Heart Diseases Prediction for Optimization based Feature Selection and Classification using Machine Learning Methods

N. Rajinikanth[1], Dr. L. Pavithra[2]

Associate Professor, CMS College of Science and Commerce
Associate Professor, Department of Computer Science, Dr. N. G. P. Arts and Science College, Coimbatore, india

Abstract—**Globally, heart disease is considered to be the major cause of death. As per statistics, 17.9 million people are losing their lives every year worldwide. Chronic Kidney Disease (CKD) and Breast Cancer takes the next positions in the list. Disease classification is an important issue that needs more attention now. Making use of an optimized technique for such classification would be a better option. In this heart disease classification, initially, feature selection was done using Teaching learning based Optimization based (TLO) and Kernel Density. TLO is based on the process of classroom teaching, which involves too much iteration that leads to time complexity. Similarly, a certain level of misclassifications has been observed by using Kernel Density (KD). In the proposed method, K-Nearest Neighbour (KNN) is used to address the issue of NaN values and Density based Modified Teaching Learning based Optimization (DMTLO) is used for feature selection. Finally the classification process is done by considering Support Vector Machine (SVM) and Ensemble (Adaboosting method). SVM categorizes data bydissimilar class names by defining a group of support vectors that are part of the group of training inputs that plan a hyper plane in the attribute space. Ensemble method is used to solve statistical, computational and representational problems. Experimental outcomes have proved that the projected DMTLOovertakes the existing methodologies with required quantity of attributes.**

Keywords—Teaching learning based optimization; kernel density; support vector machine; k-nearest neighbour; ensemble learning

I. INTRODUCTION

Nowadays, datasets are tremendously accumulated with enormous quantity of data sources. Such high dimensional data rises the calculationrate and diminishes the results of a ML model if the dataset has inappropriate, duplicate and unwanted attributeswhich is notfavourable to the improvement of an analytical model. The issue of over fitting with vast number of features could be addressed by using Learning models. Choosing a relevant and suitable set of features could be a better way to solve this problem. Several feature selection algorithms are available in this regard. These algorithms are capable of minimizing the quantity of features in order to develop an AI model by authenticating different arrangements of features in an input dataset.

In general, wrapper based attribute selection strategies are projected to improve the competencies of classification methods. Finding a worthy arrangement of attributes is really a challenging task. Various optimization techniques are utilized for choosing proper features such as Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) by numerousscientists to advance the outcomes of the classifiers.

Parham et al., (2016) [9] established an attribute choosing strategy which is a hybridization of PSO and local search strategy. Its results were evaluated with various screen and wrapper-based strategies. It has attained notable precision results.

Hafez et al. (2015) [5] proposed an attribute choosing procedure that is dependent on Chicken swarm optimization. It replicated the performance of chicken swarms and attainedgood resultsthroughtypical datasets relatedtowards GA and PSO optimization algorithms. A methodology proposed by Panda (2017) [12] relies on elephant search optimization in aalliance with deep NN for inspecting microarray data. Venkata Rao (2016) [14], Rao (2016) [21] proposed extensive presentations of TLBO in many real time problems. The strategy of TLBO is proposed to decrease load of fixing the parameter standards during attribute choosing process.

II. RELATED WORK

Attribute selection is highly needed in various areas like categorization of emails, disease analysis, forged claims and also in the areas of credit/debit risks. In the process of developing a well-organized decision-making method, the significant step is to organize the better features which are more suitable to attain better precision results.Various scientists have made use of filter and wrapper choosing strategies Wah et al., (2018) [22] to increase the correctness of forecaststrategies. Several prevailing attribute choosing strategies have been observed to comprehend its pros and cons. Bahassine et al. (2018) [3] have projected a novel attribute choosing method for categorization of Arabic text by means of an better Chi-square technique to improve the classification outcomes. Better results have been attained by incorporating SVM classifier.

Mazini et al. (2018) [11] established a new method intended for abnormality network-based intrusion discovery model. This helps to attain a maximum detection rate with a minimum false positive rate. This model is a hybridization of both artificial bee colony and AdaBoost algorithm. The former is utilized for selecting efficient attribute whereas the latter is for classification.

Thawkar et al., (2018) [18] projected an attribute choosing method. This method was developed using Biogeography-based optimization procedureaimed atcategorization of numeral mammograms with ANN.

Wen et al. (2016) [23] developed a novel unsupervised attribute choosing techniquethat is related on L2;1-norm regularization on behalf ofidentifying certain human movements. The above said procedure achieves both attribute mining and selection instantaneously which produces ideal attributes.

Xu et al. (2017) [24] projected an innovative discriminative L2 regularization-based sparse demonstration. This procedure is exclusively for classifying input images and accomplishednotableprecisionthroughvarious inputs.

Absolute dimensionality reducing method is proposed by Lai et al. (2017) [7] that can be termed asRobust Discriminant Regression (RDR) by means of L2,1-norm as the elementarystandard in the evaluation function for attribute extraction. RDR doesn't get proper predictions for attribute selection and that is considered to be its main disadvantage.

