Journal of King Saud University – Computer and Information Sciences (2018) 30, 120–139

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

A TLBO based gradient descent learning-functional link higher order ANN: An efficient model for learning from non-linear data

Bighnaraj Naik [a,*], Janmenjoy Nayak [b], H.S. Behera [a]

[a] *Department of Computer Science Engineering & Information Technology, Veer Surendra Sai University of Technology (Formerly University College of Engineering), Burla 768018, Odisha, India*
[b] *DST INSPIRE Fellow, Govt. of India, New Delhi, India*

Received 24 May 2015; revised 11 December 2015; accepted 6 January 2016
Available online 28 March 2016

KEYWORDS

Teaching–learning based optimization;
Functional link artificial neural network;
Gradient descent learning;
Classification;
Data mining

Abstract All the higher order ANNs (HONNs) including functional link ANN (FLANN) are sensitive to random initialization of weight and rely on the learning algorithms adopted. Although a selection of efficient learning algorithms for HONNs helps to improve the performance, on the other hand, initialization of weights with optimized weights rather than random weights also play important roles on its efficiency. In this paper, the problem solving approach of the teaching learning based optimization (TLBO) along with learning ability of the gradient descent learning (GDL) is used to obtain the optimal set of weight of FLANN learning model. TLBO does not require any specific parameters rather it requires only some of the common independent parameters like number of populations, number of iterations and stopping criteria, thereby eliminating the intricacy in selection of algorithmic parameters for adjusting the set of weights of FLANN model. The proposed TLBO-FLANN is implemented in MATLAB and compared with GA-FLANN, PSO-FLANN and HS-FLANN. The TLBO-FLANN is tested on various 5-fold cross validated benchmark data sets from UCI machine learning repository and analyzed under the null-hypothesis by using Friedman test, Holm's procedure and post hoc ANOVA statistical analysis (Tukey test & Dunnett test).

1. Introduction

Classification and forecasting tasks are typical process of learning and knowledge discovery from data in the field of data mining. Many applications of classification tasks have been reported in recent years from emerging areas of science and engineering (Yang et al., 2013; Hajmohammadi et al., 2014; He et al., 2014; Uysal and Gunal, 2014; Tolambiya et al., 2010; Mei et al., 2014; Kianmehr et al., 2010; Gillies

* Corresponding author. Tel.: +91 9439152272.
E-mail address: bnaik_mca@vssut.ac.in (B. Naik).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jksuci.2016.01.001
1319-1578

Table 1 Previous literatures on teaching learning based optimization (TLBO).

Reference	Contribution	Area of application	Year
Toğan (2012)	TLBO	Engineering design	2012
Niknam et al. (2012)	MOTLBO	Economic dispatch	2012
Zou et al. (2013)	TLBO	Multi-objective optimization	2013
Roy et al. (2013)	QOTLBO	Hydro thermal scheduling	2013
Mandal and Roy (2013)	QOTLBO	Power dispatch	2013
García and Mena (2013)	MTLBO	Distributed generation	2013
Venkata Rao and Kalyankar (2013)	TLBO	Machining processes	2013
Roy (2013)	TLBO	Scheduling problem	2013
Roy and Bhui (2013)	QOTLBO	Load dispatch	2013
Singh et al. (2013)	TLBO	Power system	2013
Satapathy et al. (2013a)	WTLBO	Optimization	2013
Satapathy et al. (2013b)	OTLBO	Optimization	2013
Tuo et al. (2013)	HSTL	Optimization	2013
Theja et al. (2013)	TLBO	Power system	2013
Sultana and Roy (2014)	TLBO	Optimal capacitor placement	2014
Azizipanah-Abarghooee et al. (2014)	Gradient based modified TLBO with Black hole	Scheduling of thermal power systems	2014
Arya and Koshti (2014)	TLBO	Load shedding	2014
Khalghani and Khoob (2014)	TLBO	Power quality	2014
Niu et al. (2014)	STLBO	Fuel and solar cell models	2014
Moghadam and Seifi (2014)	Fuzzy-TLBO	Energy loss minimization	2014
Cheng (2014)	TLBO	Temperature calculations	2014
Barisal (2015)	TLBO	Load frequency control	2015
Ghasemi et al. (2015a)	GBTLBO	Power dispatch problem	2015
Chen et al. (2015)	ITLBO	Optimization	2015
Sahu et al. (2016)	TLBO	Power system	2015
Ghasemi et al. (2015b)	ITLBO	Power flow	2015

et al., 2013; Lai and Garibaldi, 2013), which has given motivation and direction to the application of the classification task in data mining. Although a good number of traditional classification methods are proposed by many researchers (Quinlan, 1993; Yung and Shaw, 1995; Hamamoto et al., 1997; Yager, 2006), for the first time, Zhang (2000) realized that artificial neural network (ANN) models are alternatives to various conventional classification methods which are based on statistics. ANNs are capable of generating complex mapping between input and the output space and thus they can form arbitrarily complex nonlinear decision boundaries. Along the way, there are already several artificial neural networks, each utilizing a different form of learning or hybridization. As compared to higher order neural network, classical neural networks (Example: MLP) are suffering from slow convergence and unable to automatically decide the optimal model of prediction for classification. In the last few years, to overcome the limitations of conventional ANNs, some researchers have focused on higher order neural network (HONN) models (Redding et al., 1993; Goel et al., 2006) for better performance.

In this paper, it is an attempt to design HONN model with competitive learning based on a new meta-heuristic optimization algorithm for classification of benchmark data sets considered from well known machine learning data repository. The encouraging results and absence of the typical algorithm dependent parameters of the recently developed TLBO algorithm (Rao et al., 2011) inspired us to develop the model: TLBO-FLANN: a hybrid TLBO based gradient descent learning-functional link higher order ANN (FLANN) model for learning from non-linear data.

2. Literature survey

A good number of research papers is reported on various FLANN models and its application in classification, prediction and forecasting in recent years. Also excellent efforts are made to improve the performance of these FLANN models by using various optimization techniques. This section presents various previous works reported on FLANN models in classification, prediction and forecasting in data mining. A classification method by using FLANN with least complex architecture as compared to multilayer perceptron is anticipated by Mishra and Dehuri (2007) and the proposed method is found to be efficient in terms of ability of handling linearly non-separable classes by increasing the dimension of input space. In most cases, the performance and processing time of FLANN model is found to be better than the other models. A survey on FLANN is made and PSO based back propagation learning is proposed by Dehuri and Cho (2009). The basic concept of FLANN, associated basis functions, learning schemes and development of FLANNs over time are discussed in this paper. Also authors have proposed a Chebyshev-FLANN with hybrid PSO-back propagation learning for classification. The proposed model is proven to be better as compared to FLANN by testing with benchmark data sets. Patra et al. (2009) suggested an efficient FLANN model for making stock price prediction of the closing price of US stocks and the proposed model is compared with MLP-based prediction model through several experiments. This proposed trigonometric FLANN has shown superior performance over MLP by making more accurate predictions of stock prices. Prediction of the causing genes in gene diseases by FLANN

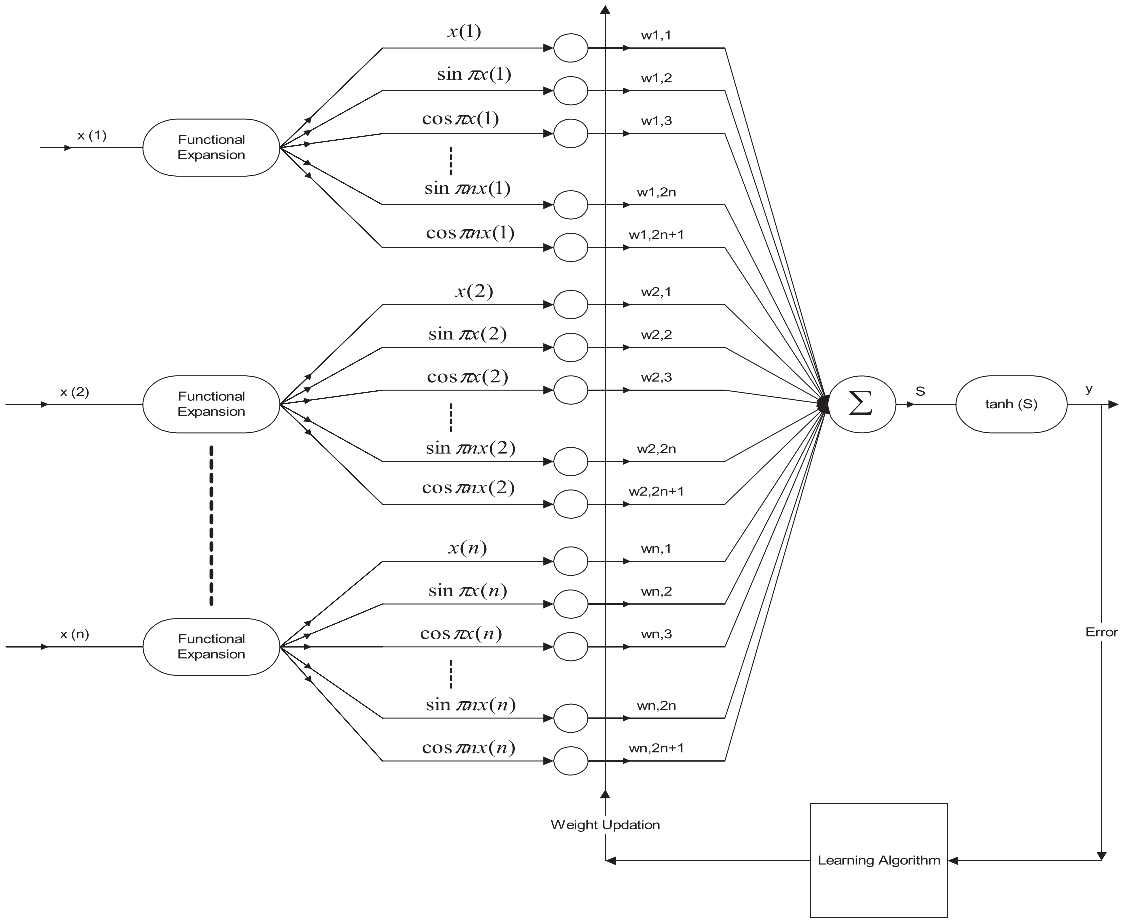

Figure 1 Functional link higher order artificial neural network architecture.

Table 2 Various notations and symbols used in algorithmic design of proposed scheme.

