Wrapper based Feature Selection using Integrative Teaching Learning Based Optimization Algorithm

Mohan Allam and Nandhini Malaiyappan
Department of Computer Science, Pondicherry University, India

Abstract: *The performance of the machine learning models mainly relies on the key features available in the training dataset. Feature selection is a significant job for pattern recognition for finding an important group of features to build classification models with a minimum number of features. Feature selection with optimization algorithms will improve the prediction rate of the classification models. But, tuning the controlling parameters of the optimization algorithms is a challenging task. In this paper, we present a wrapper-based model called Feature Selection with Integrative Teaching Learning Based Optimization (FS-ITLBO), which uses multiple teachers to select the optimal set of features from feature space. The goal of the proposed algorithm is to search the entire solution space without struck in the local optima of features. Moreover, the proposed method only utilizes teacher count parameter along with the size of the population and a number of iterations. Various classification models have been used for finding the fitness of instances in the population and to estimate the effectiveness of the proposed model. The robustness of the proposed algorithm has been assessed on Wisconsin Diagnostic Breast Cancer (WDBC) as well as Parkinson's Disease datasets and compared with different wrapper-based feature selection techniques, including genetic algorithm and Binary Teaching Learning Based Optimization (BTLBO). The outcomes have confirmed that FS-ITLBO model produced the best accuracy with the optimal subset of features.*

Keywords: *Feature Selection, Integrative Teaching Learning based Optimization, Genetic Algorithm, Breast Cancer.*

Received May 15, 2019; accepted April 10, 2020
https://doi.org/10.34028/iajit/17/6/7

1. Introduction

The rapid growth in the automation of technology produces vast data with a great number of features [15]. The high dimensionality data causes the rise in the computational cost and training time of a classifier [11]. Learning models are facing many challenges with unnecessary information in huge data generated from various resources. The ability of learning models such as prediction accuracy will be reduced with unrelated features in the dataset [12]. Searching for the best combination of features for a large dataset is a challenging task. One possible fix for this issue is selecting the best combination of variables from feature space using soft computing techniques.

Feature selection models will handle the high dimensional data problems by filtering the irrelevant and redundant attributes from the data and maximize the performance and minimize the complexity of the classification models [6]. Feature Selection (FS) algorithms will search for more informative attributes (optimal features) from the overall feature set that can contribute to the results of the learning models [8]. Classification models will produce good predictive results with the optimal feature subset. FS techniques based on the wrapper approach uses the efficiency of a classification model in choosing the best features from a dataset [14]. In the wrapper technique, a learning model is used to assess the performance of the selected group of features from the dataset. The outcome of the classifier is used as feedback to the feature subset generation process in each iteration. However, continuous interaction with a learning model for feature evaluation makes the wrapper model complex and computationally expensive, specifically for high-dimensional data. Different searching techniques are adopted to boost the ability of existing feature selection algorithms.

Many authors employed several optimization algorithms in the feature selection for searching the best informative attributes. Genetic Algorithm (GA) is extensively adopted in several attribute selection techniques to boost the results of the classifiers. The combination of GA and k-Nearest Neighbors (k-NN) classifier which provides loss function as a fitness value boosts the accuracy of prediction models [18]. The GA-SVM combination also used for wrapper-based feature selection models [32] for better results. GA selected the best combination of features by randomly searching the solution space with crossover and mutation strategies. Moradi and Gholampour [17] suggested a novel feature selection process using Particle Swarm Optimization (PSO) and a local search approach to improve the performance. They used a wrapper-based procedure to evaluate the system and achieved improved accuracies. Hafez et al. [9] used a chicken swarm optimization algorithm to design a new feature selection technique. This new method was tested against standard datasets and accomplished good outcomes compare to GA and PSO based methods. Most of the researchers are using GA for feature selection due to its simplicity. A new system has been proposed by panda [19] to explore microarray data using the elephant search optimization algorithm and a deep neural network. Rodrigues et al. [25] addressed a new wrapper-based approach by involving the bat algorithm for feature selection.

The optimization algorithms use many controlling parameters throughout the selection process in the wrapper approach. These parameters need to be tuned for getting better results. Rao et al. [21] developed a simple and efficient swarm intelligence-based algorithm named Teaching Learning Based Optimization (TLBO) [24] and applied in numerous applications of different domains [21, 22] TLBO requires very less number of attributes compare to other optimization algorithms. Existing feature selection methods based on the TLBO and Binary TLBO algorithms obtained an optimal number of variables and accomplished moderate results with the Wisconsin Diagnostic Breast Cancer (WDBC) dataset. But, sometimes Binary Teaching Learning Based Optimization (BTLBO) converges very fast due to the problem of a local minimum for classification error. In this work, we proposed a novel supervised wrapper FS technique that incorporates an integrative TLBO (FS-ITLBO) algorithm to streng then the classification ability of various learning models. The performance of learning models will be measured using classification accuracy or error rate. The proposed algorithm was improved by introducing multiple teachers in the teaching phase and changing the default values of the teaching factor. The key contributions of the proposed method include,

a) Designing an integrative version of the TLBO algorithm with multiple teachers-Integrative Teaching Learning Based Optimization-(ITLBO) concept.
b) Developing a medical disease analysis scheme with the help of the ITLBO algorithm and several learning models.

The work is planned as follows. In the subsequent section, several existing feature selection techniques using different optimization algorithms were discussed. Section 3 presents the advanced model of FS-ITLBO to achieve better outcomes. The fourth section deals with the responses of the proposed model by comparing the outcomes by means of the original complete feature set and existing feature selection methods. The last section contains the summary and the future scope of our work.