Mafarja et al. (2017) [8] utilized the Dual Dragonfly Procedure. This is in the direction ofpicking a subdivision of attributes taken fromUCI repository and attainedimprovedout comesequated with GA and PSO algorithms.

Sayed et al. (2017) [15] recommended a fresh meta-heuristic techniquewhich is similar to crow search procedure for picking proper attributes and appealedhealthieroutcomes throughstandard datasets.

Sayed et al. (2018) [16] established a hybridized technique which is a combination of swarm algorithm for attribute selection and with chaos theory. This addresses the issues of confined optima and little convergence problems.

Agrawal et al., (2015) [1] projected a novel attribute selection strategy that is dependent on Artificial Bee Colony and K-NN algorithms. This is used for categorizing theCT images of cervical cancer.

Marie-Sainte et al., (2018) [10] recommended an innovative attribute choosing method for categorizing Arabic text with the help of firefly algorithm. This obviously improves classification performance. The researchershave made trials on OSAC dataset and accomplished 0.994 accuracy rate.

Shahbeig et al. (2016) [17] designated a subcategory of interrelatedDNAscollected from the input of breast cancer microarraythrough the support of transformed fuzzy adaptive PSO incorporated with TLBO procedure and confirmed the correctnessby SVM classifier.

Tuo et al. (2017) [19] established an original hybrid HSTLBO technique that stabilizes the convergence difficulty of distinct TLBO and Harmony Search procedures.

III. EXISTING SYSTEM

Feature selection can be done in dualways; Teaching Learning based Optimization (TLO) and Kernel Density (KD)

A. *Feature Selection using Teaching Learning based Optimization (TLO)*

TLO is familiar technique towards choosing the ideal sub division of features. This has binary segments. First segment covers an optimization Technique, which can be utilized to choose ideal set of attributes. Various classification models are covered in the latter phase. These segments arerecurrenttill anendingcondition has seen. Stopping criteria can be taken as astaticamount of iterations. Improved precision with various classification models cannot be adopted in Teaching Learning based Optimization (TLO) and also this TLO cannot be hybridized with any other feature selection strategies.

B. *Feature Selection using Kernel Density (KD)*

Kernel Density (KD) isa non-parametric and it doesn't make any conventions with respect to data distribution. It always chooses attributes that capture the performance of usual data by separating the outliers. A forward search strategy is used for estimating standards. This is highly capable of discovering outliers when compared to other familiar strategies. Incorporating other search techniques would be a more challenging factor in terms of attribute selection since it exploits the parallelism. Also, no proper studies have been done so far to ensure the value of the features.

IV. PROPOSED SYSTEM

In the proposed system pre-processing to remove the Nan is done using KNN method, feature selection using Density based Modified Teaching Learning based Optimization (DMTLO) and Kernel Density (KD) based method. Classification is done using Classification using SVM and Ensemble (Adaboosting method).

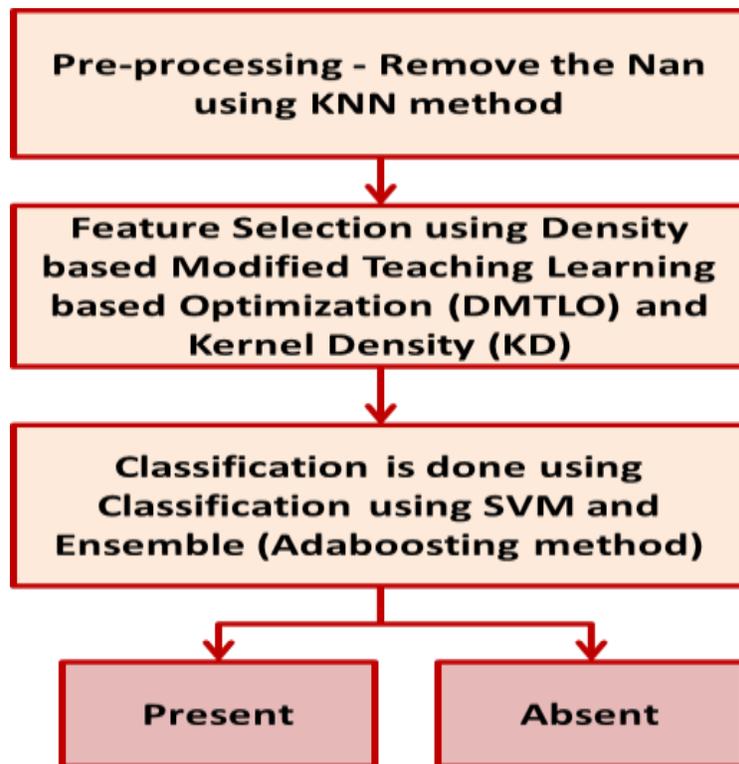Fig. 1 represents the architecture of the proposed system.