Notation used	Meaning
X	Population of weight-sets
x_i	Each individual weight-set in X
$w_{i,j}$	jth weight value of ith weight-set (x_i)
RMSE	Root mean square error
F_{x_i}	Fitness of x_i (ith weight-set in X)
X_{mean}	Mean of the population X
$x_{teacher}$	Selected weight-set as teacher
T_F	Teaching Factor
X(i)	ith weight-set of the population X
$X_{next}(i)$	ith weight-set of the next population X
Φ	Functionally expanded input values
T	Target values
μ	Gradient descent learning parameter
Y(k)	Output of FLANN model for kth pattern
t(k)	Target value of kth pattern
e(k)	Error value of kth pattern
ΔW	Error term of the network

model is proposed by Sun et al. (2009) and compared with multilayer perceptron (MLP) and support vector machines (SVM). In this study, three classifiers (i.e. MLP, SVM, FLANN) have been implemented and the performance of FLANN classifier is found to be better than MLP and SVM. Chakravarty and Dash (2009) have proposed functional link neural fuzzy (FLNF) model to predict the stock market indices and compared with FLANN model in terms of root mean square error. The simulation results have been demonstrated to claim better performance of FLNF over FLANN. Also the local minima problem which may arise in back propagation learning algorithm is addressed by using particle swarm optimization (PSO). A FLANN based model with least mean square and recursive least square learning algorithm is employed by Majhi et al. (2009) for forecasting of the stock markets. Between least mean square learning based FLANN and recursive least square learning FLANN, the recursive least square learning FLANN is found to be efficient and has less computational complexity. A compact and accurate hybrid FLANN classifier (HFLNN) has been proposed by Dehuri and Cho (2010) by selecting an optimal subset of favorable input features by eliminating features with fewer or no predictive information and this method is found to be better as compared to FLANN and RBFN. The MLP, FLANN and PSO-FLANN classification models are used and tested by Mishra et al. (2012) for classification of biomedical data. In this paper, an efficient dynamic classifier fusion (DCF) has been proposed along with principal component analysis (PCA) scheme which is used to extract important input features. Extracted features are then supplied to LMS classifiers, with learning based on back propagation and PSO algorithm. Although MLP is a traditional ANN, surprisingly, in this

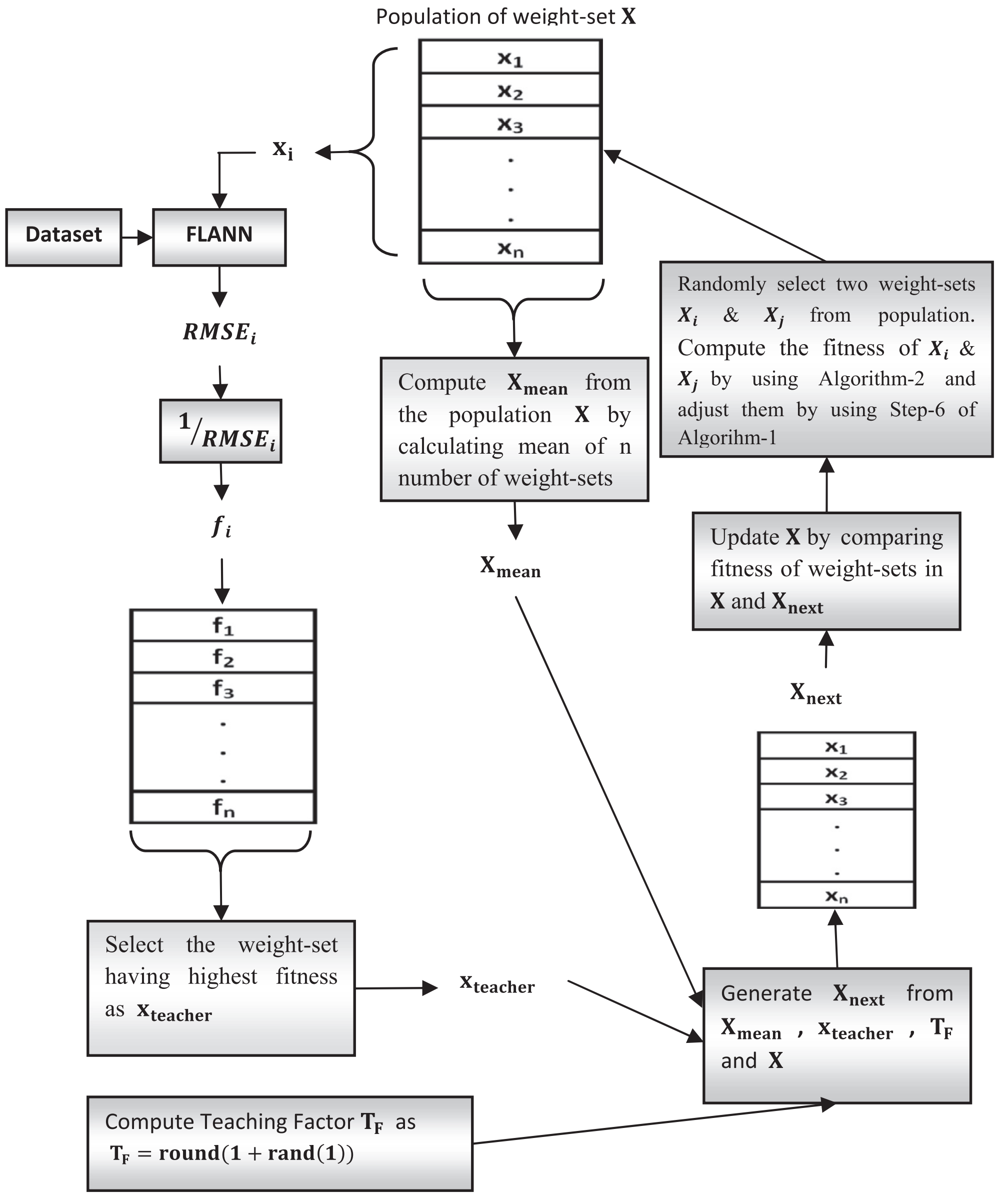

Figure 2 Working procedure of the proposed scheme.

study, they found MLP is better as compared to FLANN and PSO-FLANN. Dehuri et al. (2012) have proposed an IPSO (improved PSO) based FLANN classifier (IPSO-FLANN) and compared with MLP, support vector machine (SVM), RBFN, FLANN with gradient descent learning and fuzzy swarm net (FSN) model. Initially, the set of weight values of FLANN are optimized using IPSO and then those are supplied to FLANN classifier along with functionally expanded (using trigonometric basis functions) input patterns for classification. The ISO-FLANN is simple in architecture and found to be better as compared to MLP, SVM, FLANN with gradient descent learning and FSN. An efficient classification method based on FLANN and a hybrid learning scheme based on PSO and GA have been proposed by Naik et al. (2015a) and it is found to be relatively better in performance as compared to other alternatives. The PSO, GA and the gradient descent search are used iteratively to adjust the parameters of FLANN until the error is less than the required value, which helps the FLANN model to get better classification accuracy. Naik et al. (2015b) have designed a honey bee mating optimization (HMBO) based learning scheme for FLANN classifier and compared with FLANN, GA based FLANN, PSO based FLANN and HS based FLANN classification model. The proposed method mimics the iterative mating process of honey bees and strategies to select eligible drones for the mating process, selection of best weights for FLANN classifiers. A novel FLANN with a harmony search (HS) algorithm-based GDL method is proposed by Naik et al. (2014), which requires less mathematical computation as compared to other similar methods. The HS based FLANN (HS-FLANN) is more efficiently able to classify the data than PSO-based FLANN and GA based FLANN.

In this paper, a FLANN model with hybrid TLBO based gradient descent learning (GDL) method for learning from non-linear data for classification task has been proposed and compared with previously available alternatives.

Table 3 Details on data sets used for experiments.

Data sets	Number of pattern	Number of features (excluding class label)	Number of classes
Wine	178	13	03
Iris	150	04	03
Hayesroth	160	04	03
Monk 2	256	06	02
Ionosphere	351	33	02
Hepatitis	80	19	02
Pima	768	08	02
New Thyroid	215	05	03
Bupa	345	06	02
Dermatology	256	34	06
Heart	270	13	02

3. Design issues identification

All the FLANN models discussed in the literature survey implement some form of learning methods which learns from past data in forecasting and classification tasks in data mining. Almost all HONNs including FLANN are sensitive to random initialization of weights and rely on the learning algorithm adopted. Although a selection of efficient learning algorithms for HONNs helps to improve the performance, initialization of weights with optimized weights rather than random weights also play important roles in efficiency of HONNs. In the literature survey, it was noticed that, almost all the previously published works have addressed the issue of random initialization of weights in FLANN by using various optimization algorithms like genetic algorithm (GA) (Holland, 1992; Goldberg, 1989), particle swarm optimization (PSO) (Kennedy et al., 1995), harmony search (HS) (Geem et al., 2001), honey-bee mating optimization (HBMO) (Abbass et al., 2001) etc. In these papers, the optimization algorithms (GA, PSO, Improved PSO, HS, HMBO etc.) are used to select the best set of weights for FLANN models for various applications like prediction, classification and forecasting. Although it is reported that these optimization techniques are successfully used in FLANN models for implementation in improved models like GA based FLANN (GA-FLANN) (Dehuri et al., 2012), PSO based FLANN (PSO-FLANN) (Dehuri et al., 2012), IPSO based FLANN (IPSO-FLANN) (Dehuri et al., 2012), HBMO based FLANN (HBMO-FLANN) (Naik et al., 2015b) and HS based FLANN (HS-FLANN) (Naik et al., 2014), some weakness of these implementations are the requirement of various algorithmic controlling parameters like: (i) selection of mutation rate and type of crossover operator in GA in GA-FLANN, (ii) choosing of appropriate values of inertia weight (λ), coefficient c1 and c2 in PSO in PSO-FLANN and IPSO-FLANN (iii) types of crossover operator selection and choosing drone and worker ratio in HBMO in HBMO-FLANN and (iv) selection of pitch adjustment rate, harmony memory consideration rate and bandwidth in HS in HS-FLANN. The performances of these models are dependent upon such controlling parameters and any changes in these parameters may not only lead to increase the effort to develop the program but also the time and space complexity of the algorithm increases.

Keeping this in view, a new teaching and learning inspired algorithm (TLBO) is used in FLANN learning model with gradient descent learning (GDL) (Rumelhart et al., 1986) scheme for classification. This paper attempts to address the intricacy in adjusting the set of weights of the FLANN model by using an appropriate learning algorithm. Here the problem solving approach of the TLBO along with learning ability of the GDL is used to obtain the optimal set of weights of FLANN model. In this paper, a FLANN model has been designed with TLBO algorithm which does not require any algorithmic specific parameters, rather it requires some of the common independent parameters like number of populations, number of

Table 5 Comparison between GA-FLANN, PSO-FLANN, HS-FLANN and TLBO-FLANN models based on the fitness.

Datasets	Fitness obtained by various hybrid models			
	GA-FLANN	PSO-FLANN	HS-FLANN	TLBO-FLANN
Wine	2.462398	2.462398	2.534043	3.057658
Iris	5.972679	5.97268	5.97268	9.411738
Hayesroth	1.888252	1.869618	1.895258	3.118014
Monk 2	2.024735	2.033361	2.037116	3.057657
Ionosphere	1.677268	1.696773	1.67727	2.382388
Hepatitis	2.58331	2.813839	3.276094	2.349537
Pima	2.216361	2.216474	2.217241	2.220777
New Thyroid	1.888252	1.869618	2.675806	3.118014
Bupa	1.54692	1.547116	1.548419	2.162731
Dermatology	1.888967	1.981951	2.267907	3.690652
Heart	2.128121	2.127275	2.128124	2.696475

Table 4 Data sets prepared in 5-fold cross validation for training and testing phase.

Dataset	Data files	Number of Pattern	Task	Number of Pattern in class-1	Number of Pattern in class-2	Number of Pattern in class-3
New Thyroid	Newthyroid-5-1tra.dat	172	Training	120	28	24
	Newthyroid-5-1tst.dat	43	Testing	30	07	06
	Newthyroid-5-2tra.dat	172	Training	120	28	24
	Newthyroid-5-2tst.dat	43	Testing	30	07	06
	Newthyroid-5-3tra.dat	172	Training	120	28	24
	Newthyroid-5-3tst.dat	43	Testing	30	07	06
	Newthyroid-5-4tra.dat	172	Training	120	28	24
	Newthyroid-5-4tst.dat	43	Testing	30	07	06
	Newthyroid-5-5tra.dat	172	Training	120	28	24
	Newthyroid-5-5tst.dat	43	Testing	30	07	06

Table 6 Comparison between GA-FLANN, PSO-FLANN, HS-FLANN and TLBO-FLANN models based on average fitness on 11 numbers of data sets.

	Various hybrid models			
	GA-FLANN	PSO-FLANN	HS-FLANN	TLBO-FLANN
Average RMSE on 11 data sets	0.473349	0.468362	0.442128	0.345602

Table 7 Friedman's ranks of various models on the data sets based on the maximum fitness obtained.