2. Related Works

In machine learning, the performance of a particular classifier is mainly depending on the correlation of the features with the problem used in training. Selecting a relevant group of features will improve the accuracy of a decision model that is used in different applications. We have studied various optimization techniques involved in optimal feature selection. Mafarja *et al*. [16] addressed a feature selection procedure by adopting binary dragonfly algorithm and evaluated the performance on UCI data. Sayed *et al*. [27] used a crow search algorithm to develop a novel feature selection method and verified the performance on different benchmark problems. Sayed *et al*. [28] addressed the problems of local optima and proposed a new solution with the salp swarm algorithm. The model also used chaos theory to deal with the problem of low convergence and achieved better performance with benchmark data. Rajamohana and Umamaheswari [20] used an improved version of binary PSO and shuffled frog leaping algorithms to manage the high dimensionality problems for helping the customers in the identification of fake reviews.

Many researchers used optimization-based feature selection algorithms for medical image diagnosis [1, 2, 29]. Hossam *et al*. [10] used the PSO algorithm as a search method for selecting informative predictors in the Breast Cancer (BC) dataset and verified the efficiency with various classifiers. A novel FS technique was presented by Sridevi and Murugan [31] using a modified correlation rough set for medical investigation and selected a group of variables from the BC dataset. Bhardwaj *et al*. [5] used genetic programming for the selection of features from WBC and WDBC datasets to diagnose breast cancer. Arora and Anand [4], proposed a FS method using the butterfly optimization approach with two alternatives and achieved better efficiency with the BC dataset.

The TLBO algorithm was deployed in various FS methods. Shahbeig *et al*. [30] recommend a hybrid FS technique based on TLBO in combination with a modified PSO algorithm. They used Support Vector Machine (SVM) classifier to assess the efficiency of the new model with the breast dataset. Allam and Nandhini [3] stated a binary version of the TLBO algorithm for FS and achieved better efficiency with the BC dataset. A new multi-objective FS algorithm was presented by Kiziloz *et al*. [13] based on the TLBO algorithm to classify the binary datasets with the machine learning methods. Tuo *et al*. [33] used the harmony search approach to provide a solution to the high dimensionality problem and developed a hybrid TLBO algorithm. Rao and Patel [23] enhanced the TLBO algorithm by assigning more than one teacher to a class. The class is fragmented into multiple groups based on the knowledge (fitness) of the learners (solutions). Each group was assigned to a different teacher to improve the results of the learners.

After surveying many research papers we recognized that majority of the researchers employed optimization algorithms for feature selection. Most of these algorithms are producing local optimum solutions with faster convergence problems. A robust integrative algorithm is proposed to overcome the above problem. The FS-ITLBO model is discussed in the next section.

3. Materials and Method

Feature selection is used to filter the solution space by selecting informative attributes from the dataset. Initially, fitness is measured for every solution in the population and new solutions are created for the next generations. After a few generations, the population contains better solutions compared to the previous generations. Our proposed feature selection model contains two sections as shown in Figure 1. In the first section, the wrapper-based FS is performed using an optimization algorithm. In the second section, the efficiency of the new method is measured with various training models on the dataset. In this proposal, an integrated version of the TLBO algorithm has been developed to achieve more performance in selecting a group of attributes from the data. This proposal deals with the cancer data from which the FS-ITLBO model selects an optimal subset of features for classification.

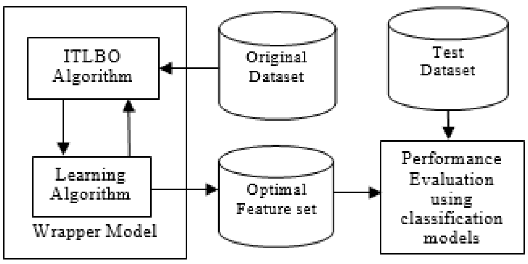


Figure 1. Wrapper-based FS model using the ITLBO algorithm.

The FS-ITLBO model has been developed to achieve better performance than BTLBO for feature selection. BTLBO is a modified version of TLBO which contains binary solutions (feature string) in the population devised for a feature selection problem. TLBO algorithm adopted the concept of student learning in a school. First, the teacher shares their knowledge with learners, and later, learners gain knowledge from the neighbor learners in the school.

TLBO needs only the information of the number of solutions and iterations. But, the ITLBO algorithm uses one more parameter called "number of teachers" to train the learners in each iteration. ITLBO algorithm can be discussed in two levels named the teaching phase followed by the learning phase which is displayed in Figure 2. A classification algorithm is required to approximate the fitness of the learners in the wrapper-based FS method.

Integrative TLBO Algorithm: FS-ITLBO model is the enhanced variation of the existing binary TLBO algorithm in which multiple number of teachers will be involved iteratively in the first phase of the algorithm. This concept will make the model very strong by exploring the entire search space without hanging in the local minima. The two phases of the algorithm will be looped until an ending condition has been encountered. The fixed count of iterations is considered as a termination criterion for this algorithm.

In each iteration (g), 'v' represents features {v=1, 2…m}, and 'i' represents solutions ({i=1, 2…p}). $S_{v,I,g}$ represents a subset of features. The flow of the ITLBO procedure is stated below.

- *Step* 1: Initialize the size of the population, number of generations, and the count of teachers.
- *Step* 2: Calculate the mean of every attribute for solutions ($M_{v,g}$).
- *Step* 3: Find the best solution and compute the fitness of individual solutions by means of Equation (1).

$$\text{Fitness}\left(S_{v,i,g}\right) = \text{Accuracy}\left(S_{v,i,g}\right) \tag{1}$$

Teacher Phase:

- *Step* 4: Reform individual solutions by means of teachers.

  a) Choose the top most fitness solution (The large value of accuracy or a low value of error rate) from the group as a trainer.
  b) Find the mean-variance concerning the best solution as mentioned in Equation (2).

$$\text{Diff}_{\text{Mean}_{v,i,g}} = r_g\left(S_{v,\text{ibest},g} - T_F M_{v,g}\right) \tag{2}$$