Fig. 1. Architecture of the Proposed System

A. Pre-Processing to Remove the Nan using KNN Method

In familiar data mining tasks like, classification and regression Altman (1992) [2], K-Nearest Neighbour (K-NN) is considered to be a constraintfree approach. It is a method of instance-based learning and it is likewise termed as lazy learning. Local approximations are done on the functions and the calculations are suspended until classification. It is considered to be the basic way of all AI techniques.

Its outcomes determine the classification or regression. The characteristicscomprise ease to take outcomes, calculation time and analyticalcompetence.If K-NN is utilized for classification, the results give the class membership.

Objects are categorizedby means ofconsidering the vote attained from neighbours. All those objects are allocated to a class which is more obvious in KNN.In the phase of regression, the outcome provides the stuff of object which is the average of the values of KNNs.

B. Feature Selection using Density based Modified Teaching Learning based Optimization (DMTLO)

Density based Modified Teaching Learning based Optimization (DMTLO) is adopted inorder to streamline the conventional TLBO in the calculation of evaluation function. The size of input and design variables is considered to be theinput parameters to discover the biasedgroup of attributes.

DMTLO starts by fixing the populacescope, t i.e., the quantity of learners (ps= 1, 2, 3 ….. n) and the design variable, s i.e., the quantity of subjects (su=1, 2, 3…..m) which aretrained.

The representation of objective function is given below.

Minimum $f(y) = \sum nr=1 \ [y2 \ r - 10 \cos (2\pi yr) + 10]$ (1)

1) Teacher phase: The best learner would be chosen in this phase. Teacher tries to take an attempt in order to enrich the understanding of rest of the learners by maximizing their average mean. Throughout this phase, final iteration can be represented as.

y^{th}Iteration for (y=1, 2, 3….m)

Subject x(x=1, 2, 3…..n)

Mean value for individual subject is considered and it could be demonstrated as ms(x,y)

In this phase, variancesare taken to modernize the standards in the resolutionpool by totalling the value of differences to the present solution and the algorithm continues to the learner phase.

Chebyshev distance metric is taken to modernize the values in output space. Differences are denoted as Ds, Dchebyshev distance as Dc.

$Ds = v \ (O\_new,s \ -TFOs)$ (2)

$Dc \ (yi, yj) = \max \ (|yi-yj|)$ (3)

$X'new= f(y) + Dc(yi-yj)$ (4)

2) Learner phase: By making interaction with the peers, the understandability of individual learners can be improved.

For y=1:tr

Choose additional learner arbitrarily Xx, such that y≠x

If f(Xx) < f(Xx)

X"new,y=X'new,y + ry(Xy– Xx)

Else

X"new,y=X'new,y+ ry(Xx – Xy)

End If

End For

Admit'X"new', when a function value is superior to its earlier value. The attributes that shows enhanced outcomes based on the latest evaluation function through the every cycle is accumulated in attribute subset. This algorithm finishes when each and every attributes are taken for evaluation.

C. Classification using SVM and Ensemble (Adaboostingmethod)

Classification is done using Classification using SVM and Ensemble (Adaboosting method).

1) Support Vector Machine (SVM): One of the newest procedures aimed at pattern classification is SVM. It is extensively used in various fields. It is a supervised learning technique connected with learning procedures to examine data and to distinguish patterns. Fixing up the kernel factor for SVM in training phase will definitely influence the correctness of classification results. SVMs were initially recommended by Vapnik (1995) [20]. It is widely used in various applications like image recognition Pontil&Verri (1998) [13], bioinformatics Yu et al. (2003) [25] and text classification Joachims (1998) [6].

Class labels are used to classify the input data. This is possible via defining a group of support vectors which are considered to be a part of training inputs.

Along with linear classification, SVMs are well relevant for random classification with the help of data, indirectly plotting their inputs on high-dimensional attribute spaces.

2) Ensemble classification: Ensemble learning helps in enlightening the outcomes of Machine Learning (ML) by linking several models. This strategy produces a notable outcome in contrast to a solitary model. A group of classifiers acquire and then cast their vote. The extrapolative correctness is upgraded but it is challenging to comprehend them Dietterich (2002) [4]. It is beneficial in solving statistical, computational and representational problems. It is not essential to find more precise models, but build models with errors. Ensemble models built to perform classification can misclassify initially.

There are different methods of building ensembles.

- Maximum Vote
- Bagging and Random Forest (RF)
- Chance Injection

- Feature choice Ensembles
- Error Correcting Output Coding (ECOC)

The algorithm is shown below.

Step 1: Form the test set 'T' using 'n' documents in 'X'

Step 2: Form the training set 'TR' using the residual documents in 'X'

Step 3: for every classifier in 'C'.

Make use of classified documents to train the classifier in 'T'.

Utilize the trained classifier to group the documents in 'S'.

Store the resultant labels in the particular class.