Datasets	Fitness obtained by various hybrid models			
	GA-FLANN	PSO-FLANN	HS-FLANN	TLBO-FLANN
Wine	2.462398 (3)	2.462398 (3)	2.534043 (2)	3.057658 (1)
Iris	5.972679 (3)	5.97268 (2)	5.97268 (2)	9.411738 (1)
Hayesroth	1.888252 (3)	1.869618 (4)	1.895258 (2)	3.118014 (1)
Monk 2	2.024735 (4)	2.033361 (3)	2.037116 (2)	3.057657 (1)
Ionosphere	1.677268 (4)	1.696773 (2)	1.67727 (3)	2.382388 (1)
Hepatitis	2.58331 (3)	2.813839 (2)	3.276094 (1)	2.349537 (4)
Pima	2.216361 (4)	2.216474 (3)	2.217241 (2)	2.220777 (1)
New Thyroid	1.888252 (3)	1.869618 (4)	2.675806 (2)	3.118014 (1)
Bupa	1.54692 (4)	1.547116 (3)	1.548419 (2)	2.162731 (1)
Dermatology	1.888967 (4)	1.981951 (3)	2.267907 (2)	3.690652 (1)
Heart	2.128121 (3)	2.127284 (4)	2.128124 (2)	2.696475 (1)
Friedman's rank in average	3.455	3	2	1.272

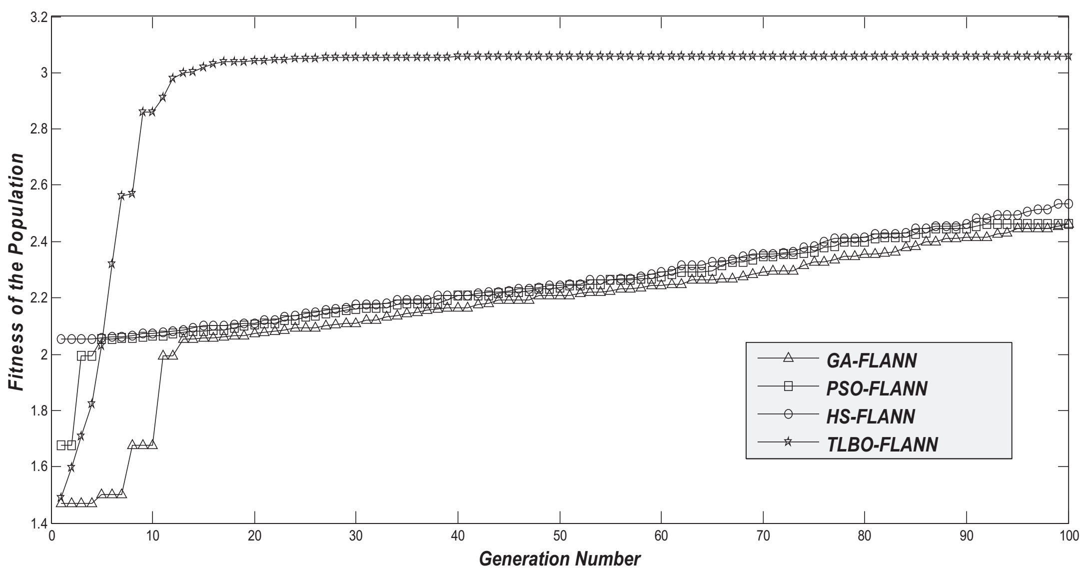

Figure 3 Observation on improvements in fitness of population in various iterations on Wine dataset.

iterations and stopping criteria (like other optimization techniques). Starting from the development of TLBO, it has been of keen interest among the diversified researchers and has been used in various real life applications (Vedat Toğan, 2012; Niknam et al., 2012; Zou et al., 2013; Roy et al., 2013; Mandal and Roy, 2013; García and Mena, 2013; Venkata Rao and Kalyankar, 2013; Roy, 2013; Roy and Bhui, 2013; Singh et al., 2013; Satapathy et al., 2013a,b; Tuo et al., 2013; Theja et al., 2013; Sultana and Roy, 2014; Azizipanah-Abarghooee et al., 2014; Arya and Koshti, 2014; Khalghani and Khoob, 2014; Niu et al., 2014; Moghadam and Seifi, 2014; Cheng, 2014; Barisal, 2015; Ghasemi et al., 2015a; Chen et al., 2015; Sahu et al., 2016; Ghasemi et al., 2015b) (Table 1). Although Črepinšek et al. (2012) had raised some issues relating the performance of TLBO such as: (i) incorrect use of formulae to evaluate various fitness functions,(ii) controlling parameters, (iii) improper experiment settings etc.; later on Waghmare (2013) has correctly reexamined, commented and tested on the above raised issues on TLBO. He has not only tested the algorithm by considering a number of both constrained and unconstrained functions and found the algorithm outperforms the other existing evolutionary algorithms, but he has also been able to prove the misinterpretation and misconception created by Črepinšek et al. about the understanding point of view of TLBO . Also, he has properly justified the less controlling parameter required to supply for TLBO; in fact only the common controlling parameters are required during the experimental analysis of TLBO, which is claimed by testing the functions and their improved results. However, in this paper, after all the required experimental settings, proper tuning of the common parameters and from the experimental results, it is found that TLBO outperforms some other existing techniques.

The remaining of the paper is organized as follows: Section 4 describes the preliminaries like TLBO algorithm, concept of FLANN and GDL. Section 5 includes the proposed TLBO-FLANN learning model. Section 6 contains experimental setup & data preparation. Section 7 discusses the experimental study and result analysis of the proposed method. Section 8 outlines different statistical tests. Section 9 presents time complexity analysis. Section 10 concludes our work with the future development aspects followed by References.

4. Preliminaries

This section presents basic concepts of TLBO, FLANN and GDL which are the background for the development of the proposed model.

4.1. Teaching learning based optimization

TLBO (Rao et al., 2011) is a recently developed and competent optimization technique which is inspired by natural teaching–learning process in an education system. TLBO mimics the strategy used by a teacher to teach a group of students effectively and captures the effect of teacher on the learning process of the students. The working procedure of TLBO suggested by Rao et al., 2011 can be realized as follows:

Teacher phase
Step-1. Initialize population of students X (candidate solutions) randomly.
Step-2. Calculate the mean of each student in the population (Xmean).
Step-3. Compute the fitness of each student in the population and find out the best solution (Xteacher).
Step-4. Generate a new population by modifying the solutions in initial population based on best solution (teacher), mean of students in the population (mean) and teaching factorT_F.
  for i = 1:1: nos of student in the population X
    $T_F = \text{round}(1 + \text{rand}(1))$
    $X_i(\text{new}) = X_i(\text{old}) + \text{rand}(1)(X_{\text{teacher}} - T_F * X_{\text{mean}})$
  Endfor
Student phase
Step-5. Update population of student X by comparing fitness of students in old population X and new population Xnew.
  for i = 1:1: nos of student in the population X
    if (Xi(old) < Xi (new))
      Xi = Xi(new)
    Else
      Xi = Xi(old)
    endif
  endfor
Step-6. Randomly select two no. of student from the population and improvise them.
  Select ith and jth student X_i and X_j randomly from the population.
  If (fitness of X_i < fitness of X_j)
  $X_i(new) = X_i(old) + rand(1)(X_j - X_i)$
  Else
  $X_j(new) = X_j(old) + rand(1)(X_i - X_j)$
  Ifend
Step-7. Check for termination criteria. If reached stop. Else go to step-2.
Step-8. Exit

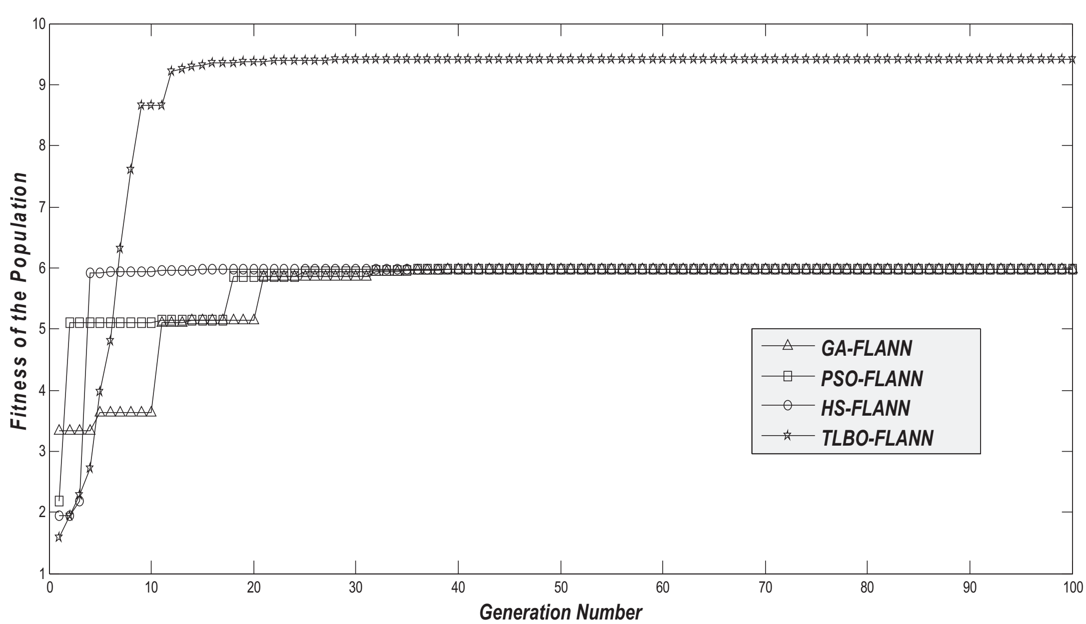

Figure 4 Observation on improvements in fitness of population in various iterations on Iris dataset.

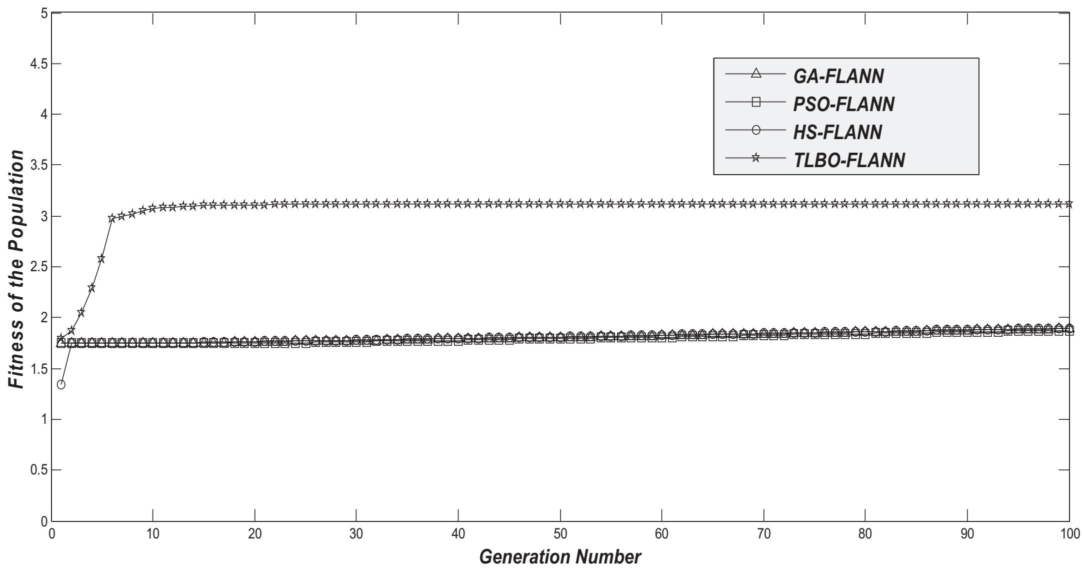

Figure 5 Observation on improvements in fitness of population in various iterations on Hayesroth dataset.

