method, the parameters of the differential evolution algorithm are manually initialized and then DE algorithm is used for weighting classifiers. The structure of each chromosome in this method is shown in Fig. 2. In this figure, the parameters w_i are the weights assigned to each classifier.

w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}

Fig. 2. The structure of chromosome in the differential evolution algorithm

The ensemble classification based on original DE algorithm is defined by Pseudocode 2.

Pseudocode 2: Ensemble Classification Based on DE

```
Defining problem parameters and algorithms
Establishing initial population and evaluating its members
Repeat the following steps until the termination terms have not been met:
The following steps are taken for each member of the population:
A temporary response is generated using the mutation operator.
Using the crossover operator, the new response is created and evaluated.
If the new response is better than the current one, then the current response will be replaced by the new one; otherwise, the current response is kept in its place.
The best solution ever found is returned as output.
```

3.4.2 LUS Based DE Algorithm

Due to the possibility of trapping the original LUS in local minima, the following modification is done.

. Initially, the value of e is defined using Eq. 13.

$$e = \frac{e}{1 + b \times iter} \tag{14}$$

In the above equation, b and e are randomly initialized by small values between 0 and 1. Also parameter r is initialized to uniformly distributed random number with value that is inside search range [0,1]. If $r > e$, search is performed using LUS algorithm and exploitation. However, if $r < e$, each parameter chooses a value randomly. This procedure prevents the algorithm to trap in local optimal. The accuracy of these values is calculated. In this case, the previous values will be replaced with the random values. As the value of e decreases at each iteration, the algorithm occurs at the beginning of the exploration and goes toward the exploitation. The flowchart of the modified LUS algorithm is shown in Fig. 3.

4. SIMULATION RESULTS

As stated in previous sections, performance of DE is mainly influenced by its control parameters (i.e. maximum number of generation ($MaxGen$), population size (NP), scaling factor (F) and crossover probability (CR)). To find the suitable control parameters, the user uses the trial-and-error scheme. Although it is possible to determine the parameters by trial-and-error, however, however, this is not efficient and inappropriate configurations of these parameters can cause stagnation or premature convergence. Because the quality of the control parameters inserted by the user into DE strongly affects its effectiveness, also to avoid manually tuning these, we have used the modified LUS method as a meta-optimizer for finding high-quality parameters.

First the range and possible values of the DE parameters are chosen as $NP \in \{1,2,\ldots,200\}$, $MaxGen \in \{1,2,\ldots,300\}$, $CR \in [0,1]$, $F \in [0,2]$. Then MLUS is performed trying to find the optimal choice of DE parameters. After that, the objective values of a given choice of DE parameters are computed. The LUS has a number of iterations which is chosen 30 iterations. Using the above the above settings, the best performing parameters found for the DE method are as follows:

$NP = 18,\ MaxGen = 81,\ CR = 0.21,\ F = 0.66.$

The performance of each algorithm can be evaluated using various performance measures like Accuracy, Sensitivity, Specificity, Precision [19].These measures are defined by four decisions: True Positive (TP), True Negative (TN), False Positive (FN) and False Negative (FN). TP decision occurs when malignant instances predicted rightly.TN decision benign instances predicted rightly. FP decision occurs when benign instances predicted as malignant. FN decision occurs when malignant instances predicted as benign [19]. Accuracy, Sensitivity, Specificity and Precision can be calculated as:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}} \tag{15}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{P}} \tag{16}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{N}} \tag{17}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

The confusion matrix for the data set is then computed using these values into above equations to find Accuracy, Sensitivity, Specificity and Precision.

In this work, the performance of the proposed scheme has been evaluated using the dataset of Wisconsin Breast Cancer (Diagnostic) [2]. This cancer dataset contains 699 instances, with 458 benign (65.5%) and 241 (34.5%) malignant cases. Each instance is described by 9 attributes with integer value in the range 1-10 and a binary class label. The dataset is randomly divided into two subsets, one with about 70% of the instances to training, and another with around the remaining 30% of instances to testing. The performance the proposed method is evaluated using four measures: Classification Accuracy, Sensitivity, Specificity and Precision.