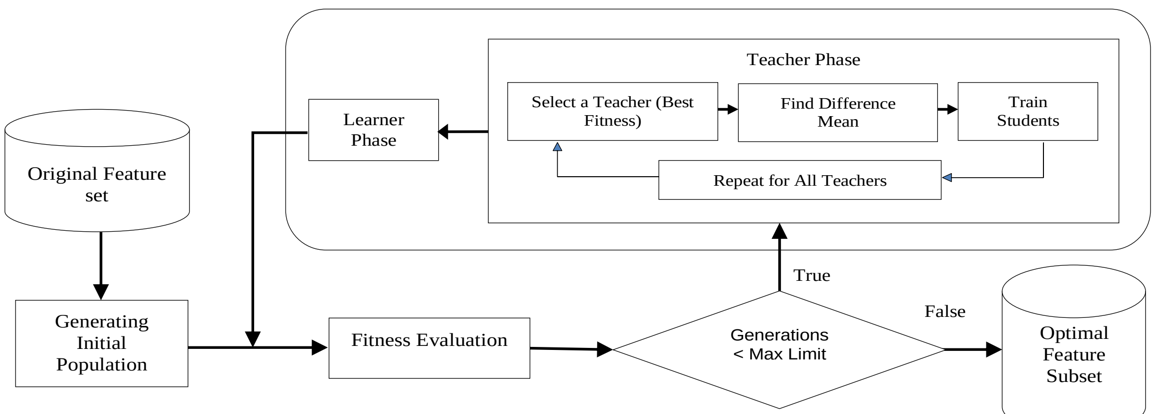


Figure 2. Workflow of the FS-ITLBO mode.

Where, $X_{v,ibest,g}$ is the best solution, T_F is called as a teaching factor which ranges from 1 to 2, rg is an arbitrary value that exists between 0 and 1.

c) The teacher trains the remaining solutions with the Equation (3).

$$\begin{aligned} S'_{v,i,g} &= 0 \quad \text{if } S_{v,i,g} + \text{Diff}_{\text{Mean}_{v,g}} < rand(1) \\ S'_{v,I,g} &= 1 \quad \text{if } S_{v,I,g} + \text{Diff}_{\text{Mean}_{v,g}} \geq rand(1) \end{aligned} \tag{3}$$

Where, $S'_{v,I,g}$ is the final modified value of $S_{v,I,g}$

d) If $S_{v,I,g}$ is well qualified than $S'_{v,I,g}$ Continue the original solution else, Substitute with the new solution
e) Repeat the steps from (a) to (d) for all teachers.

In the teaching stage, the mean and difference mean values gives the likelihood of a specific attribute in the population and attribute variation between teacher and remaining solutions. Equation (3) will resolve the presence of a particular attribute in the forthcoming generations.

Learner Phase:

- *Step* 5: Update every solution in the population by means of neighbor solutions as follows,

  a) Select two solutions A and B from the population such that $S'_{total-A,g} \neq S'_{total-B,g}$ at random.

Where, $S'_{total-A,g}$, $S'_{total-B,g}$ are updated solutions for original solutions ($S_{total-A,g}$, $S_{total-B,g}$).

  b) The solution with the highest fitness rate will affect the remaining solutions based on the Equations (4) and (5).

If the fitness of the first solution ($S'_{total-A,g}$) is better than the second solution ($S'_{total-B,g}$) then update using Equation (4).

$$\begin{aligned} S''_{v,A,g} &= 0 \quad \text{if } S'_{v,A,g} + r_g\left(S'_{v,A,g} - S'_{v,B,g}\right) < rand(1) \\ S''_{v,A,g} &= 1 \quad \text{if } S'_{v,A,g} + r_g\left(S'_{v,A,g} - S'_{v,B,g}\right) \geq rand(1) \end{aligned} \tag{4}$$

else update the solutions using Equation (5),

$$\begin{aligned} S''_{v,A,g} &= 0 \quad \text{if } S'_{v,A,g} + r_g\left(S'_{v,B,g} - S'_{v,B,g}\right) < rand(1) \\ S''_{v,A,g} &= 1 \quad \text{if } S'_{v,A,g} + r_g\left(S'_{v,B,g} - S'_{v,A,g}\right) \geq rand(1) \end{aligned} \tag{5}$$

  c) If the updated solution ($S''_{v,A,g}$) is better than the previous solution ($S'_{v,A,g}$),

Then substitute the previous solution with the new solution in the population for the next generation. Else, continue the old solution in the population for the next generation.

- *Step* 6: If the final ending criteria satisfied, Then output the final solution else, proceed to Step 2 for the next generation.

The population is characterized by binary solutions in the FS-ITLBO procedure. The bits ‘1’ and ‘0’ describes the existence and non existence of a specific attribute in the solution. The number of bits in the solution string is identical to the dimension of each record. The number of attributes participating in the solution is identical to the sum of attributes in the record.

In the initial stage, the best solutions act as teachers and train outstanding solutions by considering them as beginners to expand their knowledge. Here, all learners (solutions) are trained by every teacher of the class in the corresponding periods (iterations). It helps the learners to gain more knowledge (explore maximum feature space) from all subjects. Wrapper-based models use classifiers to compute the fitness of solutions in the population. Classification accuracy or error will be used as an aptness for maximization and minimization problems respectively. The solution with the maximum correctness value or minimum error rate will be considered as a trainer for this stage. The trainer educates the students to carry their solutions in the direction of best mean differences (v). Different classification models are used to measure classification errors and accuracies. The teaching factor could be taken in a range of 1 to 2. The Classification Error (CE) can be defined as shown in Equation (6).

$$\text{CE} = \frac{\text{Incorrectly Classified Instances}}{\text{Total Instances}} \tag{6}$$

In the second stage, every student (solution) in the population will gain knowledge through other randomly nominated students from the population. After each iteration, the population holds the top most solutions employing the highest fitness cost. In the last iteration, the solution with the highest competence will be taken for selecting the subgroup of features from the population. The iteration count will be treated as termination criteria for our model. The dataset formed with the new subdivision of features will be used to build classification models to achieve better prediction results. Figures 3 and 4 show the flow of the TLBO and ITLBO algorithms respectively. The key contributions of this proposed algorithm are,

1) The teacher phase involves multiple teachers in every generation to explore the full feature space and escape of the local minimum problem. It also makes the proposed model converge with average speed.
2) Direct value of 1 or 2 is used for the teaching factor instead of using the random values of 1 or 2. It allows the algorithm to exploit the solution for a global maximum.
3) The process of updating solutions is improved in teaching and learning phases using the randomly generated limit value instead of using the default value (0.5) to cover maximum solution space.