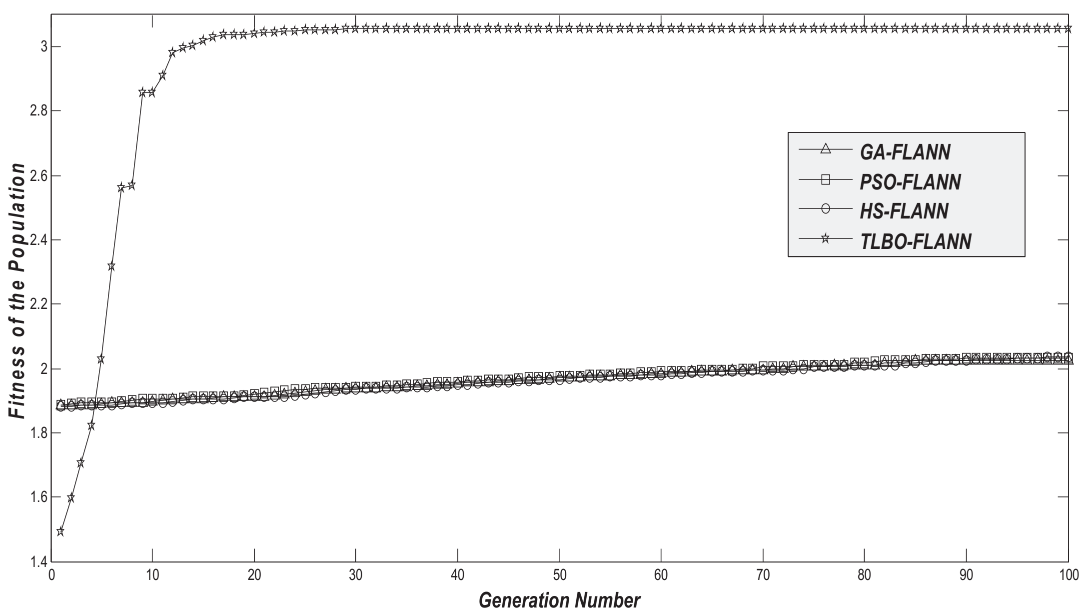

Figure 6 Observation on improvements in fitness of population in various iterations on Monk2 dataset.

4.2. Functional link artificial neural network

FLANN (Pao, 1989; Pao and Takefuji, 1992) is a class of higher order neural network that makes use of a higher combination of its inputs. Even if it has a single-layer network, still it is capable of handling non-linear separable classification tasks as compared to MLP. The error of FLANN is calculated based on net output and given target value. A suitable learning method can be adapted to adjust weight values of FLANN based on error of the network. Fig. 1 depicts the basic architecture of FLANN. The complete concepts of FLANN and its implementations may be found from recently published related work (Naik et al., 2015a, b, 2014).

4.3. Gradient descent learning

Gradient descent learning (GDL) (Rumelhart et al., 1986) is the most commonly used training method in which weights are changed in such a way that network error is declined as rapidly as possible. The complete implementations of GDL for FLANN can be found from recently published related work (Majhi et al., 2009; Naik et al., 2015a,b, 2014). Basically,

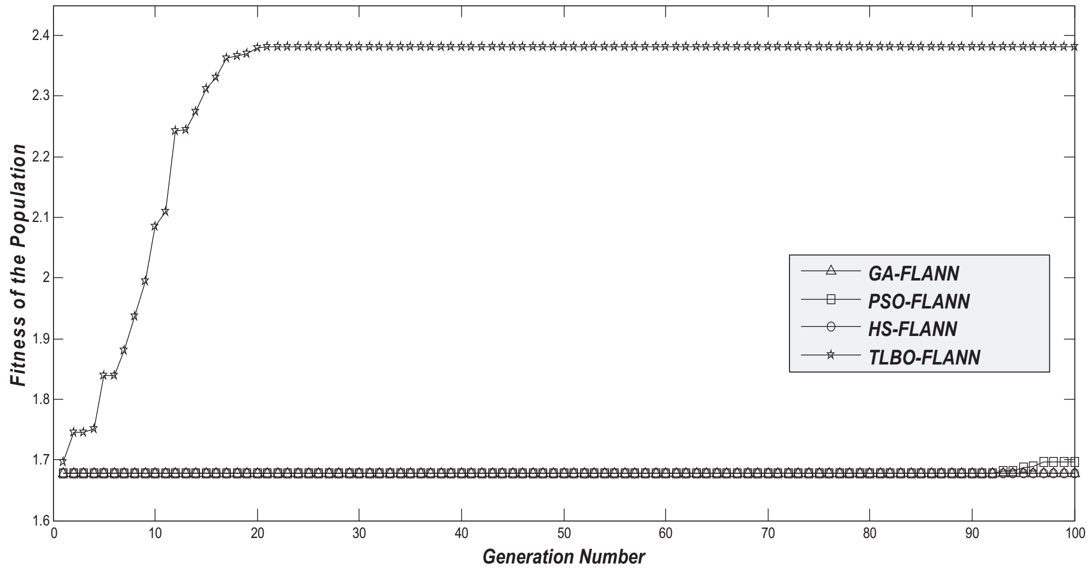

Figure 7 Observation on improvements in fitness of population in various iterations on Ionosphere dataset.

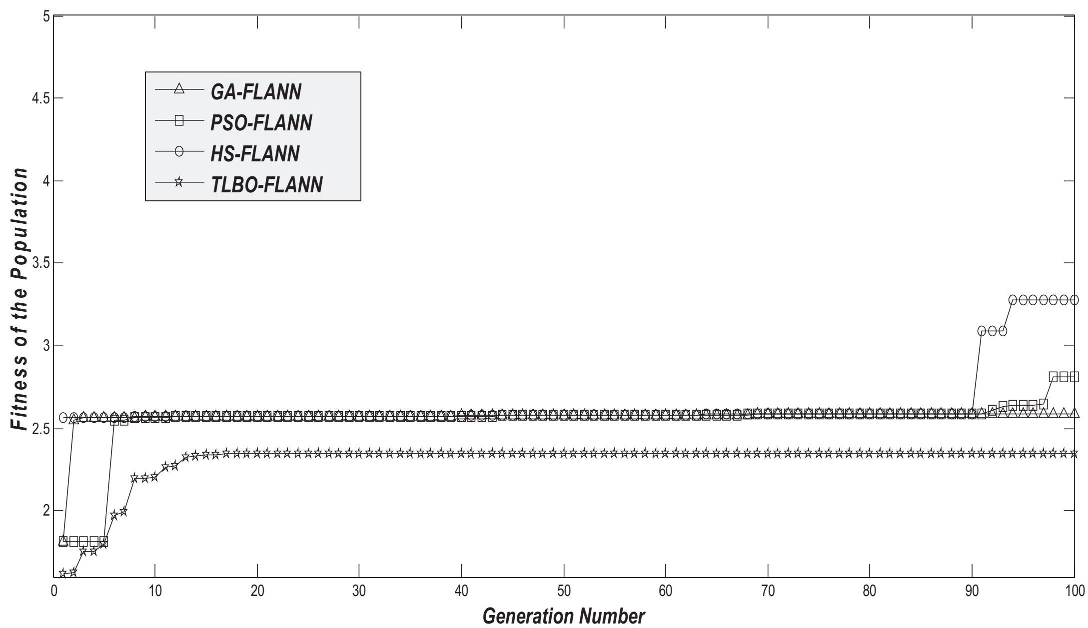

Figure 8 Observation on improvements in fitness of population in various iterations on Hepatitis dataset.

a better learning algorithm helps the ANN models for faster convergence. Further, a use of competitive optimization technique can improve the performance of a learning algorithm in terms of fast convergence and accuracy of the ANN model.

5. Proposed method

In this paper, a new metaheuristic algorithm called TLBO is used to obtain better FLANN learning model. The problem solving strategy of TLBO along with Gradient Descent Learning is used to find the best set of weights for FLANN model for classification task. The objective is to investigate performance of TLBO to enhance learning capability of FLANN classification model as compared to GA, PSO and HS. The pseudo codes developed during implementation of TLBO based FLANN (TLBO-FLANN) is presented in this section. In this paper, the simulation results and the comparisons of performance of these hybrid FLANN models (GA-FLANN, PSO-FLANN, HS-FLANN and the proposed TLBO based FLANN (TLBO-FLANN)) are represented and discussed. Table 2 represents various notations used for description of the proposed schemes and designing algorithms.

5.1. TLBO based gradient descent learning FLANN

Initially, the population of weight-sets **X** (population of students) is initialized with ‘n’ no.s of weight-sets for FLANN. Each weight-set in the population **X** is a vector of weights

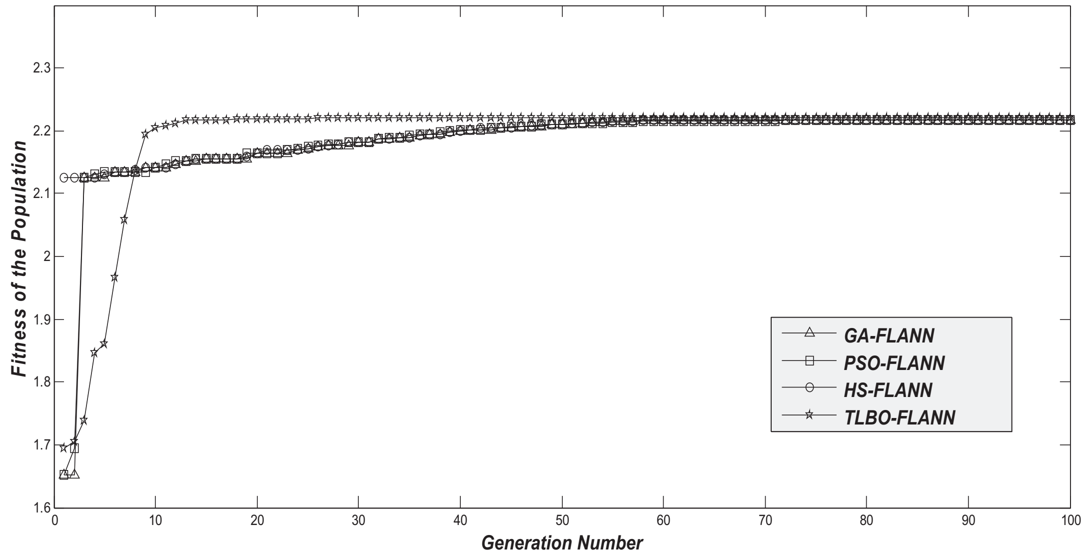

Figure 9 Observation on improvements in fitness of population in various iterations on Pima dataset.

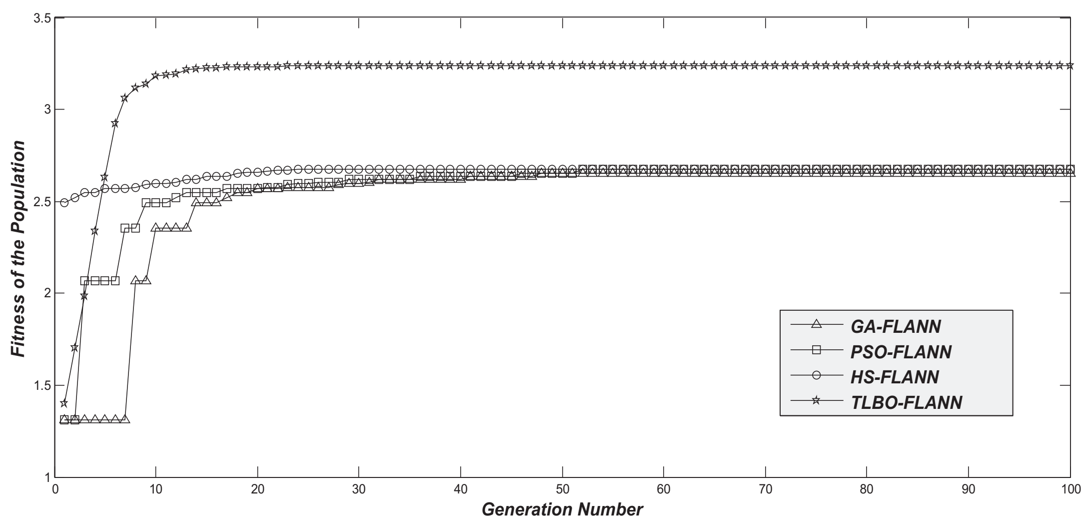

Figure 10 Observation on improvements in fitness of population in various iterations on New Thyroid dataset.