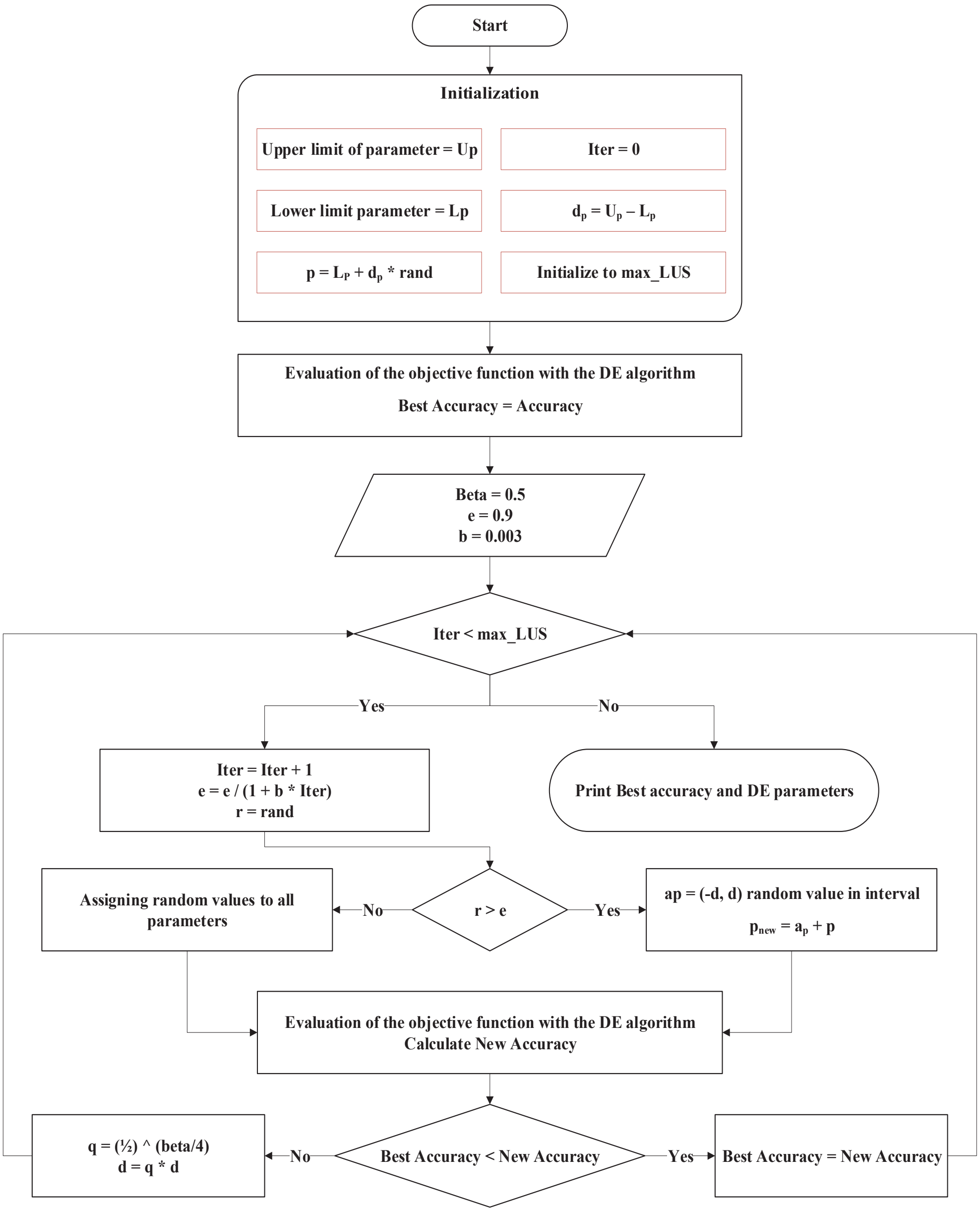

Fig. 3. The modified LUS algorithm

Fig. 4 shows the performances of the SVM, KNN, DE and MLUS_DE classifiers in terms of classification accuracy. Note that accuracy for base experts (i.e. SVM and KNN) is relatively high. The accuracy of base experts varied in correspondence to selected value of RBF or k parameters and obtained expert's accuracies are in the range from 85.30% to 96.20% and 96.80% to 98.57% for SVM and KNN respectively. Well-selected training data, good choice of RBF and k parameters and wide range of variation in selected parameters are some of the factors that produced high accuracies of base classifiers.