The proposed FS with the above-mentioned conditions accomplished enhanced conclusions on the datasets related to total features, GA and BTLBO selected subset.

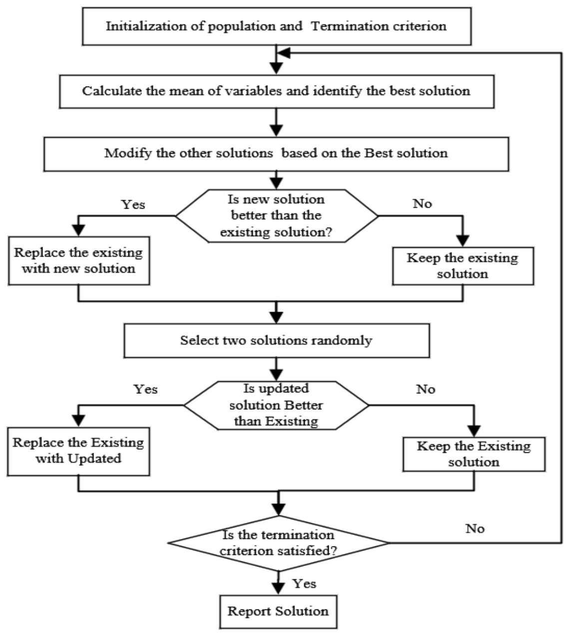


Figure 3. Workflow of the TLBO algorithm.

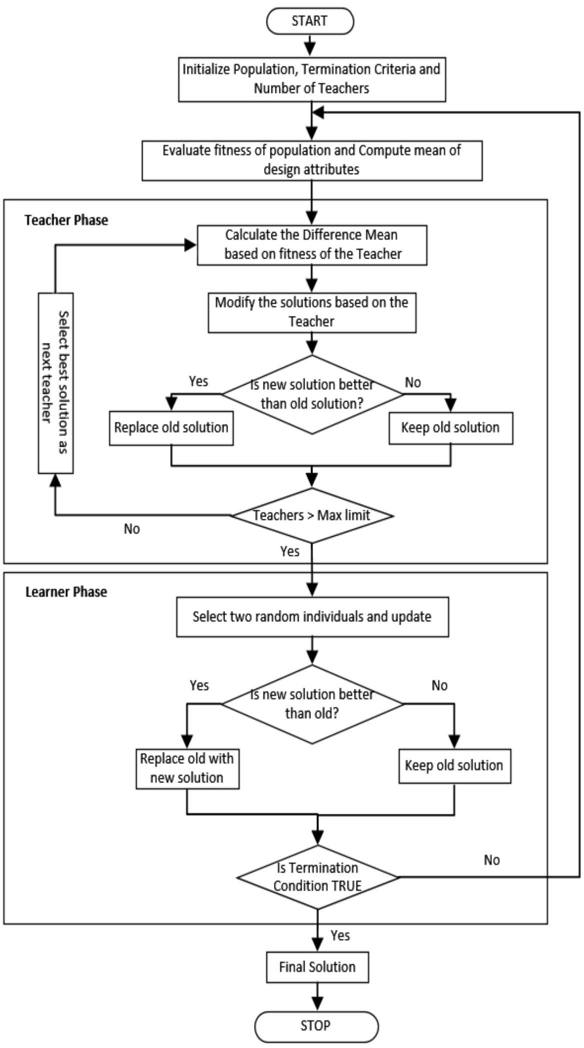


Figure 4.Workflow of the ITLBO algorithm.

4. Results and Discussion

Our recommended algorithm ITLBO is a novel FS method to improve the performance with a reduced subset of features. Classification accuracy has been used to evaluate the efficiency of the algorithm and classification error as an assessment point to identify the optimal group of features. The algorithm was initialized with a population of size 30 and iterated for 50 generations. Several classification models were developed with all dataset features and a subset of features selected using BTLBO along with the proposed algorithm. The records in the dataset were shared to train and validate classifiers for testing different feature selection algorithms.

The efficiency of the ITLBO algorithm was tested through various classification algorithms (SVM, Naïve Bayes, Decision Trees, k-NN, and Linear Discriminant Analysis (LDA)) and compared with the responses of BTLBO algorithms for binary classification. The performance of the FS algorithms was assessed in the form of classification accuracy and error for WDBC [7] and Parkinson's Disease (PD) [26] datasets. Wisconsin (Diagnostic) BC dataset has been utilized to estimate the efficiency of the FS-ITLBO model. The dataset contains 569 patient records in which 357 related to malignant tumors and 212 related to benign tumors. There are 30 features in each record that are extracted from images of BC patients. The PD dataset was prepared from the speech tapes of the patients. Several speech signal processing techniques have been used to obtain effective data for disease critical analysis. The data is collected from 252 people in which 188 are suffering from the disease. The dataset contains 756 samples and each sample provides 754 features.

4.1. Performance Evaluation of FS-ITLBO on WDBC Dataset

Table 1 displays the comparison of prediction accuracies generated by various classification models with the overall attributes of WDBC and ITLBO selected attributes. Figure 5 gives the graphical illustration of the performance improvement in terms of classification results for the proposed model. The highest improvement in the classification accuracy (98.06%) is attained with ITLBO selected subset of features using a Naïve Bayes classifier when compare to the initial value of 94.02%. The subset of features generated with the SVM classifier achieved only a nominal performance hike from 98.76% to 99.29% accuracy. The k-NN classifier also provided better accuracy values of 94.72% and 97.71% with the original and ITLBO selected features respectively. The Decision Tree model produced the overall best classification accuracy of 99.96% with the optimal subset of attributes. Finally, the discriminant analysis classifier achieved 96.48% accuracy with the overall attributes and 97.89% with the subgroup of features.

