Ema-Qpso Based Feature Selection and Weighted Classification by Ls-Svm for Diabetes Diagnosis

Fawzi Elias Bekri, A. Govardhan

Abstract:- In accordance to the fast developing technology now a days, every field is gaining it's benefit through machines other than human involvement. Many changes are being made much advancement is possible by this developing technology. Likewise this technology is too gaining its importance in bioinformatics especially to analyse data. As we all know that diabetes is one of the present day deadly diseases prevailing. With the motivation of our earlier model OFW-ITS-LSSVM, here in this paper we introduce weighted classification with LSSVM to diagnose the diabetes in given blood sample datasets. We derived and proposed a swarm intelligence technique called Escalated Mediocre Agent based Quantum Particle Swarm Optimization EMA-QPSO for feature selection. The feature weights will be identified using a technique DFWQ (dynamic feature weight quantization) that derived from HITS algorithm, which uses in web mining. In contrast to our earlier model OFW-ITS-LSSVM the proposed model is not using pre defined ontology. Further, considering the patient's details we can predict where he has a chance to get diabetes, if so measures to cure or stop it.

Keywords:- machine learning, SVM, Feature reduction, feature optimization, tabu search, Tabu search.

I. INTRODUCTION

In the present situation we can say that diabetes has no cure. In real, it happens due to the lack of insulin which has to do along with glucose in our body. It has to be supplied into our body during loss conditions externally. Indirectly, I is the main cause for fatal heart, kidney, eye and nerve diseases, which can be overcome or prevented by good food habits and body exercises[1]. Over all this, the difficult thing is to differentiate between disease diagnosis and interpretation of diabetes data. For doing this, we are with Support vector Machine(SVM) which was developed by Vepnik[2]. Its work has been tested in many ways[3][12][4]. The utmost advantage with it is that it can even work good with nonlinear functions and it contains Radial basis Functions(RBF) which is even more precise than polynomial and linear kernel functions. The comparison of this SVM with other methods like Combined Neural Networks (CNNs), Mixture of Experts (MEs), Multilayer Perceptrons (MLPs), Probabilistic Neural Networks (PNNs) also revealed that svm methods are perfect. In estimating the diabetes features, Feature Selection is applied. By the above analysis, if we are left with 8 features, we can come down to 4 by this feature selection. It can take out the factors not conserned with the feature set.[5][6][7][8].

PCA (Principal Component analysis) is one of the feature selection method recently gaining its importance being used in image recognision, signal processing, face recognition etc. By applying SVM to disease datasets, it can grab a large circumference of data sets even relevant or not relevant to the diagnosing the disease. But by such features with variation, the diagnosing will not be perfect and so weighted factors are to be developed . And they were contributed by Zhichao Wang[9] giving them weights by their ontological relevance. The LS-SVM technique at last works with 2 parameters for accurate results. Out of many datasets and values came from SVM, the choosing of 2 parameters is very important, If very high features are chosen, some datasets will be missed and if chosen with utter accuracy and care, leads to under-fitting[6,13]. Hence in our earlier work [34], two optimised solutions are to be found out possibly by Intensified Tabu Search(ITS)[14]. The working of this ITS involves 3 phases. PCA, discussed above, is to get rid of irrelevant evidences given by SVM. Then OFW is to calculate the weight of each factor which PCA thought relevant. Then comes ITS which can find out the best possible 2 parameters for SVM so that it may not be under fit or over fit. In this paper we attempt to find a navel model to measure the feature weights without using any prior assigned ontology. And also attempt to derive a scalable swarm intelligence technique called Escalated Mediocre Agent based QPSO, which assumes to be scalable than Intensified Tabu Search. To have a quick look on what paper contains, we shall see the initial data sets of diabetes in section II, then our 1st step of PCA reduction in section III, dynamic feature weight quantization process in section IV, then Escalated Mediocre Agent based QPSO in section V followed by experimental results in later sections.

II. DATASET OVERVIEW

The initial data sets are gathered form UCI Machine Learning Repository[16]. It contains almost 8 categories on a whole and 768 sub categories which is really a very large database. The attributes are choose from these large data sets may be either discrete or continuous with an interval[17]. The large data base, provided now is from the following:

- Pregnant: Number of times of pregnant
- Plasma-Glucose: Plasma glucose concentration measured using a two-hour oral glucose tolerance test. Blood sugar level.
- BMI: Body mass index (w in kg/h in m)
- DPF: Diabetes pedigree function
- TricepsSFT: Triceps skin fold thickness (mm)
- Serum-Insulin: 2-hour serum insulin (mu U/mt)
- DiastolicBP: Diastolic blood pressure (mmHg)
- Age: Age of the patient (years)
- Class: Diabetes onset within five years (0 or 1)

Manuscript Received on June 22, 2012.

Fawzi Elias Bekri, Department of Computer Science & Engineering JNTU, Hyderabad- 500 085, Andhra Pradesh, India

A. Govardhan, Professor of Computer Science & Engineering Principal College of engineering, JNTUH, Jagityal,

III. FEATURE SELECTION

The 1st method which runs for reducing the data base is feature selection. The complexity of data can be reduced so that we can be left eith less datasets and can be more precise. Then comes PCA helping the classification to happen further with the help of statistical measures. The simplification of data by PCA is as follows:
D n-dimension dataset.