Step 4: for every 'x' in the range 1 to s

for every 'y' in the range 1 to s

for every 'z' in the range 1 to k

for every 'n' in the range z+1 to k

if (class[z,x] == class[n,y])

if(M[x,y]==0)

M[x,y]=1;

else

M[x, y] =M[x, y]*2;

Step 5: 'm' is served into the k-means procedure to form document groups.

Step 6: Apply SVM-linear algorithm on 'T' for document categorization.

Step 7: Select the classes conforming to the clusters by finding the class attained in the preceding step.

D. Datasets

The datasets are taken from UCI machine learning repository.

Nearly 76 features are present in the heart disease dataset, but most of the researchers have made use of 14 in the list. The objective of this dataset is to conclude whether a patient is having a heart disease or not. It is numerical value that ranges from 0 to 4. Investigations with the Cleveland database have focused on simply attempting to differentiate existence (values 1, 2, 3, 4) from non-existence (value 0) of heart disease.

In the heart diseases dataset there are 14 attributes 304 Instances, whereas in Chronic Kidney Disease dataset there are 25 attributes 400 Instances and Breast cancer dataset includes 32 attributes 569 Instances. Each has an attribute that is a class like present and not present.

Chronic Kidney Disease dataset includes blood tests and various other measures collected from the patients either with the presence or absence of CKD. The details are collected from nearly 400 patients who were in observation for over period of 60 days. Out of 400 patients, 250 were diagnosed with Chronic

Kidney Disease and 150 were without Chronic Kidney Disease. This variation is represented as "Class" in the dataset. Few important attributes of this dataset are age, Hyper tension, Diabetic, Blood Glucose Random, Blood Urea, Haemoglobin etc.

Wisconsin Diagnostic Breast Cancer (WDBC) is one of the standard datasets considered for Breast cancer diagnosis. It has nearly 699 instances, in which 458 are benign and 241 are malignant with 11 attributes that includes a class attribute.

V. RESULTS AND DISCUSSION

The following figures (Fig. 2-7) show the performance of the benchmarked and the proposed schemes. Table I shows the quantity of Features taken using TLO, KD and DMTLO. Table II illustrates the attributes selected using TLO, KD and DMTLO.

Fig. 2 shows the Accuracy, Precision, Recall, F-measure for the Heart Disease Dataset. It is seen that the proposed DMTLO_Adaboosting offers 6%, 4%, 4%, 2%, and 2% better Accuracy in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboostingand DMTLO_SVM respectively. Similarly it offers 5%, 3%, 3%, 1%, and 1% better Precision in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. The Recall of DMTLO_Adaboosting is 4%, 4%, 1%, 2% and 2% improved when compared to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM respectively. Similarly, the F-Measure of DMTLO_Adaboosting is 4%, 3%, 3%, 2% and 1% improved when compared to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively.

Fig. 3 shows the Time Period and Error rate for the Heart Disease Dataset. DMTLO_Adaboosting offers 80.95%, 61.90%, 61.90%, 33.33% and 23.81% better Time period in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM respectively. Similarly it involves 2.76, 2.46, 2.23, 1.85 and 1.38 times lesser error rate in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively.

Fig. 4 shows the Accuracy, Precision, Recall, F-measure for the Chronic Kidney Disease Dataset. It is seen that the proposed DMTLO_Adaboosting offers 6%, 3%, 7%, 4%, and 2% better Accuracy in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM respectively. Similarly it offers 6%, 3%, 7%, 5%, and 2% better Precision in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. The Recall of DMTLO_Adaboosting is 5%, 3%, 7%, 4% and 2% improved when compared to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM respectively. Similarly, the F-Measure of DMTLO_Adaboosting is 5%, 3%, 6%, 3% and 1% improved when compared to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively.

Fig. 5 shows the Time Period and Error rate for the Chronic Kidney Disease Dataset. DMTLO_Adaboosting offers 65.21%, 39.13%, 86.95%, 60.86% and 26.08% better Time period in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. Similarly it involves 2, 1.58,2, 1.75 and 1.33 times lesser error rate in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively.

Fig. 6 shows the Accuracy, Precision, Recall, F-measure for the Breast Cancer Dataset. It is seen that the proposed DMTLO_Adaboosting offers 5%, 1%, 5%, 3%, and 2% better Accuracy in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. Similarly it offers 5%, 2%, 4%, 3%, and 1% better Precision in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. The Recall of DMTLO_Adaboosting is 5%, 2%, 5%, 4% and 1% improved when compared to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. Similarly, the F-Measure of DMTLO_Adaboosting is 3%, 1%, 6%, 1% and 1% improved when compared to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively.