initialized randomly between −1 and 1 which are the potential candidate weight-sets of FLANN model of a particular data set. Each individual weight-set in **X** can be defined as:

$$x_i = (w_{i,1}, w_{i,2} \ldots w_{i,m\times a\times(2\times k+1)}) \tag{1}$$

$$\mathrm{X} = (x_1, x_2 \ldots x_\mathrm{n}) \tag{2}$$

In Eq. (1), the $(2 \times k + 1)$ is the No. of functionally expanded values for a single value in input pattern (for a chosen value of k), 'a' is the number of values (attributes) in a single input pattern, 'm' is the number of patterns in the data set and 'n' is the number of weight-sets in the population X. The set of weight-sets in X are represented as in Eq. (2). Here, the objective is to prune out optimal weight-sets for the FLANN model to obtain better classification accuracy. Each weight-set x_i is set to FLANN individually and the FLANN model is trained with a particular data set. Based on the output of FLANN and given target value, the error of the model is obtained. For a specific data set, the root mean square error (RMSE) (Eq. (3)) for each weight-set x_i is computed using output of the FLANN and given target value (Algorithm 2). Based on RMSE's, fitness of the weight-sets are computed by using Eq. (4). The RMSE of predicted output values $\widehat{\mathrm{y}}_\mathrm{i}$ of a target variable y_i is computed for 'n' different predictions as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$F_{x_i} = \frac{1}{RMSE_i} \tag{4}$$

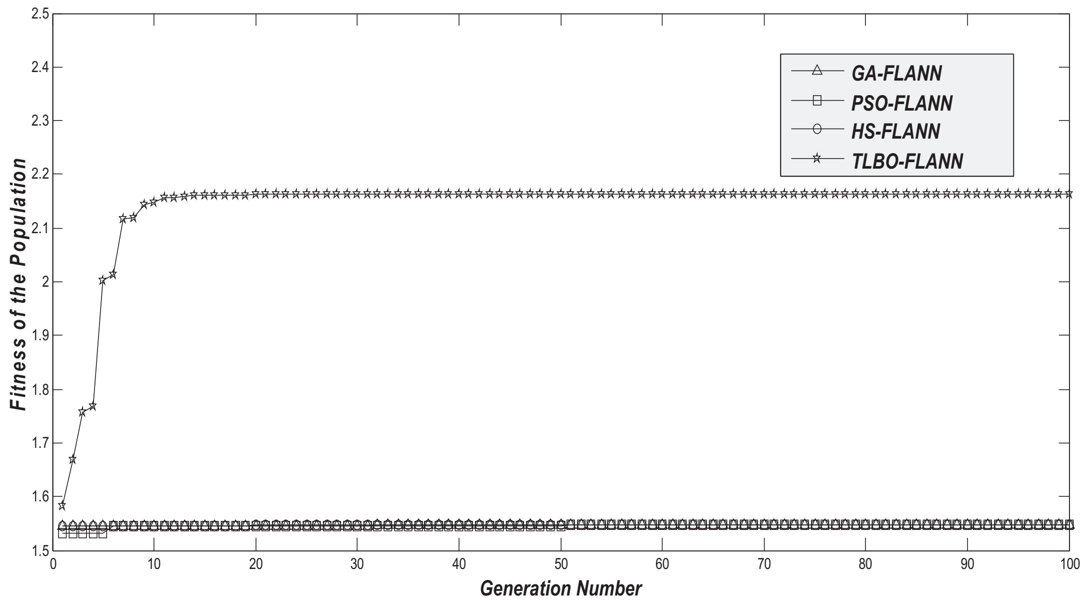

Figure 11 Observation on improvements in fitness of population in various iterations on Bupa dataset.

In Eq. (4), x_i is the ith weight-set in the population, $RMSE_i$ is the root mean square error of ith weight-set and F_{x_i} is the fitness of ith weight-set x_i.

After evaluation of fitness values for each weight-set in X, the weight-set with maximum fitness is selected as Teacher ($\mathbf{x_{teacher}}$). From the population of X, the mean of the weight-sets ($\mathbf{X_{mean}}$) is computed by calculating the mean of all the weight-sets in X. After the calculation of teaching factor ($\mathbf{T_F}$) (Algorithm-1), the next population $\mathbf{X_{next}}$ is generated from **X**, $\mathbf{X_{mean}}$, $\mathbf{x_{teacher}}$ and $\mathbf{T_F}$ (Step-4 of Algoritm-1). Then the weight-sets in initial population **X** are updated by comparing fitness of weight-sets in **X** and $\mathbf{X_{next}}$ (Step-5 of Algoritm-1). The resultant population of weight-sets **X**' goes through improvisation steps (Step-6 of Algoritm-1) in which two weight-sets are randomly selected from the population **X** and best among them are chosen as weight-set for the next generation $\mathbf{X_{next}}$ by comparing their fitness, thereby giving more chances to migrate better weight-sets for the next generation. These processes are continued until maximum iteration is reached or increase in fitness of weight-sets in **X** is not significant. The complete flow of execution can be realized in Fig. 2.

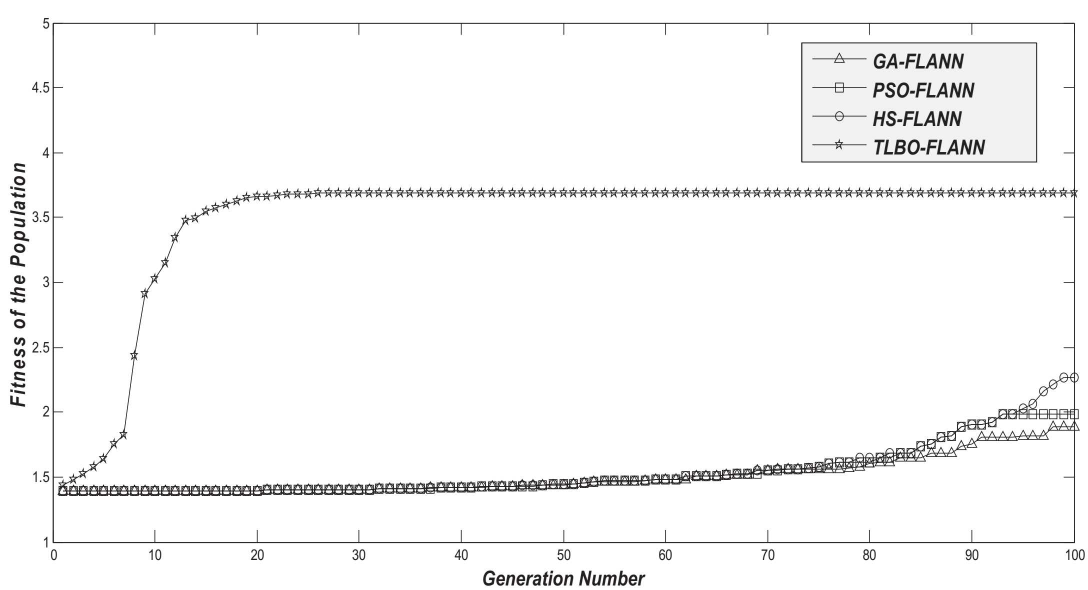

Figure 12 Observation on improvements in fitness of population in various iterations on Dermatology dataset.

Algorithm 1. TLBO based gradient descent learning FLANN (TLBO-FLANN)

1. Initialize n number of weight-sets (Population of students). X = $\{x_1, x_2, \ldots, x_n\}$, where each x_i is a potential weight-set of FLANN model which is initialized randomly as $x_i = \{w_{i,1}, w_{i,2}, \ldots, w_{i,m\times a\times(2\times k+1)}\}$.
2. Calculate mean of all the weight-sets (x_{mean}) in the population X.
3. Compute fitness of all the weight-sets in X by using Algorithm-2 (fitfromtrain procedure) and select weight-set with maximum fitness as best weight-set ($x_{teacher}$).
4. Generate next population X_{next} by using weight-sets in old population X, X_{mean}, $x_{teacher}$ and T_F:
   for i = 1:1: nos of weight-sets in the population X
      $T_F = round(1 + rand(1))$
      $x_i = x_i + rand(1)(x_{teacher} - T_F * X_{mean})$
      $X_{next}(i) = x_i$
   endfor
5. Update population of weight-sets X by comparing fitness of weight-sets in X and X_{next}:
   for i = 1:1: nos of weight-sets in the population X
      if (X(i) < X_{next}(i))
         X(i) = X_{next}(i)
      endif
   endfor
6. Randomly select two weight-sets from the population and improvise them:
   for k = 1:1: nos of weight-sets in the population X
      Select *i*th and *j*th weight-sets x_i and x_j randomly from population X.
      Compute fitness of x_i by using algorithm-2 as Fi = **fitfromtrain** ($\boldsymbol{\varphi}$, $\boldsymbol{x_i}$, **t**, $\boldsymbol{\mu}$).
      Compute fitness of x_j by using algorithm-2 as Fj = **fitfromtrain** ($\boldsymbol{\varphi}$, $\boldsymbol{x_j}$, **t**, $\boldsymbol{\mu}$).
      If (Fi < Fj)
      $X_{new}(i) = x_i + rand(1)(x_j - x_i)$
      Else
      $X_{new}(j) = x_j + rand(1)(x_i - x_j)$
      Ifend
   Endfor
7. Check for termination criteria:
   if (maximum nos of generation reached)
      then go to step-8.
   else go to step-2
   endIf
8. **Exit**

Algorithm 2. fitfromtrain procedure

Function F = **fitfromtrain** ($\boldsymbol{\varphi}$, $\boldsymbol{x_i}$, **t**, $\boldsymbol{\mu}$)
Let '$\boldsymbol{\varphi}$' be the functionally expanded input data and '$\boldsymbol{x_i}$' be the selected weight-set from the population. The vector S is computed as S = $\boldsymbol{\varphi}$. $\boldsymbol{x_i}$.
The output vector 'y' is computed by using tanh activation function as y = tanh(S).
The errors of the network 'e' is computed by using target vector 't' and output vector 'y' as e = t − y
Compute error term $\delta(k) = \left(\frac{1-y_k^2}{2}\right) \times e(k)$, *for k = 1,2…L*
where L is the number of patterns
If $\varphi = (\varphi_1, \varphi_2 \ldots \varphi_L)$, $e = (e_1, e_2 \ldots e_L)$ *and* $\delta = (\delta_1, \delta_2 \ldots \delta_L)$ *are vectors which represent set of functional expansion, set of errors and set of error terms respectively, then weight factor of w '*ΔW*' is computed as follows:* $\Delta W_q = \left(\frac{\sum_{i=1}^{L} 2\times\mu\times\varphi_i\times\delta_i}{L}\right)$
Compute root mean square error (RMSE) by using Eq. (3) from target vector 't' and output vector 'y'
Compute fitness 'F' of the network instance of FLANN model as F = 1/RMSE (Eq. (4)).
end

6. Experimental setup & data preparation

In this section, the environment for simulation, the data set used for training & testing phase and parameter settings for proposed methods during simulation are presented. All the hybrid models (GA-FLANN, PSO-FLANN, HS-FLANN and TLBO-FLANN) are implemented in MATLAB (Version 9.0) in a system with Window XP operating system. After obtaining the results of simulation, statistical analysis has been carried out using SPSS statistical tool (Version 16.0). The benchmark data sets (Table 3) used for training and testing phase for classification are originated from UCI machine learning repository (Bache et al., 2013) and processed by KEEL software (Alcalá-Fdez et al., 2011). The detailed descriptions about all these data sets can be obtained at 'http://archive.ics.uci.edu/ml/' and 'http://keel.es/'.