Table 1. Classification accuracies with Total and FS-ITLBO selected features.

Classification Models	Overall Features	FS-ITLBO Selected Features
Naive Bayes	94.02	98.06
SVM	98.76	99.29
k-NN	94.72	97.71
Decision Tree	98.94	99.96
LDA	96.48	97.89

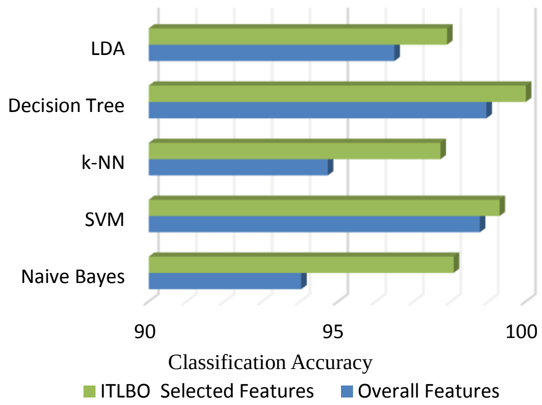


Figure 5. Comparison of classification accuracy on WDBC Dataset.

Table 2 demonstrates the subset of features selected as binary strings with various classification models.

Table 2. Final selected features bit vectors.

Classification models	Features bit vectors
Naive Bayes	010010000000000011000111100100
SVM	000001110011010101011101100110
K-NN	000010111000001100100000000000
Decision Tree	101001110000000001110110000010
LDA	011001000000001110011001000110

The proposed wrapper-based FS system selected the least number of features using different classifiers as cost functions (Fitness function) to find the fitness (classification error rate) of solutions and also trained various classification models with a significant combination of optimal features as shown in Table 3.

Table 3. FS-ITLBO selected optimal features and fitness values.

Classification Model	Error Rate	Features Count	Selected Optimal Features
Naive Bayes	1.93	9	2, 5, 17, 18, 22, 23, 24,25,28
SVM	0.7	15	6, 7, 8, 11, 12, 14, 16, 18, 20, 21,22, 24, 25, 28, 29
K-NN	2.28	7	5, 7, 8, 9, 15, 16, 19
Decision Tree	0.04	11	1, 3, 6, 7, 8, 18, 19, 20,22, 23, 29
Discriminant Analysis	2.1	11	2, 3, 6, 15, 16, 17, 20, 21, 24, 28, 29

The feature selection algorithm produced 9 features (2, 5, 17, 18, 22, 23, 24, 25, and 28) out of 30 using the NB classifier with the fitness value of 1.93 as shown in Table 3. SVM classifier has given the fitness of 0.7 and selected a subset of 15 features (6, 7, 8, 11, 12, 14, 16, 18, 20, 21, 22, 24, 25, 28 and 29). The k-NN classifier generated a substring with 7 features (5, 7, 8, 9, 15, 16, and 19) from the dataset which is the best solution compared to the remaining classifiers as fitness functions. Decision tree and discriminant analysis suggested 11 features (1, 3, 6, 7, 8, 18, 19, 20, 22, 23 and 29) and (2, 3, 6, 15, 16, 17, 20, 21, 24, 28 and 29) with the cost of 0.04 and 2.1 respectively with our improved feature selection algorithm.

4.2. Comparison of FS-ITLBO Results on WDBC Dataset

The learner's update process of the BTLBO algorithm is improved using multiple best solutions in the ITLBO algorithm. So, the results of ITLBO are compared with BTLBO for a better understanding of the improvement in the performance of the feature selection process. The results also compared with GA which is proved and widely used for feature selection. GA and BTLBO algorithms are evaluated along with ITLBO on BC dataset for automatic diagnosis of BC and achieved better results with all the classification models as depicted in Table 4.

Table 4. The analogy of classification accuracies for GA, BTLBO, and ITLBO algorithms.

	Overall Features	GA based Features	BTLBO based Features	ITLBO based Features
Naive Bayes	94.02	97.36	97.36	98.06
SVM	98.76	98.94	99.12	99.29
k-NN	94.72	96.30	96.66	97.71
Decision Tree	98.94	99.82	99.64	99.96
LDA	96.48	97.36	97.53	97.89

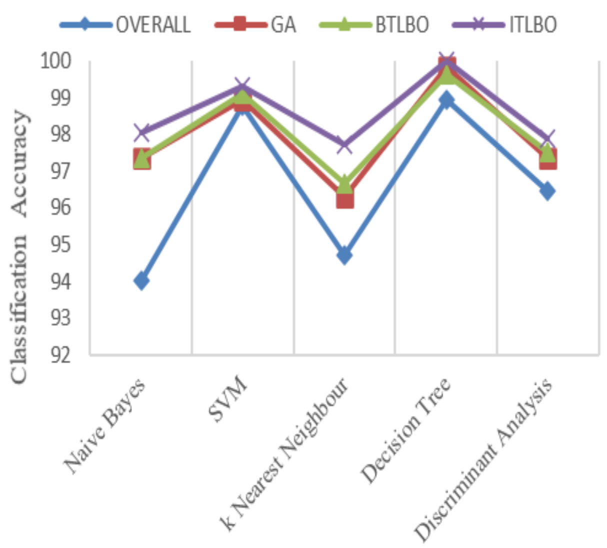
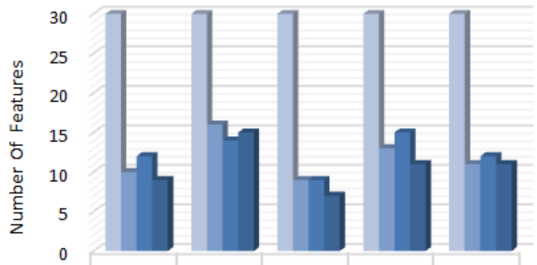


Figure 6. Comparison of FS-ITLBO performance with other Feature Selection methods.