TABLE I. NUMBER OF FEATURES SELECTED USING TLO, KD AND DMTLO

Dataset	No. of Attribute Selection	Feature Selection using TLO	Feature selection using KD	Feature Selection using DMTLO
Heart disease	14	10	8	12
CKD	25	18	16	19
Breast Cancer	32	20	23	26

TABLE II. ATTRIBUTES SELECTED USING TLO, KD AND DMTLO

| Dataset	Selected Attributes of TLO	Selected Attributes of KD	Selected Attributes of DMTLO
Heart diseases	1,2,3,4,5,6,7,10,12,13	4,5,6,12,13,10,7,3,	12,10,11,8,13,2,4,7,6,5,9
CKD	2,3,4,5,6,10,17,18,19,14,15,11,13,12,9,8,16,20	3,4,5,10,11,12,15,16,19,18,11,8,9,2,14,6	2,3,10,4,5,17,18,19,14,15,6,7,11,12,13,8,9,21,23
Breast Cancer	12,13,11,27,28,8,7,29,6,18,17,16,19,10,15,14,22,21,26	11,12,13,14,17,27,28,29,30,15,16,17,23,22,21,18,19,2,3,4,5,24,8,	12,13,11,27,28,29,26,8,7,25,9,5,18,30,17,16,19,10,15,2,14,1,22,21,6,4
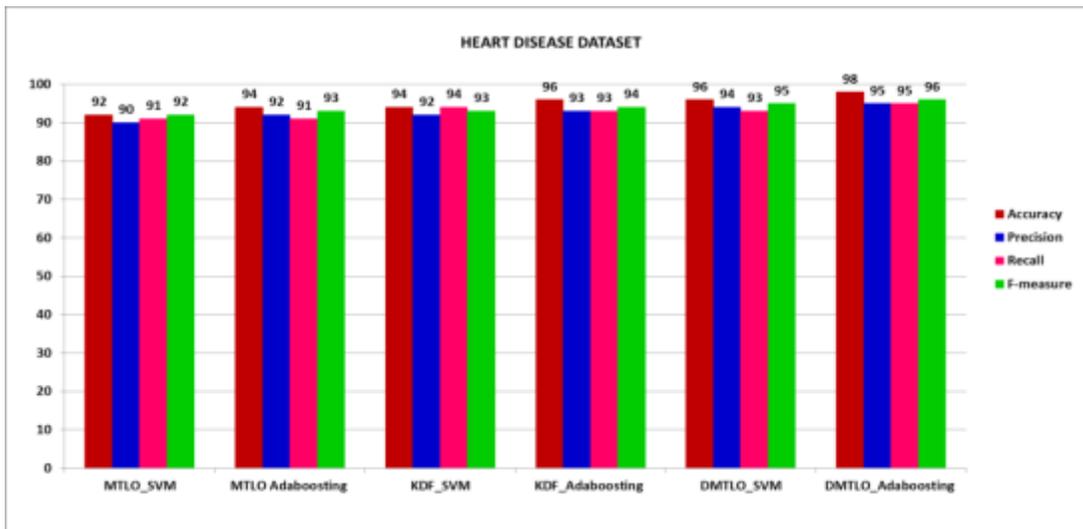
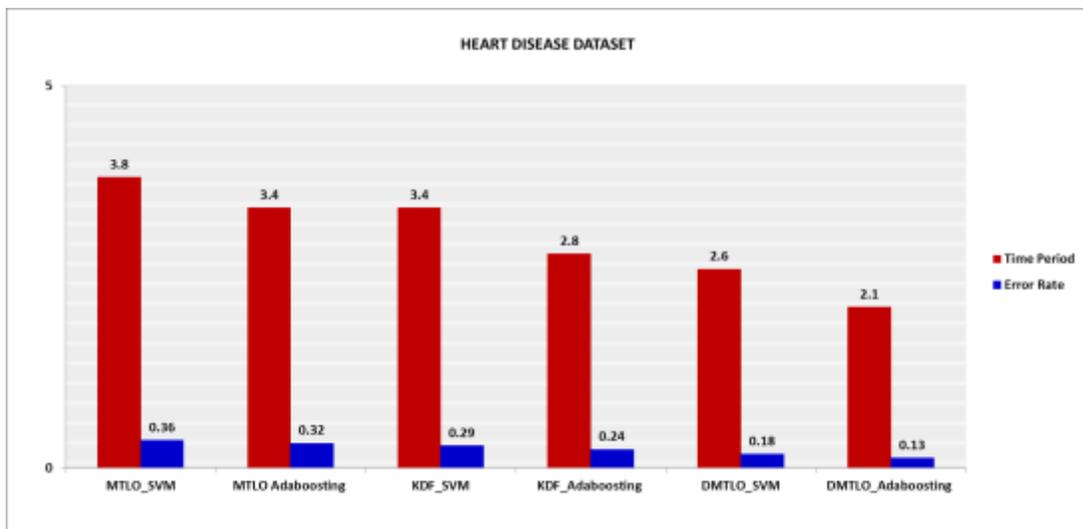Fig. 2. Accuracy, Precision, Recall, F-Measure for the Heart Disease Dataset.