6.1. Cross validation

In this paper, all the data sets used for classification are prepared for cross validation by using 5-fold cross validation technique (Larson, 1931; Mosteller et al., 1968). During the preparation of data sets for the 5-fold cross validation, 5 pairs of data set samples are created and each pair contains data sets for the training and testing phase. For example (Table 4), the 'newthyroid-5-1tra.dat' and 'newthyroid-5-1tst.dat' data are a pair of data sets sample of New Thyroid data set which are used for the training and testing phase for a single run respectively. As 5-fold cross validation is employed, the New Thyroid data sets contain 5 such pairs of data set samples for training and testing the algorithms. The 5-fold cross validated data set for New Thyroid data set is presented in Table 4. All other data sets are prepared for 5-fold cross validation in the same way and collected from KEEL data set repository. The maximum fitness from their respective RMSEs on 5-fold cross validation data set during training and testing phase is listed in Table 5. The maximum fitness values obtained by various algorithms on training data ('newthyroid-5-1tra.dat', 'newthyroid-5-2tra.dat', 'newthyroid-5-3tra.dat', 'newthyroid-5-4tra.dat' and 'newthyroid-5-5tra.dat') for training phase and testing data ('newthyroid-5-1tst.dat', 'newthyroid-5-2tst.dat', 'newthyroid-5-3tst.dat', 'newthyroid-5-4tst.dat' and 'newthyroid-5-5tst.dat') for testing phase and are posted. This process is repeated for all other data sets in Table 3 and maximum fitness is posted in Table 5.

6.2. Parameter setting used for simulation

The FLANN parameters and TLBO parameters are set to suggested values from previous related works (Rao et al., 2011; Dehuri et al., 2012; Naik et al., 2014; Pao et al., 1989) on trial and error basis.

FLANN parameter:	TLBO parameter:
For FLANN, the learning parameter 'μ' is set to 0.13 in gradient descent learning by testing the models in the range 0–3. For functional expansion in FLANN, the value of n is set to 5, thereby each value in the input pattern is expanded to 11 number of functionally expanded input values (As in FLANN model, it is suggested (Pao et al., 1989; Pao and Takefuji, 1992) to generate 2n + 1 number of functionally expanded input values for a single value in the input pattern). The number of functionally input values increases hugely if a larger value of n is selected and the small value of n is unable to handle the non-linear nature of real world data sets. To make a balance between them, we have used the tested and suggested value of n (Dehuri et al., 2012; Naik et al., 2014) for functional expansion	For TLBO, we have set the common algorithmic parameters by testing the model by considering the suggested values (Population size = 40; Number of generations = 100; Stopping Criteria = Maximum number of generation) from previous related works (Rao et al., 2011)

7. Result obtained

Table 5 describes the comparison of maximum fitness obtained by all models and Table 6 represents average RMSEs obtained by four models: GA-FLANN, PSO-FLANN HS-FLANN and TLBO-FLANN on all 11 no of data sets. In this study, the performance of GA, PSO, HS and TLBO are examined in order to know the improvement of weight-sets in the population by these algorithms in various iterations. The changes in fitness of weight-sets in different iterations are observed in all the 11 number of data sets and the Figs. 3–13 demonstrates the improvements of fitness of weight-sets in the population.

8. Proof of statistical significance of the results

In this section, the statistical comparison of all the models over multiple data sets (Demsar, 2006) is presented to argue that the projected method is statistically better and significantly different from other alternative methods by using Friedman test (Friedman, 1937, 1940). List of results on which these tests have been carried out is presented in Table 7.

8.1. Friedman test

The Friedman test is a non-parametric statistical method which computes average ranks of algorithms by using Eq. (5) and compares them.

$$R_j = \frac{1}{N}\sum_i r_i^j \qquad (5)$$

In Eq. (5), r_i^j is the rank of the *j*th of k number of models on *i*th of N number of data sets. In Table 7, the ranks of each model on various data sets are shown in braces. Based on r_i^j, the average ranks of four models are found out from equation 5. The average ranks for all models: GA-FLANN,

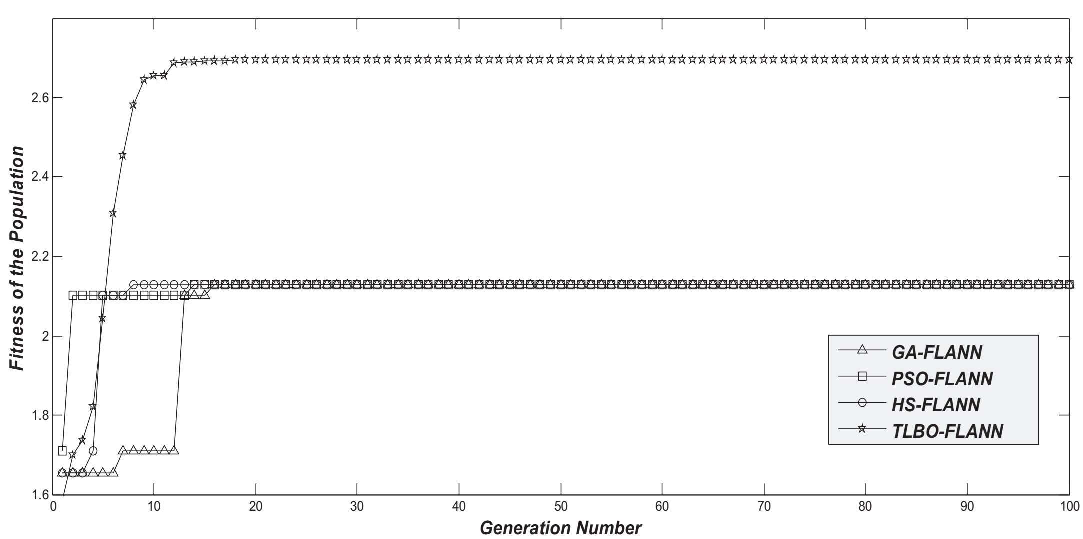

Figure 13 Observation on improvements in fitness of population in various iterations on Heart dataset.

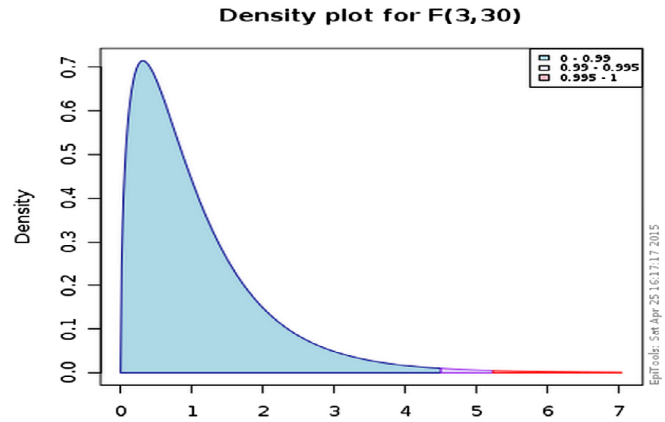

Figure 14 Density plot for 4.514 $F_F$statistic and (3, 30) degree of freedom by selecting α = 0.01.

Table 8 Result of Holm statistical test.

i	Learning models	z-values	p-values	$\frac{\alpha}{(k-i)}$
1	TLBO-FLANN : GA-FLANN	3.969	0.000036	0.0033
2	TLBO-FLANN : PSO-FLANN	3.141	0.000842	0.005
3	TLBO-FLANN : HS-FLANN	1.32	0.093418	0.01

PSO-FLANN HS-FLANN and TLBO-FLANN are found as $\{R_1 = \mathbf{3.455}, R_2 = 3, R_3 = 2, R_4 = \mathbf{1.272}\}$ respectively. The X_F^2 value is computed from average rank R_j of each model by using Eq. (6). In this study, we got the value of X_F^2 as 10.263. From the value of X_F^2, the Friedman statistics F_F is computed by Eq. (7) and found as 4.514. The Friedman statistics are distributed according to X_F^2 with (k − 1) degree of freedom under the null-hypothesis (H_0) and the critical value of the F-distribution can be obtained from F_F with (k − 1) and (k − 1) * (N − 1) degree of freedom. In our case, for the four number of classifiers and 11 number of data sets, the F_F = 4.514 with 4 − 1 = 3 and (4 − 1) * (11 − 1) = 30 degrees of freedom and the crucial value = 4.510 is obtained from suitably selecting α = 0.01. Density plot for degree of freedom (3, 30) is obtained and displayed in Fig. 14. The null-hypothesis is clearly rejected, as the critical value 4.510 is less than F_F statistic (4.514).

H_0: All the learning models have the same rank and differences between them are merely random, hence they are equivalent.

$$X_F^2 = \left(\frac{12N}{k(k+1)}\right)\left(\sum_j R_j^2 - \frac{k(k+1)^2}{4}\right) \tag{6}$$

Descriptives

Sample

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
GA-FLANN	110	2.4494	1.17060	.11161	2.2282	2.6706	1.55	5.97
PSO-FLANN	110	2.4689	1.16482	.11106	2.2488	2.6891	1.55	5.97
HS-FLANN	110	2.5278	1.16891	.11145	2.3069	2.7487	1.55	5.97
TLBO-FLANN	110	3.7371	2.45325	.23391	3.2735	4.2007	2.06	11.27
Total	440	2.7958	1.67557	.07988	2.6388	2.9528	1.55	11.27

ANOVA

Sample

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		130.310	3	43.437	17.182	.000
	Linear Term	Contrast	84.595	1	84.595	33.464	.000
		Deviation	45.714	2	22.857	9.042	.000
Within Groups			1102.195	436	2.528		
Total			1232.505	439			

Figure 15 Results of a one-way-ANOVA statistical test.

Multiple Comparisons

Dependent Variable:Sample

	(I) Algorithm	(J) Algorithm	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Tukey HSD	GA-FLANN	PSO-FLANN	-.01951	.21439	1.000	-.5724	.5334
		HS-FLANN	-.07840	.21439	.983	-.6313	.4745
		TLBO-FLANN	-1.28765*	.21439	.000	-1.8406	-.7348
	PSO-FLANN	GA-FLANN	.01951	.21439	1.000	-.5334	.5724
		HS-FLANN	-.05889	.21439	.993	-.6118	.4940
		TLBO-FLANN	-1.26814*	.21439	.000	-1.8210	-.7152
	HS-FLANN	GA-FLANN	.07840	.21439	.983	-.4745	.6313
		PSO-FLANN	.05889	.21439	.993	-.4940	.6118
		TLBO-FLANN	-1.20925*	.21439	.000	-1.7622	-.6563
	TLBO-FLANN	GA-FLANN	1.28765*	.21439	.000	.7348	1.8406
		PSO-FLANN	1.26814*	.21439	.000	.7152	1.8210
		HS-FLANN	1.20925*	.21439	.000	.6563	1.7622
Dunnett t (2-sided) [a]	GA-FLANN	TLBO-FLANN	-1.28765*	.21439	.000	-1.7930	-.7823
	PSO-FLANN	TLBO-FLANN	-1.26814*	.21439	.000	-1.7735	-.7628
	HS-FLANN	TLBO-FLANN	-1.20925*	.21439	.000	-1.7146	-.7039

*. The mean difference is significant at the 0.05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Figure 16 Results of Tukey and Dunnett statistical test.

Sample

	Algorithm	N	Subset for alpha = 0.05	
			1	2
Tukey HSD [a]	GA-FLANN	110	2.4494	
	PSO-FLANN	110	2.4689	
	HS-FLANN	110	2.5278	
	TLBO-FLANN	110		3.7371
	Sig.		.983	1.000
Duncan [a]	GA-FLANN	110	2.4494	
	PSO-FLANN	110	2.4689	
	HS-FLANN	110	2.5278	
	TLBO-FLANN	110		3.7371
	Sig.		.733	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 110.000.

Figure 17 Observations on models in their homogeneous group based on their level of significance.

Table 9 Various notations used in time complexity analysis on various models.

Symbol used	Meaning
T(n)	Time complexity of the models on input of n number of solution vector (weight-set)
O	Big-oh asymptotic notation
n	Number of weight-sets in the population.
nc	Number of attributes in weight-sets
c	Constant time
j, l	Number of functionally extended input patterns
k	Number of features in each functionally extended input pattern

$$\mathrm{F_F} = \frac{((N-1)X_F^2)}{(N(K-1)-X_F^2)} \tag{7}$$

After the rejection of null-hypothesis from Friedman test, the post hoc test has been carried out by using the Holm procedure (Demsar, 2006; Luengo et al., 2009; Garcia et al., 2010) in order to compare the performance of the proposed model with other models based on z-score value and p-value.