When considering Majority Vote (MV) and Weighted Majority Vote (WMV) methods, we observe no strikingly high rage accuracy. This is due to the relatively low accuracy of base classifiers. MV and WMV methods were proven unable to overcome difficulties presented by variations between testing and training data and perform poorer than base classifier with the highest accuracy. Hence, we concluded that combining base classifiers into an ensemble classifier with the methodology proposed in [16] has not proved to be very successful.

According to the results presented in Fig. 4, the best result we obtained was with MLUS_DE. On average we have accuracy improvement of 0.89% compared to the accuracy of the best base classifier. The success of MLUS_DE-based ensemble can be attributed to the sets of generated weights, which were further optimized to produce results with best possible accuracy. The weights generated by WMV and MV are constrained to lie between 0 and 1. Unlike these, weights generated by MLUS_DE do not have such restrictions and can take on a positive or negative value. The success of LUS_DE can be attributed to the sets of generated weights, which were further optimized to produce results with best possible accuracy. The poor accuracy of MWV may be corrected by defining different parameters for WMV. Even then, we could not hope to match the MLUS_DE results since weights generated by WMV would be strictly positive values.

A comparison of the accuracy between DE and LUS_DE based ensemble indicates that MLUS_DE achieves an average improvement of 1.36% compared to DE. In fact in this study the accuracy is improved from 98.10% to 99.46% by using LUS method. Consequently, LUS is an effective method to combine base classifiers into an ensemble for breast cancer diagnosis. It is worth mention that, although MLUS_DE has better performance than DE algorithm, however, LUS come at the cost of much longer runtimes.

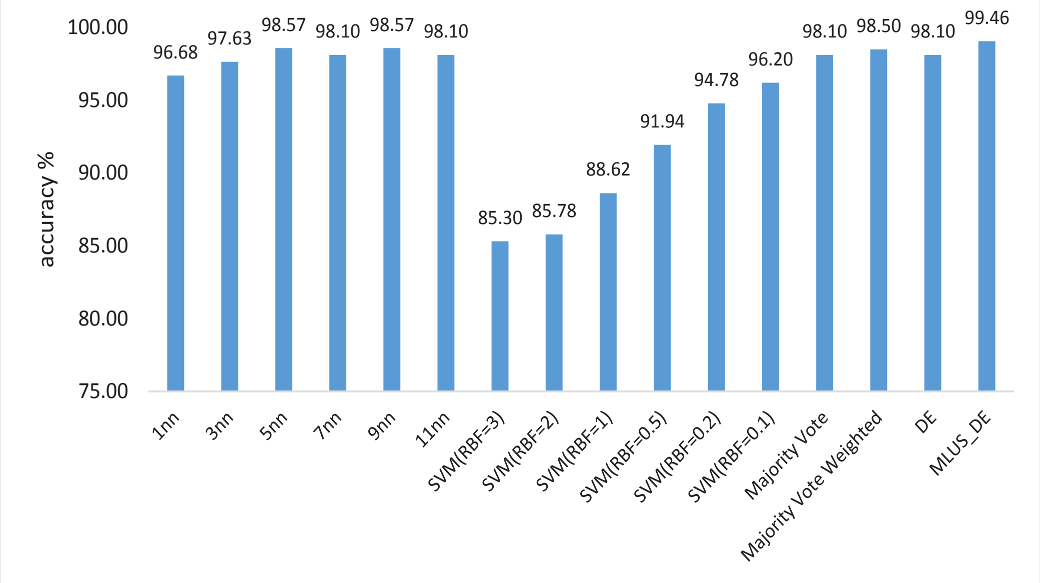

Fig. 4. Comparison between the various types of ensemble algorithms

Fig. 5 shows the performances of the DE and MLUS_DE classifiers in terms of sensitivity, specificity and precision.