Fig. 3. Time Period and Error Rate for the Heart Disease Dataset.
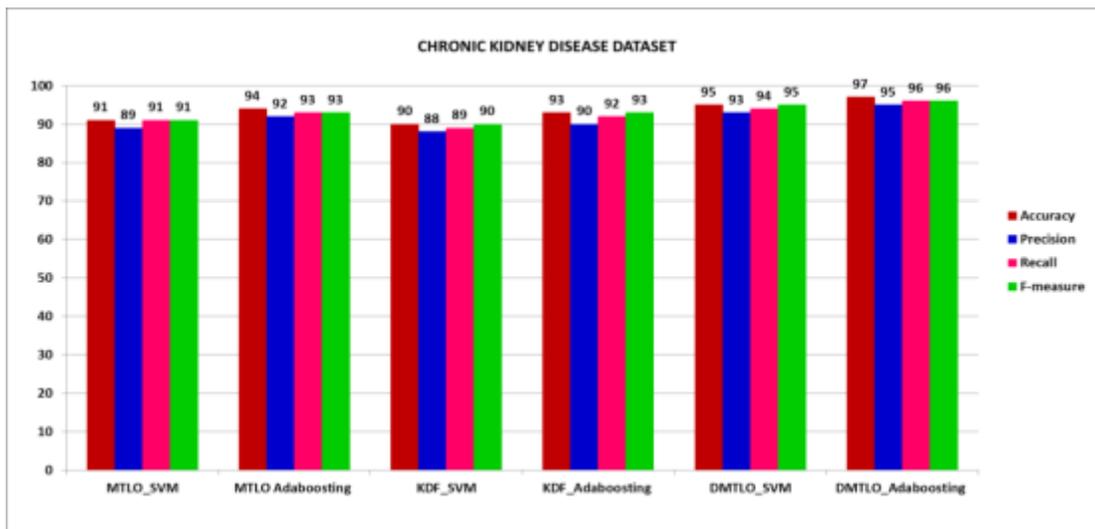

Fig. 4. Accuracy, Precision, Recall, F-Measure for the Chronic Kidney Disease Dataset.
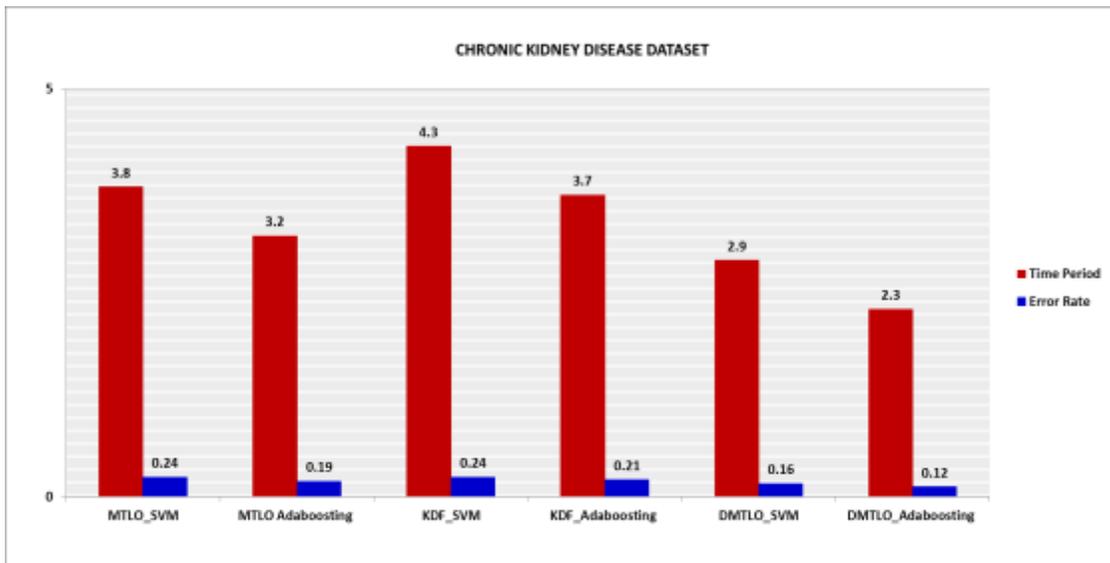
Fig. 5.   Time Period and Error Rate for the Chronic Kidney Disease Dataset.
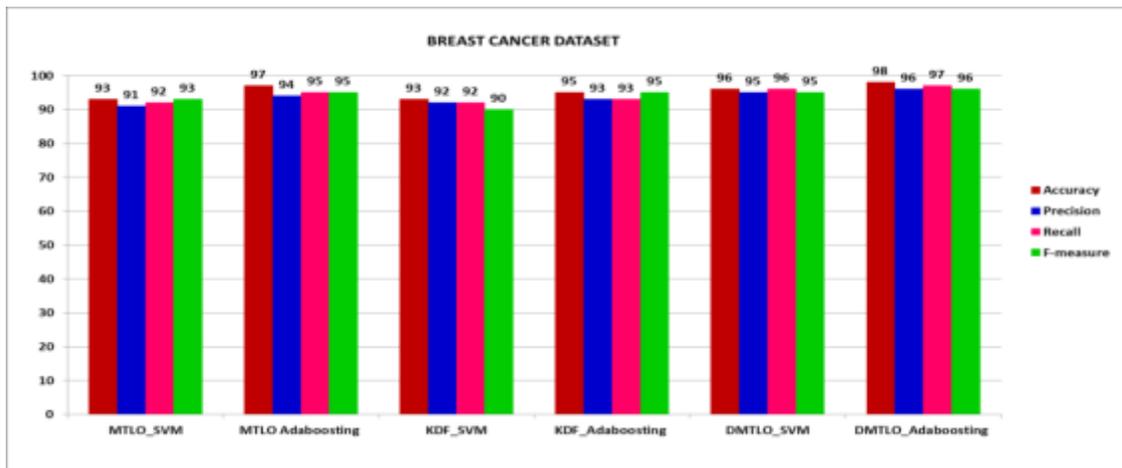