8.2. Holm procedure

In this section, the Holm procedure (Holm, 1979) is used for pair wise comparison based on their z-score value and p-value. The Eq. (8) is used to obtain z-value and p-value is computed from the z-value and table of the normal distribution.

$$z = \frac{(R_i - R_j)}{\sqrt{\frac{k(k+1)}{6N}}} \tag{8}$$

Here, z is the z-score value, k is the number of models and N is the number of data sets. The R_i and R_j are average ranks of ith and jth model respectively. All four models are compared based on z-value, p-value and $\frac{a}{(k-i)}$ (Table 8), where 'i' is the model's number. By using the Holm test, while comparing p_i values with corresponding $\frac{a}{(k-i)}$ values, we observed that, in almost all cases p_i is less than $\frac{a}{(k-i)}$ (except comparison between TLBO-FLANN with HS-FLANN). Hence, on this basis the null-hypothesis is rejected. Thus, the proposed classification model TLBO-FLANN is statistically better and significantly different from other models (except HS-FLANN). While comparing with HS-FLANN, the TLBO-FLANN is found to be better but not significantly better than HS-FLANN because, p-value is not less than respective $\frac{a}{(k-i)}$ *value*.

8.3. Post-hoc ANOVA statistical analysis (Tukey test & Dunnett test)

After the rejection of the null-hypothesis from the Friedman test in the Section 8.1 and analysis under the Holm procedure in Section 8.2, in this section, the ANOVA statistical analysis (Fisher, 1959) has been carried out by using Tukey Test & Dunnett test to get generalized statistics on the performance of all models.

In this paper, the statistics on all model's performance is computed under ANOVA test by using SPSS (Version: 16.0) statistical tool. All the methods are executed for ten no.s of runs on each data set. The test has been carried out with 95% confidence interval, 0.05 significant level and linear polynomial contrast (Fig. 15). To get the differences between the performances of classifiers, the Post-hoc test is conducted by using Tukey test (Tukey, 1949) and Dunnett test (Dunnett, 1980). The Tukey test is carried out for comparisons of performance of all models with each other and the Dunnett test for comparison of all models with base classification model (proposed model). The results from Tukey test and Dunnett test are presented in Fig. 16. As a conclusion of these tests, we noticed that, the mean differences (between-classification models variability) among models are larger than the standard errors (between-error variability) (Fig. 16). Also in Dunnett test (Fig. 16), we observed the same as that of Tukey test. Hence the null-hypothesis can be clearly rejected. Furthermore, all the models in their homogeneous group based on their level of significance are presented in Fig. 17.

9. Time complexity analysis

In this section, all the models that we have considered are analyzed by comparing their time complexities. The notations used for time complexity analysis are listed in Table 9. Tables 10–13 represent the number of program steps taken by the models: GA-FLANN, PSO-FLANN, HS-FLANN and TLBO-FLANN respectively. In this analysis, we noted that time complexity T(n) of GA-FLANN, PSO-FLANN, HS-FLANN and TLBO-FLANN are dominated by the factor $(n*(c*((j*k*c)+(l*c))))$, $(c*(n*(c*((j*k*c)+(l*c)))))$, $(c*(n*(c*((j*k*c)+(l*c)))))$ and $(c*(n*(c*((j*k*c)+(l*c)))))$ respectively. The time complexities

Table 10 Time complexity analysis on GA based FLANN (*GA-FLANN*) model.

Algorithm step	Time complexity
Fitness evaluation in FLANN model	$(n*(c*((j*k*c)+(l*c))))$
Mating pool construction	$(n+(n+c))$
Crossover	$((n/2)*c)$
Selection	$(n*c)$
Stopping criteria check	$(n*c)$
Overall time complexity	$T(n)=(n*(c*((j*k*c)+(l*c))))+(n+(n+c))+((n/2)*c)+(n*c)+(n*c)$ $T(n)=\mathrm{O}((n*(c*((j*k*c)+(l*c)))))$

Table 11 Time complexity analysis on PSO based FLANN *(PSO-FLANN)* model.

Algorithm step	Time complexity
Fitness evaluation in FLANN model	$(n*(c*((j*k*c)+(l*c))))$
Local best selection	$(2*(n*(c*((j*k*c)+(l*c))))+(n*c))$
Global best selection	$(n*c)$
Velocity calculation	$(n*nc*c)$
Particle position updation	$(n*c)$
Stopping criteria check	$(n*c)$
Overall time complexity	$T(n)=(n*(c*((j*k*c)+(l*c))))+(2*(n*(c*((j*k*c)+(l*c))))$ $+(n*c))+(n*c)+(n*nc*c)+(n*c)+(n*c)$ $T(n)=\mathrm{O}(c*(n*(c*((j*k*c)+(l*c)))))$

Table 12 Time complexity analysis on HS based FLANN *(HS-FLANN)* model.

Algorithm step	Time complexity
Initialization of harmony memory (HM), harmony memory size, harmony memory consideration rate, pitch adjustment rate and, bandwidth	c
Computes MCP (memory consideration probability), PAP (pitch adjustment probability) and RP (randomization probability).	c
Improvise HM to get new harmony memory (NHM)	$(HMS*c)$
Update the HM based on comparison between solution vectors of HM and NHM	$(HMS*c*(n*(c*((j*k*c)+(l*c)))))$
Fitness evaluation in FLANN model	$(n*(c*((j*k*c)+(l*c))))$
Stopping criteria	1
Overall time complexity	$T(n)=c+c+(HMS*c)+(HMS*c*(n*(c*((j*k*c)$ $+(l*c)))))+(n*(c*((j*k*c)+(l*c))))+1$ $T(n)=\mathrm{O}(c*(n*(c*((j*k*c)+(l*c)))))$

Table 13 Time complexity analysis on TLBO based FLANN *(TLBO-FLANN)* model.

Algorithm step	Time complexity
Mean calculation	$(n*c)$
Fitness evaluation in FLANN model	$(n*(c*((j*k*c)+(l*c))))$
Teacher selection	$(n*c)$
Next population generation	$(n*c)$
Population updation	$(n*c)$
Improvisation of population	$((n*2*(c*((j*k*c)+(l*c))))+(n*c))$
Stopping criteria	1
Overall time complexity	$T(n)=(n*c)+(n*(c*((j*k*c)+(l*c))))+(n*c)+(n*c)+(n*c)$ $+((n*2*(c*((j*k*c)+(l*c))))+(n*c))+1$ $T(n)=\mathrm{O}(c*(n*(c*((j*k*c)+(l*c)))))$

Table 14 Comparison of various models based on time complexity.

Implemented algorithms	Time complexity
GA-FLANN	$\mathrm{O}((n*(c*((j*k*c)+(l*c)))))$
PSO-FLANN	$\mathrm{O}(c*(n*(c*((j*k*c)+(l*c)))))$
HS-FLANN	$\mathrm{O}(c*(n*(c*((j*k*c)+(l*c)))))$
TLBO-FLANN	$\mathrm{O}(c*(n*(c*((j*k*c)+(l*c)))))$

of all the algorithms are listed in Table 14. Here, we conclude that the time complexity of all the models is dominated by the time complexity of the basic step '*Fitness Evaluation in FLANN model*' (Algorithm-2) which is $(n*(c*((j*k*c)+(l*c))))$. The proposed method TLBO-FLANN is equivalent with other models (GA-FLANN, PSO-FLANN and HS-FLANN) in terms of time complexity but it is better in learning ability as compared to other alternatives GA-FLANN, PSO-FLANN and HS-FLANN.

10. Discussion & conclusion

The proposed TLBO-FLANN model can be computed at a low cost due to that less complex architecture of FLANN and TLBO requires few common algorithmic specific parameters and mathematical computations. Also TLBO is free from complicated operators (like crossover in GA) and parameters (like c1, c2 in PSO and HMCR, PAR & bw in HS). The performance of TLBO-FLANN can be improved by adopting other improved versions of TLBO.

In this study, we analyzed the proposed TLBO-FLANN by using various statistical methods like Friedman test, Holm procedure, Tukey & Dunnett Test post hoc test and one-way-ANOVA. As a result, a clear rejection of null-hypothesis is noticed in Friedman Test and Tukey & Dunnett test post hoc test, but in the Holm procedure, the null-hypothesis is rejected except between TLBO-FLANN and HS-FLANN. This is because; the proposed TLBO-FLANN model is statistically better and significantly different from GA-FLANN and PSO-FLANN models. On the other hand, the TLBO-FLANN outperforms HS-FLANN in terms of accuracies but not significantly different and higher than HS-FLANN. From rigorous tests under well known statistical methods (Friedman test, Post-hoc test by Holm procedure, Tukey test, Dunnett test and one-way-ANOVA), we claim the proposed TLBO-FLANN model is better and outperforms other alternative hybrid models (GA-FLANN, PSO-FLANN and HS-FLANN).

In time complexity analysis, it is observed that, the proposed method TLBO-FLANN is similar to other alternatives (GA-FLANN, PSO-FLANN and HS-FLANN), but the performance of TLBO-FLANN is better than GA-FLANN, PSO-FLANN and HS-FLANN in term of learning capability. The future work is comprised of integration of other new and improved variants of TLBO with FLANN in diverse applications of data mining.

Acknowledgement

We would like to thank anonymous reviewers for their valuable comments and suggestions which have certainly improved the quality of this paper.

This work is supported by Department of Science & Technology (DST), Ministry of Science & Technology, New Delhi, Govt. of India, under Grants No. DST/INSPIRE Fellowship/2013/585.

References

Abbass, H.A., 2001. A monogenous MBO approach to satisfiability. In: International Conference on Computational Intelligence for Modelling Control and Automation, CIMCA 2001, Las Vegas, NV, USA.

Alcalá-Fdez, J., Fernandez, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F., 2011. KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. J. Multiple-Valued Logic Soft Comput. 17 (2–3), 255–287.

Arya, L.D., Koshti, Atul., 2014. Anticipatory load shedding for line overload alleviation using teaching learning based optimization (TLBO). Electr. Power Energy Syst. 63, 862–877.

Azizipanah-Abarghooee, Rasoul, Niknam, Taher, Bavafa, Farhad, Zare, Mohsen, 2014. Short-term scheduling of thermal power systems using hybrid gradient based modified teaching–learning optimizer with black hole algorithm. Electr. Power Syst. Res. 108, 16–34.

Bache, K., Lichman, M., 2013. UCI Machine Learning Repository <http://archive.ics.uci.edu/ml]>, Irvine, CA: University of California, School of Information and Computer Science.

Barisal, A.K., 2015. Comparative performance analysis of teaching learning based optimization for automatic load frequency control of multi-source power systems. Electr. Power Energy Syst. 66, 67–77.

Chakravarty, S., Dash, P.K., 2009. Forecasting stock market indices using hybrid network. In: IEEE World Congress on Nature & Biologically Inspired Computing. pp. 1225–1230.

Chen, Debao, Zou, Feng, Li, Zheng, Wang, Jiangtao, Li, Suwen, 2015. An improved teaching–learning-based optimization algorithm for solving global optimization problem. Inf. Sci. 297, 171–190.

Cheng, Yu-Huei, 2014. Estimation of teaching–learning-based optimization primer design using regression analysis for different melting temperature calculations. IEEE Trans. Nano Biosci. (99) http://dx.doi.org/10.1109/TNB.2014.2352351.

Črepinšek, Matej, Liu, Shih-Hsi, Mernik, Luka, 2012. A note on teaching–learning-based optimization algorithm. Inf. Sci. 212, 79–93.

Dehuri, S., Cho, S., 2009. A comprehensive survey on functional link neural networks and an adaptive PSO–BP learning for CFLNN. Neural Comput. Appl. 19 (2), 187–205.

Dehuri, S., Cho, S.-B., 2010. Evolutionarily optimized features in functional link neural network for classification. Expert Syst. Appl. 37, 4379–4391.