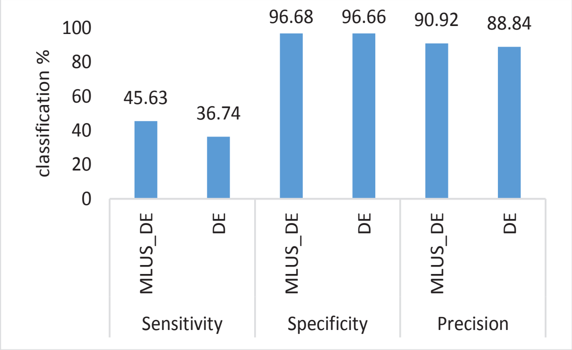

Fig. 5. The performance of DE and MLUS_DE classifiers

As can be seen, the MLUS_DE classifier performs the best for all of the evaluation metrics. In particular, the best performance is obtained by MLUS_DE for classification sensitivity 45.63%, specificity 96.68%, and precision 90.92%. Moreover, there are no big performance differences between these two methods in terms of specificity (96.68% and 96.66%).

A comparison between the results of the proposed method and the method presented in literature regarding the accuracy are also shown in Table 1. As can bee seen from this Table, the accuracy for all methods are relatively high. The accuracies are in the range from 89.71% to 99.46%. Well-selected training data, good choice of classifier parameters and wide range of variation in selected parameters are some of the factors that produced high accuracies of classifiers. The proposed method systematically provides a promising performance, scoring better or equivalently to the competitor methods on breast cancer datasets, and generally ranking among the best methods. SVM classification is the most widely used classification method for breast cancer dataset. It is suitable for structure of attributes, so it gives the best result. According to the results presented in Table 1, SVM produced overall accuracies ranging from 97.53% to 99.00% [26-30, 32] compared to 96.05% to 95.57% [31, 33, 35, 36] and 89.71% [34] for the KNN and Neural network classifiers, respectively. Neural network gives a valid result but it is not as good as SVM. KNN based classifier is as good as SVM but this approach needs modification [31]. The changes with the same methods depend on kernel specifications and importance order of attributes. If we have a look at our results, we can notice that the best accuracy is from our proposed classification and its overall accuracy, 99.46% is higher than the average value of other studies. In conclusion, we can state that weights generated with proposed metaheuristic optimization algorithm can yield improved accuracy for breast cancer diagnosis. Although, MLUS_DE based ensemble requires more times to complete classification, however improvement of accuracy is significant.

Table 1. Comparison of the obtained accuracy in this paper and the previous works

Proposed method	[39]	[38]	[37]	[36]	[35]	[34]	[33]	[32]	[31]	[30]	[29]
99.46	96.84	95.10	89.71	95.57	98.53	96.05	98.24	98.16	97.53	98.60	99.00

5. CONCLUSION

Breast cancer is the most common cancer among women throughout the world, which is one of the causes of women's mortality. The most effective way to reduce the risk of breast cancer mortality is early diagnosis and a precise and reliable method is required for early diagnosis. Machine learning is one of the methods used to detect or predict a variety of cancers, including breast cancer. This paper proposed a novel ensemble method that uses DE generated weights to create ensemble of classifiers for breast cancer diagnosis. Through experimental results, we have demonstrated that classification accuracy can be improved by ensemble of multiple classifiers into one using an ensemble approach. We combined the classifiers opinions using the DE, MLUS_DE, and the WMA approaches. The results show that weights generated with MLUS_DE can yield improved accuracy for breast cancer diagnosis system. The accuracy gains contributed by LUS come at the cost of longer runtimes. This suggests that further improvement could be achieved by implementing different optimization algorithms. Therefore, a task that could be considered for future work is to implement and compare different metaheuristic optimization algorithms for generating weights.

REFERENCES

[1] W. N. L. W. H. Ibeni, M. Z. M. Salikon, A. Mustapha, S. A. Daud, and M. N. M. Salleh, "Comparative analysis on bayesian classification for breast cancer problem," Bulletin of Electrical Engineering and Informatics, vol. 8, pp. 1303–1311, 2019.