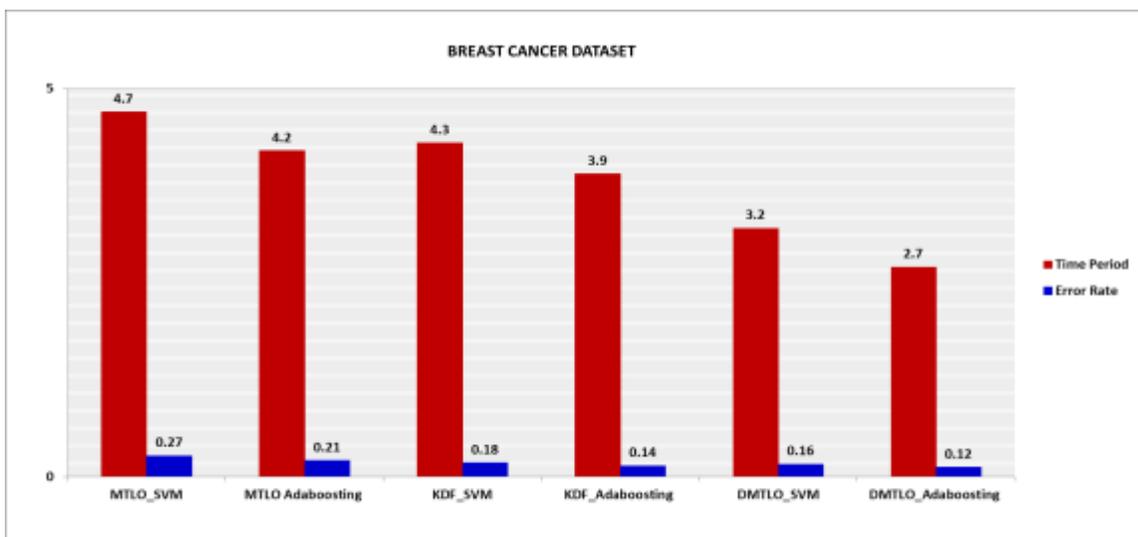Fig. 6.   Accuracy, Precision, Recall, F-Measure for the Breast Cancer Dataset.

Fig. 7.   Time Period and Error Rate for the Breast Cancer Dataset.

Fig. 7 shows the Time Period and Error rate for the Breast Cancer Dataset. DMTLO_Adaboosting offers 74.07%, 55.55%, 59.25%, 44.44% and 18.51% better Time period in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively. Similarly it involves 2.25, 1.75, 1.5, 1.66 and 1.33 times lesser error rate in contrast to MTLO_SVM, MTLO Adaboosting, KDF_SVM, KDF_Adaboosting and DMTLO_SVM, respectively.

VI. Conclusion

In this paper, the outcomes of the proposed system are evaluated for 3 various datasets like Heat disease, chronic kidney disease and Breast cancer. The experimental results are compared with existing Teaching Learning optimization and Kernel Density. The results are analysed in terms of Accuracy, Precision, Recall, F-measure, Time Period and Error Rate. Based on this, it is noticeable that the proposed DMLTO overtakes the existing methodologies.

References

[1] Agrawal, Vartika, Chandra, Satish, 2015. ''Feature Selection using Artificial Bee Colony Algorithm for Medical Image Classification''. In: International Conference on Contemporary Computing (IC3). IEEE, pp. 171–176.

[2] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician, 46(3), 175-185.

[3] Bahassine, Said, Madani, Abdellah, Al-Sarem, Mohammed, Kissi, Mohamed, 2018. Feature selection using an improved Chi-square for Arabic text classification. J.King Saud Univ. – Comput. Inf. Sci. https://doi.org/10.1016/j.jksuci.2018.05.010.

[4] Dietterich, T. G., 2002, "Ensemble learning, The handbook of brain theory and neural networks", MA Arbib, vol. 2, pp. 110-125.