The features opted by BTLBO algorithm achieved better accuracy values compared to the results of the original feature set and got only comparable results against GA. The FS-ITLBO model improved all the classification models prediction results by providing an optimal subset of features that can be found in the graph as observed in Figure 6. The ITLBO selected a smaller number of features to compare with GA and BTLBO algorithms for most of the classification models as shown in Figure 7. The k-NN and SVM classifiers selected the least and highest average number of features with different feature selection methods. FS-ITLBO generated only 7 features out of 30 from the dataset.

In feature selection, reducing the number of features and achieving the best classification accuracy with a classifier are two distinct issues. A feature selection model needs to balance the trade-off between these two objectives. BTLBO improved the classification accuracy to some level but ineffective in reducing the number of features when compared with GA as shown in Table 4. Where as, the ITLBO algorithm explored the maximum feature space by incorporating multiple teachers in the search process to identify the minimal subset of features.



	Naive Bayes	SVM	k-NN	Decision Tree	LDA
OVERALL	30	30	30	30	30
GA	10	16	9	13	11
BTLBO	12	14	9	15	12
ITLBO	9	15	7	11	11

Figure 7. Comparison of features selected by the FS-ITLBO model on WDBC data.

ITLBO reduced the number of features in addition to improving the classification accuracies when compared with BTLBO as shown in Figure 7. As more priority has been given to the prediction accuracy of classifiers than the number of features selected, the ITLBO achieved slight improvement in the accuracy and selected one more attribute than the BTLBO FS model for the SVM classifier.

The Convergence of FS-ITLBO and BTLBO algorithms are represented on graphs in which X-axis specifies generations and Y-axis represents prediction errors. Figure 8 shows the convergence of feature selection algorithms regarding the Naive Bayes classifier. The error curve in the graph specifies the faster convergence of the BTLBO algorithm with a value of 0.02636 which indicates the solution at local best. The proposed feature selection algorithm delayed the convergence by searching the global best solution. SVM classifier converged at 0.008 and 0.007 error values with BTLBO and ITLBO algorithms respectively for less difference in iterations as depicted in Figure 9.

The proposed algorithm brought down the classification error by exploring the solution space in a moderate number of generations compares to the BTLBO algorithm as shown in Figure 10. The decision tree classifier converged very fast (in few iterations) at 0.0035 error value through the BTLBO feature selection process. BTLBO algorithm took a few more generations for choosing the best subset of features with the least error value (0.004) as shown in Figure 11. Discriminant analysis classifiers took many iterations for converging with the BTLBO algorithm but improved algorithm gradually converged at 0.021 with in a minimum number of generations.

The BTLBO algorithm converged very fast for most of the classifiers and very slow with LDA as shown in Figure 12. Coming to the FS-ITLBO model, all the classification models converged moderately indicating the consistency in searching the feature space for choosing the optimal features from the information.

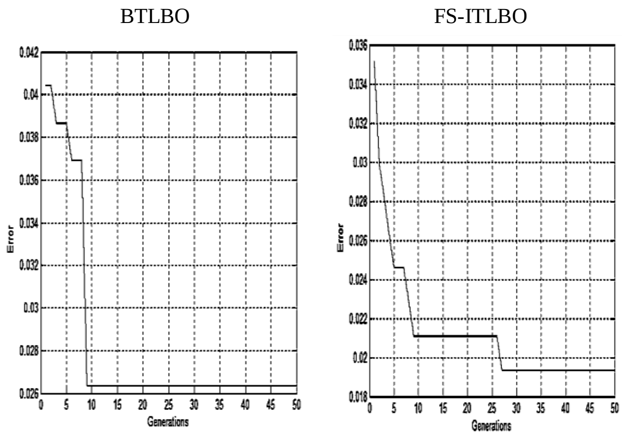


Figure 8. The analogy of convergences with the Naïve Bayes model.

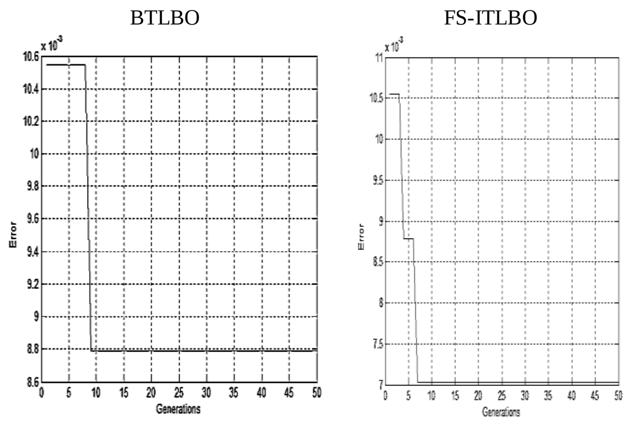


Figure 9. The analogy of convergences with the SVM model.

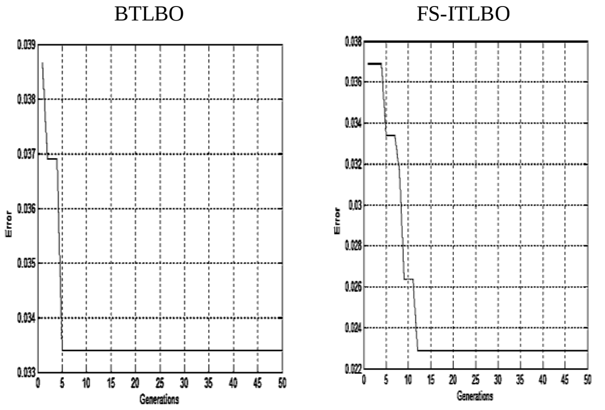


Figure 10. The analogy of convergences with the k-NN model.