[2] B. Zheng, S. W. Yoon, and S. S. Lam, "Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms," Expert Systems with Applications, vol. 41, pp. 1476-1482, 2014.

[3] A. M. Abdel-Zaher and A. M. Eldeib, "Breast cancer classification using deep belief networks," Expert Systems with Applications, vol. 46, pp. 139-144, 2016.

[4] J. Torrents-Barrena, D. Puig, J. Melendez, and A. Valls, "Computer-aided diagnosis of breast cancer via Gabor wavelet bank and binary-class SVM in mammographic images," Journal of Experimental & Theoretical Artificial Intelligence, vol. 28, pp. 295-311, 2016.

[5] A. T. Azar and S. A. El-Said, "Performance analysis of support vector machines classifiers in breast cancer mammography recognition," Neural Computing and Applications, vol. 24, pp. 1163-1177, 2014.

[6] S. A. Medjahed, "Applying Different Support Vector Machines Approaches in the Breast Cancer Classification," Universal Journal of Computers & Technology, vol. 1, pp. 42-48, 2016.

[7] E. Aličković and A. Subasi, "Breast cancer diagnosis using GA feature selection and Rotation Forest," Neural Computing and Applications, vol. 28, pp. 753-763, 2017.

[8] E. J. Sutton, B. Z. Dashevsky, J. H. Oh, H. Veeraraghavan, A. P. Apte, S. B. Thakur, et al., "Breast cancer molecular subtype classifier that incorporates MRI features," Journal of Magnetic Resonance Imaging, vol. 44, pp. 122-129, 2016.

[9] R. Sheikhpour, M. A. Sarram, and R. Sheikhpour, "Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer," Applied Soft Computing, vol. 40, pp. 113-131, 2016.

[10] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Computational and structural biotechnology journal, vol. 13, pp. 8-17, 2015.

[11] R. Fusco, M. Sansone, S. Filice, G. Carone, D. M. Amato, C. Sansone, et al., "Pattern recognition approaches for breast cancer DCE-MRI classification: a systematic review," Journal of

medical and biological engineering, vol. 36, pp. 449-459, 2016.

[12] P. T. Nguyen, A. Abbosh, and S. Crozier, "Three-dimensional microwave hyperthermia for breast cancer treatment in a realistic environment using particle swarm optimization," IEEE Transactions on Biomedical Engineering, vol. 64, pp. 1335-1344, 2017.

[13] N. Hamdi, K. Auhmani, and M. M. R. Hassani, "A new approach Based on Quantum Clustering and Wavelet Transform for breast cancer Classification: Comparative study," International Journal of Electrical and Computer Engineering (IJECE), vol. 5, pp. 1027-1034, 2015.

[14] U. Raghavendra, U. R. Acharya, H. Fujita, A. Gudigar, J. H. Tan, and S. Chokkadi, "Application of Gabor wavelet and Locality Sensitive Discriminant Analysis for automated identification of breast cancer using digitized mammogram images," Applied Soft Computing, vol. 46, pp. 151-161, 2016.

[15] N. I. Yassin, S. Omran, E. M. El Houby, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review," Computer methods and programs in biomedicine, vol. 156, p. 21, 2018.

[16] T. Wan, J. Cao, J. Chen, and Z. Qin, "Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features," Neurocomputing, vol. 229, pp. 34-44, 2017.

[17] W. Sun, T.-L. B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," Computerized Medical Imaging and Graphics, vol. 57, pp. 4-9, 2017.

[18] M. Peikari, M. J. Gangeh, J. Zubovits, G. Clarke, and A. L. Martel, "Triaging diagnostically relevant regions from pathology whole slides of breast cancer: A texture based approach," IEEE transactions on medical imaging, vol. 35, pp. 307-315, 2016.

[19] W. Zeng and H. Gao, "Nonlinear Classifier Design Research Based on SVM and Genetic Algorithm," Indonesian Journal of Electrical Engineering and Computer Science, vol. 13, pp. 159-165, 2015.