[5] Hafez, A.I., Zawbaa, H.M., Emery, E., Mahmoud, H.A., Hassanien, A.E., 2015. ''An innovative approach for feature selection based on chicken swarm optimization''. 2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR), pp. 19–24.

[6] Joachims, T 1998, 'Text categorization with support vector machines: Learning with many relevant features', Springer Berlin Heidelberg, pp. 137-142.

[7] Lai, Zhihui, Mo, Dongmei, Wong, Wai Keung, Yong, Xu., Miao, Duoqian, Zhang, David, 2017. ''Robust Discriminant Regression for Feature Extraction''. IEEE Trans. Cybernetics. 10.1109/TCYB.2017.2740949.

[8] Mafarja, M.M., Eleyan, D., Jaber, I., Hammouri, A., Mirjalili, S., 2017. ''Binary Dragonfly Algorithm for Feature Selection,''. International Conference on New Trends in Computing Sciences (ICTCS), pp. 12–17.

[9] Moradi, Parham, Gholampour, Mozhgan, 2016. ''A hybrid particle swarm optimization for feature subset selection by integrating a novel local search strategy''. Appl. Soft Comput. 43, 117–130.

[10] Marie-Sainte, Larabi, Alalyani, S., 2018. (in press), N. Firefly Algorithm based Feature Selection for Arabic Text Classification. J. King Saud Univ. – Comput. Inf. Sci. https://doi.org/10.1016/j.jksuci.2018.06.004.

[11] Mazini, Mehrnaz, Shirazi, BabakIrajMahdavi, 2018. (in press), ''Anomaly network based intrusion detection system using a reliable hybrid artificial bee colony and AdaBoost algorithms''. J. King Saud University – Comput. Inf. Sci. https://doi. org/10.1016/j.jksuci.2018.03.011.

[12] Panda, M., 2017. Elephant search optimization combined with the deep neural network for microarray data analysis. J. King Saud Univ. – Computer Inf. Sci. https://doi.org/10.1016/j.jksuci.2017.12.002.

[13] Pontil, M &Verri, A 1998, 'Support vector machines for 3D object recognition', IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 6, pp. 637-646.

[14] Rao, R.V., 2016. Teaching Learning Based Optimization Algorithm And Its Engineering Applications. Springer International Publishing.

[15] Sayed, G.I., Hassanien, A.E., Azar, A.T., 2017. Feature selection via a novel chaotic crow search algorithm. Neural Comput. Appl.

[16] Sayed, G.I., Khoriba, G., Haggag, M.H., 2018. ''A novel chaotic salp swarm algorithm for global optimization and feature selection''. Appl. Intell., 1–20.

[17] Shahbeig, Saleh, SadeghHelfroush, Mohammad, Rahideh, Akbar, 2016. ''A Fuzzy Multi-Objective Hybrid TLBO-PSO Approach to Select the Associated Genes with Breast Cancer''. Signal Process. 131, 58–65.

[18] Thawkar, Shankar, Ingolikar, Ranjana, 2018. (in press), ''Classification of masses in digital mammograms using Biogeography-based optimization technique''. J.King Saud Univ. – Comput. Inf. Sci. https://doi.org/10.1016/j.jksuci.2018.01.004.

[19] Tuo, Shouheng, Yong, Longquan, Deng, Fang'an, Li, Yanhai, Lin, Yong, Qiuju, Lu,2017. ''HSTLBO: a hybrid algorithm based on Harmony Search and Teaching- LearningBased Optimization for complex highdimensional optimization problems''. PLoS One 12.

[20] Vapnik V 1995, 'Support-vector networks', Machine Learning, vol.20, pp. 273–297.

[21] Venkata Rao, R., 2016. Review of applications of TLBO algorithm and a tutorial for beginners to solve the unconstrained and constrained optimization problems. Decision Sci. Lett. 5, 1–30.

[22] Wah, Y.B., Ibrahim, N., Hamid, H.A., Abdul-Rahman, S., Fong, S., 2018. Feature selection methods: case of filter and wrapper approaches for maximising classification accuracy. Pertanika J. Sci. Technol. 26 (1), 329–340.

[23] Wen, Jiajun, Lai, Zhihui, Zhan, Yinwei, Cui, Jinrong, 2016. The L2,1-norm-based unsupervised optimal feature selection with applications to action recognition.PatternRecogn. https://doi.org/10.1016/j.patcog.2016.06.006.

[24] Xu, Y., Zhong, Z., Yang, J., You, J., Zhang, D., 2017. A new discriminative sparse representation method for robust face recognition via L2 regularization. IEEE Trans. Neural Networks Learn. Syst. https://doi.org/10.1109/TNNLS.2016.2580572.

[25] Yu, GX, Ostrouchov, G, Geist, A &Samatova, NF 2003, 'An SVM-based algorithm for identification of photosynthesis-specific genome features', In Bioinformatics Conference Proceedings of the IEEE, pp. 235-243.