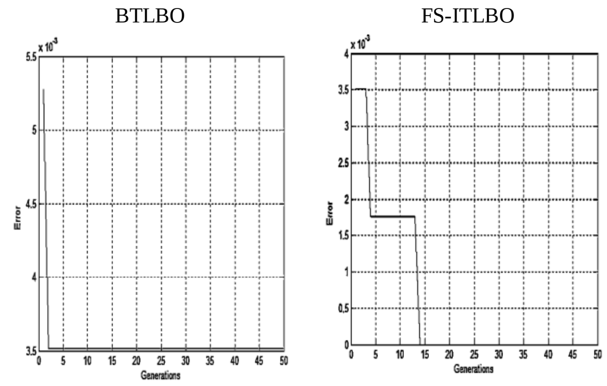


Figure 11. The analogy of convergences with the decision tree model.

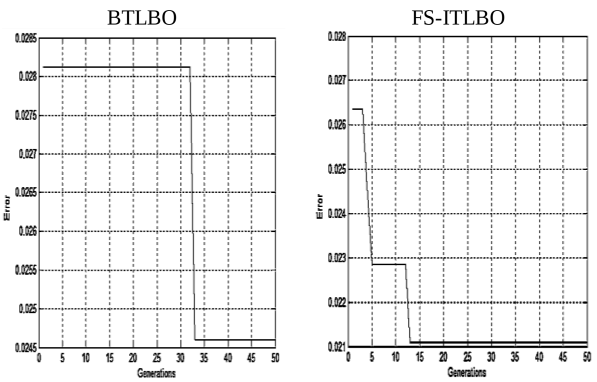


Figure 12. The analogy of convergences with the LDA model.

4.3. Performance Evaluation of FS-ITLBO on Parkinson's Disease Dataset

Table 5 demonstrates the prediction accuracies generated by several classification models with the overall attributes of PD dataset and ITLBO selected attributes. The results of ITLBO are compared with BTLBO for a better understanding of the improvement in the performance of the feature selection process. The ITLBO algorithm has been evaluated on parkinson's disease dataset in terms of classification results.

Table 5. The analogy of classification accuracies for BTLBO, and ITLBO algorithms on Parkinson's Disease dataset.

Classification Algorithm	All Features	BTLBO Features	ITLBO Features
Naive Bayes	82.11	87.41	89.41
SVM	84.76	87.07	92.71
k-NN	64.23	75.49	81.45
Decision Tree	79.47	88.74	89.40
LDA	52.31	85.43	91.39

The highest classification accuracy (92.71%) has been attained with the SVM classifier. The subset of features generated by ITLBO algorithm achieved only a nominal performance hike with naïve bayes and decision tree models. k-NN classifier also provided better accuracy value (81.45%) than the BTLBO as well as all features in the dataset. Finally, the Discriminant Analysis classifier achieved the highest difference in accuracy when compared with the overall attributes. ITLBO selected a minimal number of optimal features compared to the BTLBO algorithm and boosted the prediction responses for all the classification models as shown in Table 6.

Table 6. Comparison of features selected by the FS-ITLBO model on Parkinson's data.

Classification models	BTLBO Features	ITLBO Features
Naive Bayes	388	150
SVM	366	275
k-NN	331	256
Decision Tree	364	190
LDA	381	206

The FS-ITLBO model selected the least subset (150) out of 754 features using the naïve bayes classifier and a considerable number of features (190) with the help of the decision tree classifier.

5. Conclusions

In this paper, an integrative teaching learning based optimization algorithm is proposed for ideal feature selection. The proposed method uses multiple teachers in each iteration to explore maximum solution space. FS-ITLBO algorithm accomplished better accuracy responses compared to GA and BTLBO based FS on WDBC and PD datasets. FS-ITLBO algorithm can be applied to different datasets to keep down both the training time and error rate of a classification model by selecting optimal features. We would like to experiment with the "parameter tuning" for classification models with FS-ITLBO for achieving improved results in the future.

References

[1] Allam M. and Nandhini M., "A Study on Optimization Techniques in Feature Selection for Medical Image Analysis," *International Journal on Computer Science and Engineering*, vol. 9, no. 3, pp. 75-82, 2017.

[2] Allam M. and Nandhini M., "Feature Optimization using Teaching Learning Based Optimization for Breast Disease Diagnosis," *International Journal of Recent Technology and Engineering*, vol. 7, no. 4, pp. 78-85, 2018.

[3] Allam M. and Nandhini M., "Optimal Feature Selection Using Binary Teaching Learning Based Optimization Algorithm," *Journal of King Saud University-Computer and Information Sciences*, pp.1-13, 2018.

[4] Arora S. and Anand P., "Binary Butterfly Optimization Approaches for Feature Selection," *Expert Systems with Applications*, vol. 116, pp. 147-160, 2019.

[5] Bhardwaj H., Sakalle A., Tiwari A., Verma M., and Bhardwaj A., "Breast Cancer Diagnosis using Simultaneous Feature Selection and Classification: A Genetic Programming Approach," *in Proceedings of IEEE Symposium Series on Computational Intelligence*, Bangalore, pp. 2186-2192, 2018.

[6] Chandrasekhar A. and Rani T., "Storage and Retrieval of Large Data Sets: Dimensionality Reduction and Nearest Neighbour Search," *in Proceedings of International Conference on Contemporary Computing*, Noida, pp. 262-272, 2012.

[7] Dua D. and Graff C., "UCI Machine Learning Repository: Breast Cancer Wisconsin Data Set (Diagnostic)," Irvine, CA: University of California, 2019.

[8] Guyon I. and Elisseeff A., "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157-1182, 2003.

[9] Hafez A., Zawbaa H., Emery E., Mahmoud H., and Hassanien A., "An Innovative Approach for Feature Selection Based on Chicken Swarm Optimization," *in Proceedings of 7th International Conference of Soft Computing and Pattern Recognition*, Fukuoka, pp. 19-24, 2015.

[10] Hossam A., Harb H., and Kader H., "A Sub-Optimum Feature Selection Algorithm for Effective Breast Cancer Detection Based on Particle Swarm Optimization," *IOSR Journal of Electronics and Communication Engineering*, vol. 13, no. 3, pp. 01-12, 2018.