Dehuri, S., Roy, R., Cho, S., Ghosh, A., 2012. An improved swarm optimized functional link artificial neural network (ISO-FLANN) for classification. J. Syst. Softw., 1333–1345

Demsar, J., 2006. Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7, 1–30.

Dunnett, C.W., 1980. A multiple comparison procedure for comparing several treatments with a control. J. Am. Stat. Assoc. 50, 1096–1121.

Fisher, R.A., 1959. Statistical Methods and Scientific Inference, second ed. Hafner Publishing Co., University of Michigan, New York.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Am. Stat. Assoc. 32, 675–701.

Friedman, M.A., 1940. Comparison of alternative tests of significance for the problem of m rankings. Ann. Math. Stat. 11, 86–92.

García, Juan Andrés Martín, Mena, Antonio José Gil, 2013. Optimal distributed generation location and size using a modified teaching–learning based optimization algorithm. Electr. Power Energy Syst. 50, 65–75.

Garcia, S., Fernandez, A., Luengo, J., Herrera, F., 2010. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power. Inf. Sci. 180, 2044–2064.

Geem, Z.W., Kim, J.H., Loganathan, G.V., 2001. A new heuristic optimization algorithm: harmony search. Simulation 76, 60–70.

Ghasemi, Mojtaba, Taghizadeh, Mahdi, Ghavidel, Sahand, Aghaei, Jamshid, Abbasian, Abbas, 2015a. Solving optimal reactive power dispatch problem using a novel teaching–learning-based optimization algorithm. Eng. Appl. Artif. Intell. 39, 100–108.

Ghasemi, Mojtaba, Ghavidel, Sahand, Gitizadeh, Mohsen, Akbari, Ebrahim, 2015b. An improved teaching–learning-based optimization algorithm using Lévy mutation strategy for non-smooth optimal power flow. Electr. Power Energy Syst. 65, 375–384.

Gillies, C.E., Siadat, M., Patel, N.V., Wilson, G.D., 2013. A simulation to analyze feature selection methods utilizing gene ontology for gene expression classification. J. Biomed. Inform. 46 (6), 1044–1059.

Goel, A., Saxena, S., Bhanot, S., 2006. Modified functional link artificial neural network. Int. J. Electr. Comput. Eng. 1 (1), 22–30.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley, Reading, MA, ISBN 0201157675.

Hajmohammadi, M.S., Ibrahim, R., Selamat, A., 2014. Bi-view semi-supervised active learning for cross-lingual sentiment classification. Inf. Process. Manage. 50 (5), 718–732.

Hamamoto, Y., Uchimura, S., Tomita, S., 1997. A bootstrap technique for nearest neighbour classifier design. IEEE Trans. Pattern Anal. Mach. Intell. 19 (1), 73–79.

He, Z., Lin, S., Deng, Y., Li, X., Qian, Q., 2014. A rough membership neural network approach for fault classification in transmission lines. Int. J. Electr. Power Energy Syst. 61, 429–439.

Holland, J.H., 1992. Genetic algorithms. Sci. Am., 66–72

Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proceedings of the 1995 IEEE International Conference on Neural Networks, 4, 1942–1948.

Khalghani, Mohammad Reza, Khoob, Mohammad Hassan, 2014. A novel self-tuning control method based on regulated bi-objective emotional learning controller's structure with TLBO algorithm to control DVR compensator. Appl. Soft Comput. 24, 912–922.

Kianmehr, K., Alshalalfa, M., Alhajj, R., 2010. Fuzzy clustering-based discretization for gene expression classification. Knowl. Inf. Syst. 24 (3), 441–465.

Lai, D.T.C., Garibaldi, J.M., 2013. A preliminary study on automatic breast cancer data classification using semi-supervised fuzzy c-means. Int. J. Biomed. Eng. Technol. 13 (4), 303–322.

Larson, S., 1931. The shrinkage of the coefficient of multiple correlation. J. Educat. Psychol. 22, 45–55.

Luengo, J., Garcia, S., Herrera, F., 2009. A study on the use of statistical tests for experimentation with neural networks: analysis of parametric test conditions and non-parametric tests. Expert Syst. Appl. 36, 7798–7808.

Majhi, R., Panda, G., Sahoo, G., 2009. Development and performance evaluation of FLANN based model for forecasting of stock markets. Expert Syst. Appl. 36, 6800–6808.

Mandal, Barun, Roy, Provas Kumar, 2013. Optimal reactive power dispatch using quasi-oppositional teaching learning based optimization. Electr. Power Energy Syst. 53, 123–134.

Mei, K., Dong, P., Lei, H., Fan, J., 2014. A distributed approach for large-scale classifier training and image classification. Neurocomputing 144, 304–317.

Mishra, B.B., Dehuri, S., 2007. Functional link artificial neural network for classification task in data mining. Journal of Computer Science 3 (12), 948–955, Science Publications.

Mishra, S., Shaw, K., Mishra, D., Patnaik, S., 2012. An enhanced classifier fusion model for classifying biomedical data. Int. J. Comput. Vis. Rob. 3 (1/2), 129–137.

Moghadam, Ahmad, Seifi, Ali Reza, 2014. Fuzzy-TLBO optimal reactive power control variables planning for energy loss minimization. Energy Convers. Manage. 77, 208–215.

Mosteller, F., Turkey, J.W., 1968. Data analysis, including statistics. In: Handbook of Social Psychology. Addison-Wesley, Reading, MA.

Naik, B., Nayak, J., Behera, H.S., Abraham, A., 2014. A Harmony Search Based Gradient Descent Learning-FLANN (HS-GDL-FLANN) for Classification, Computational Intelligence in Data Mining, vol. 2, Smart Innovation, Systems and Technologies, vol. 32, 2015, pp. 525–539.

Naik, B., Nayak, J., Behera, H.S., 2015a. A novel FLANN with a hybrid PSO and GA based gradient descent learning for classification. In: Proc. of the 3rd Int. Conf. on Front. of Intell. Comput. (FICTA). Advances in Intelligent Systems and Computing, vol. 327. pp. 745–754. doi: 10.1007/978-3-319-11933-5_84.

Naik, B., Nayak, J., Behera, H.S., 2015b. A honey bee mating optimization based gradient descent learning – FLANN (HBMO-GDL-FLANN) for classification. In: Proceedings of the 49th Annual Convention of the Computer Society of India CSI – Emerging ICT for Bridging the Future. Advances in Intelligent Systems and Computing, vol. 338. pp. 211–220. doi: 10.1007/978-3-319-13731-5_24.

Niknam, T., Golestaneh, F., Sadeghi, M.S., 2012. θ-multiobjective teaching-learning-based optimization for dynamic economic emission dispatch. IEEE Syst. J. 6 (2), 341–352. http://dx.doi.org/10.1109/JSYST.2012.2183276.

Niu, Qun, Zhang, Hongyun, Li, Kang, 2014. An improved TLBO with elite strategy for parameters identification of PEM fuel cell and solar cell models. Int. J. Hydrogen Energy 39 (2014), 3837–3854.

Pao, Y.H., 1989. Adaptive Pattern Recognition and Neural Networks. Addison-Wesley Pub.

Pao, Y.H., Takefuji, Y., 1992. Functional-link net computing: theory, system architecture, and functionalities. Computer 25, 76–79.

J.C. Patra, W. Lim, N. Thanh, P. Meher, 2009. Computationally efficient FLANN-based intelligent stock price prediction system. In: IEEE Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14–19, pp. 2431–2438.

Roy, Provas Kumar, 2013. Teaching learning based optimization for short-term hydrothermal scheduling problem considering valve point effect and prohibited discharge constraint. Electr. Power Energy Syst. 53, 10–19.

Quinlan, J.R., 1993. In: C4. 5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.

Rao, R.V., Savsani, V.J., Vakharia, D.P., 2011. Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems. Comput.-Aided Des. 43, 303–315.

Redding, N., Kowalczyk, A., Downs, T., 1993. Constructive high-order network algorithm that is polynomial time. Neural Netw. 6, 997–1010.

Roy, Provas Kumar, Bhui, Sudipta, 2013. Multi-objective quasi-oppositional teaching learning based optimization for economic emission load dispatch problem. Electr. Power Energy Syst. 53 (2013), 937–948.

Roy, Provas Kumar, Sur, Aditi, Pradhan, Dinesh Kumar, 2013. Optimal short-term hydro-thermal scheduling using quasi-oppositional teaching learning based optimization. Eng. Appl. Artif. Intell. 26, 2516–2524.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (9), 533–536.

Sahu, Rabindra Kumar, Gorripotu, Tulasichandra Sekhar, Panda, Sidhartha, 2016. Automatic generation control of multi-area power systems with diverse energy sources using teaching learning based optimization algorithm. Engineering Science and Technology, An International Journal 19 (1), 113–134.

Satapathy, Suresh Chandra, Naik, Anima, Parvathi, K., 2013a. Weighted teaching–learning-based optimization for global function optimization. Appl. Math. 4, 429–439.

Satapathy, Suresh Chandra, Naik, Anima, Parvathi, K., 2013b. A teaching learning based optimization based on orthogonal design for solving global optimization problems. SpringerPlus 2, 130.

Singh, Manohar, Panigrahi, B.K., Abhyankar, A.R., 2013. Optimal coordination of directional over-current relays using teaching learning-based optimization (TLBO) algorithm. Electr. Power Energy Syst. 50, 33–41.

Sultana, Sneha, Roy, Provas Kumar, 2014. Optimal capacitor placement in radial distribution systems using teaching learning based optimization. Electr. Power Energy Syst. 54, 387–398.

Sun, J., Patra, J., Lim, W., Li, Y., 2009. Functional link artificial neural network-based disease gene prediction. In: IEEE Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14–19, pp. 3003–3010.

Theja, B.S., Rajasekhar, A., Abraham, A., 2013. An optimal design of coordinated PI based PSS with TCSC controller using modified teaching learning based optimization, World Congress on Nature and Biologically Inspired Computing (NaBIC). 99–106. doi: 10.1109/NaBIC.2013.6617845.

Tolambiya, A., Venkataraman, S., Kalra, P.K., 2010. Content-based image classification with wavelet relevance vector machines. Soft. Comput. 14 (2), 137.

Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5, 99–114.

Tuo, Shouheng, Yong, Longquan, Zhou, Tao, 2013. An Improved Harmony Search Based on Teaching-Learning Strategy for Unconstrained Optimization Problems, Mathematical Problems in Engineering. Hindawi Publishing Corporation. http://dx.doi.org/10.1155/2013/413565.

Uysal, A.K., Gunal, S., 2014. Text classification using genetic algorithm oriented latent semantic features. Expert Syst. Appl. 41 (13), 5938–5947.

Vedat Toğan, 2012. Design of planar steel frames using teaching–learning based optimization. Eng. Struct. 34, 225–232.

Venkata Rao, R., Kalyankar, V.D., 2013. Parameter optimization of modern machining processes using teaching–learning-based optimization algorithm. Eng. Appl. Artif. Intell. 26, 524–531.

Waghmare, Gajanan, 2013. Comments on 'A note on teaching–learning-based optimization algorithm'. Inf. Sci. 229, 159–169.

Yager, R.R., 2006. An extension of the naive Bayesian classifier. Inf. Sci. 176 (5), 577–588.

Yang, P., Gao, W., Tan, Q., Wong, K., 2013. A link-bridged topic model for cross-domain document classification. Inf. Process. Manage. 49 (6), 1181–1193.

Yung, Y., Shaw, M.J., 1995. Introduction to fuzzy decision tree. Fuzzy Net Syst. 69 (1), 125–139.

Zhang, G.P., 2000. Neural networks for classification: a survey. IEEE Trans. Syst. Man Cybern. Part C Appl. Rev. 30 (4), 451–462.

Zou, Feng, Wang, Lei, Hei, Xinhong, Chen, Debao, Wang, Bin, 2013. Multi-objective optimization using teaching–learning-based optimization algorithm. Eng. Appl. Artif. Intell. 26, 1291–1300.