[20] A. R. Lubis and M. Lubis, "Optimization of distance formula in K-Nearest Neighbor method," Bulletin of Electrical Engineering and Informatics, vol. 9, pp. 326-338, 2020.

[21] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "Improving Classification Accuracy Using Clustering Technique," Bulletin of Electrical Engineering and Informatics, vol. 7, pp. 465-470, 2018.

[22] K. G. Satheesh and A. N. J. Raj, "Medical Image Segmentation and Classification Using MKFCM and Hybrid Classifiers," International Journal of Intelligent Engineering and Systems, vol. 10, pp. 9-19, 2017.

[23] M. M. Saad, N. Jamil, and R. Hamzah, "Evaluation of Support Vector Machine and Decision Tree for Emotion Recognition of Malay Folklores," Bulletin of Electrical Engineering and Informatics, vol. 7, pp. 479-486, 2018.

[24] A. Khairuddin, K. A. KNF, and P. E. Kan, "A general framework for improving electrocardiography monitoring system with machine learning," Bulletin of Electrical Engineering and Informatics, vol. 8, pp. 261-268, 2019.

[25] M. Bakhshipour, F. Namdari, and N. Bahador, "New Hybrid Non-Dominated Sorting Differential Evolutionary Algorithm," Bulletin of Electrical Engineering and Informatics, vol. 5, pp. 230-241, 2016.

[26] F. Qiao and G. Zhang, "A Fuzzy Differential Evolution Scheduling Algorithm Based on Grid," Bulletin of Electrical Engineering and Informatics, vol. 1, pp. 279-284, 2012.

[27] B. K. Sahu, T. K. Pati, J. R. Nayak, S. Panda, and S. K. Kar, "A novel hybrid LUS–TLBO optimized fuzzy-PID controller for load frequency control of multi-source power system," International Journal of Electrical Power & Energy Systems, vol. 74, pp. 58-69, 2016.

[28] A. A. Aburomman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," Applied Soft Computing, vol. 38, pp. 360-372, 2016.

[29] B. M. Abed, K. Shaker, H. A. Jalab, H. Shaker, A. M. Mansoor, A. F. Alwan, et al., "A hybrid classification algorithm approach for breast cancer diagnosis," in Industrial Electronics and Applications Conference (IEACon), 2016 IEEE, 2016, pp. 269-274.

[30] K. B. Nahato, K. N. Harichandran, and K. Arputharaj, "Knowledge mining from clinical datasets using rough sets and backpropagation neural network," Computational and mathematical methods in medicine, vol. 2015, 2015.

[31] A. Mert, N. Kılıç, E. Bilgili, and A. Akan, "Breast cancer detection with reduced feature set," Computational and mathematical methods in medicine, vol. 2015, 2015.

[32] R. Setiono, "Generating concise and accurate classification rules for breast cancer diagnosis," Artificial Intelligence in medicine, vol. 18, pp. 205-219, 2000.

[33] C. A. Pena-Reyes and M. Sipper, "A fuzzy-genetic approach to breast cancer diagnosis," Artificial intelligence in medicine, vol. 17, pp. 131-155, 1999.

[34] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," Artificial intelligence in medicine, vol. 16, pp. 149-169, 1999.
[35] K. Polat and S. Güneş, "Breast cancer diagnosis using least square support vector machine," Digital Signal Processing, vol. 17, pp. 694-701, 2007.
[36] J. Abonyi and F. Szeifert, "Supervised fuzzy clustering for the identification of fuzzy classifiers," Pattern Recognition Letters, vol. 24, pp. 2195-2207, 2003.
[37] M. N. Amin and M. A. Habib, "Comparison of different classification techniques using WEKA for hematological data," American Journal of Engineering Research, vol. 4, pp. 55-61, 2015.
[38] S. Kharya and S. Soni, "Weighted naive bayes classifier: A predictive model for breast cancer detection," International Journal of Computer Applications, vol. 133, pp. 32-7, 2016.
[39] S. Aruna, S. Rajagopalan, and L. Nandakishore, "Knowledge based analysis of various statistical tools in detecting breast cancer," Computer Science & Information Technology, vol. 2, pp. 37-45, 2011.