[11] Hua J., Tembe W., and Dougherty E., "Performance of Feature-Selection Methods in The Classification of High-Dimension Data," *Pattern Recognition*, vol. 42, no. 3, pp. 409-424, 2009.

[12] Janecek A., Gansterer W., Demel M., and Ecker G., "On the Relationship Between Feature Selection and Classification Accuracy," *Proceedings of Machine Learning Research*, vol. 4, pp. 90-105, 2008.

[13] Kiziloz H., Deniz A., Dokeroglu T., and Cosar A., "Novel Multiobjective TLBO Algorithms for The Feature Subset Selection Problem," *Neurocomputing*, vol. 306, pp. 94-107, 2018.

[14] Kohavi R. and John G., "Wrappers for Feature Subset Selection," *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273-324, 1997.

[15] Li J. and Liu H., "Challenges of Feature Selection for Big Data Analytics," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 9-15, 2017.

[16] Mafarja M., Eleyan D., Jaber I., Hammouri A., and Mirjalili S., "Binary Dragonfly Algorithm for Feature Selection," *in Proceedings of International Conference on New Trends in Computing Sciences*, Amman, pp. 12-17, 2017.

[17] Moradi P. and Gholampour M., "A Hybrid Particle Swarm Optimization for Feature Subset Selection by Integrating a Novel Local Search Strategy," *Applied Soft Computing*, vol. 43, pp. 117-130, 2016.

[18] Oluleye B., Leisa A., Leng J., and Dean D., "A Genetic Algorithm-Based Feature Selection," *International Journal of Electronics Communication and Computer Engineering*, vol. 5, no. 4, pp. 899-905, 2014.

[19] Panda M., "Elephant Search Optimization Combined with The Deep Neural Network for Microarray Data Analysis," *Journal of King Saud University-Computer and Information Sciences*, 2017.

[20] Rajamohana S. and Umamaheswari K., "Hybrid Approach of Improved Binary Particle Swarm Optimization and Shuffled Frog Leaping for Feature Selection," *Computers and Electrical Engineering*, vol. 67, pp. 497-508, 2018.

[21] Rao R., "Review of Applications of TLBO Algorithm and A Tutorial for Beginners to Solve The Unconstrained and Constrained Optimization Problems," *Decision Science Letters*, vol. 5, no. 1, pp. 1-30, 2016.

[22] Rao R., *Teaching Learning Based Optimization Algorithm and its Engineering Applications*, Springer International Publishing, 2016.

[23] Rao R. and Patel V., "An Improved Teaching-Learning-Based Optimization Algorithm for Solving Unconstrained Optimization Problems," *Scientia Iranica*, vol. 20, no. 3, pp. 710-720, 2013.

[24] Rao R., Savsani V., and Vakharia D., "Teaching-Learning-Based Optimization: A Novel Method for Constrained Mechanical Design Optimization Problems," *Computer-Aided Design*, vol. 43, no. 3, pp. 303-315, 2011.

[25] Rodrigues D., Pereira L., Nakamura R., Costa A., Yang X., Souza A., and Papa J., "A Wrapper Approach for Feature Selection Based on Bat Algorithm and Optimum-Path Forest," *Expert

Systems with Applications, vol. 41, no. 5, pp. 2250-2258, 2014.

[26] Sakar C., Serbes G., Gunduz A., Tunc H., Nizam H., Sakar B., Tutuncu M., Aydin T., Isenkul M., and Apaydin H., "A Comparative Analysis of Speech Signal Processing Algorithms for Parkinson's Disease Classification and the Use of The Tunable Q-Factor Wavelet Transform," *Applied Soft Computing*, vol. 74, 2019.

[27] Sayed G., Hassanien A., and Azar A., "Feature Selection Via A Novel Chaotic Crow Search Algorithm," *Neural Computing and Applications*, vol. 31, no. 1, pp. 171-188, 2019.

[28] Sayed G., Khoriba G., and Haggag M., "A Novel Chaotic Salp Swarm Algorithm for Global Optimization and Feature Selection," *Applied Intelligence*, vol. 48, no. 10, pp. 3462-3481, 2018.

[29] Selvarajan D., Jabar A., and Ahmed I., "Comparative Analysis of PSO and ACO Based Feature Selection Techniques for Medical Data Preservation," *The International Arab Journal of Information Technology*, vol. 16, no. 4, pp. 731-736, 2019.

[30] Shahbeig S., Helfroush M., and Rahideh A., "A Fuzzy Multi-Objective Hybrid TLBO-PSO Approach to Select the Associated Genes with Breast Cancer," *Signal Processing*, vol. 131, pp. 58-65, 2016.

[31] Sridevi T. and Murugan A., "A Novel Feature Selection Method for Effective Breast Cancer Diagnosis and Prognosis," *International Journal of Computer Applications*., vol. 88, no. 11, pp. 28-33, 2014.

[32] Tan K., Teoh E., Yu Q., and Goh K., "A Hybrid Evolutionary Algorithm for Attribute Selection in Data Mining," *Expert Systems with Applications*, vol. 36, no. 4, pp. 8616-8630, 2009.

[33] Tuo S., Yong L., Deng F., Li Y., Lin Y., and Lu Q., "HSTLBO: A Hybrid Algorithm Based on Harmony Search and Teaching-Learning Based Optimization for Complex High Dimensional Optimization Problems," *PLoSONE*, vol. 12, no. 4, pp. 1-23, 2017.

Mohan Allam, Research Scholar in the Department of Computer Science, Pondicherry University, Puducherry, India, and working as Assistant Professor at Shri Vishnu Engineering College for Women. His research interests include Soft Computing and Image Processing.

Nandhini Malaiyappan, Associate Professor in the Department of Computer Science, Pondicherry University, Puducherry, India. Her research interests include Artificial Intelligence, Software Engineering, Evolutionary Algorithms, and Combinatorial Problem Optimization.