M principle axes a1,a2,…. These are orthogonal axes… then, covariance matrix is:

$$s = \left(\frac{1}{L}\right)\sum_{k=1}^{L}(x_k - p)^T(x_k - p)\ \ x_k \in D \quad (1)$$

Where m is the average of samples, and L is the number of samples. Therefore

$$sv_k = \lambda_k v_k \quad k \in 1,....,n \quad (2)$$

Where λ_k is the k^{th} largest Eigen value of s . The m principal components of a given sample $x_k \in D$ are given in the following

$$q = [q_1, q_2, ..., q_n] = A^T x_k \quad A = [a_1, a_2, ..., a_n]: \quad (3)$$

where $q_1, q_2 ...., q_n$ are the principal components of x_k .

LS-SVM: Of all the paper, we discussed the key idea of using SVM brought up by Vapnik[2]which plays a main role in collecting the wide database for our problem. It also has its use in solving pattern recognition and classification problems. The methods present in SVM other than polynomial and linear are its greatest assets which made it to lead global models containing structural risk minimization principle[19]. Though SVM sounds easy due to its extended results, finding the solution is difficult and what all can do is to find sparse solutions. Its difficulty arises from finding nonlinear equations. So as a solution, Suykens and Vandewalle [20] introduced least-squares SVM which results out linear equations. For the new type of SVM also the further proceeding like PCA, OFW and its usage in quantification and classification are applicable and reported in some works[23,24].

In calculation of linear equation, (y=wx+b), we use the 2 axes like regression(x) and dependent variable (y). And the best minimised cost function is

$$Q = \frac{1}{2}w^T w + \frac{1}{2}\gamma\sum_{i=1}^{N} e_i^2 \quad (3)$$

Subject to: $y_i = w^T\phi(x_i) + b + e_i \qquad i = 1,....N$ (4)

The formula's two parts are weight decay the 1st to generalize weights and regression error of training data is the second, where as the parameter indicated by g is to be optimized by the user.

For a better generalization model, the most important criteria is the proper selection of features for the RBF kernel and polynomial kernel.

IV. FEATURE OPTIMIZATION ESCALATED MEDIOCRE AGENT BASED QPSO [EMA-QPSO]

A new Swarm particle is used instead of least good swarm particle so as to obtain optimized QPSO. On putting a quadratic polynomial technique on best fit swarm particles a new equation is obtained, depending on which new particle is recognized. Replacement is possible if the new swarm particle obtained is better than the least good swarm particle and after each search lap the same procedure is followed.
The optimized QPSO is obtained using the following procedure:

Step 1: Begin the swarm.

Step 2: Calculate mbest

Step 3: Particles places should be updated.

Step 4: The fitness value of each particle is calculated.

Step 5: On comparing the current fitness value and the best fitness value (Pbest), whichever is better is considered.

Step 6: Update Pgbest (global best)

Step 7: A new particle is to be searched.

Step 8: On comparing the new particle with worst particle whichever is better is considered.

Step 9: Repeat step 2 till maximum iterations are obtained.

On using the below table the swarm particle can be obtained. The swarm particle can be found using the fallowing.

$t_i = \sum_{k=1}^{3} p_i^2 - q_i^2) * f(r)$	$p = a, q = b, r = c \ for\ k = 1;$ $p = b, q = c, r = a \ for\ k = 2;$ $p = c, q = a, r = b \ for\ k = 3$
$t1_i = \sum_{k=1}^{3} p_i - q_i) * f(r)$	$p = a, q = b, r = c \ for\ k = 1;$ $p = b, q = c, r = a \ for\ k = 2;$ $p = c, q = a, r = b \ for\ k = 3$

$$\overset{o}{x}_i = 0.5 * (\frac{t_i}{t1_i})$$

Where 'a' is considered as best fit swarm particle, 'b' and 'c' are considered as randomly selected swarm particles $\overset{o}{x}_i$ is considered as new swarm particle.

3.4.1 Dynamic Feature Weight Quantification Approach(DFWQ): Measuring weights using HITS algorithm

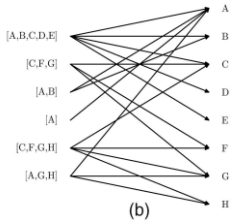
3.4.1.1. Ranking Transactions with HITS

A database of transactions can be depicted as a bipartite graph without loss of information. Let $D = \{T_1, T_2, ...T_m\}$ be a list of transactions and $I = \{i_1, i_2, ...i_n\}$ be the corresponding set of items. Then, clearly D is equivalent to the bipartite graph $G = (D, I, E)$ where

$$E = \{(T, i) : i \in T, T \in D, i \in I\}$$

TID	Transaction
100	[A, B, C, D, E]
200	[C, F, G]
300	[A, B]
400	[A]
500	[C, F, G, H]
600	[A, G, H]

(a)



(b)

Fig1: The bipartite graph representation of a database (a) Database (b) Bipartite graph

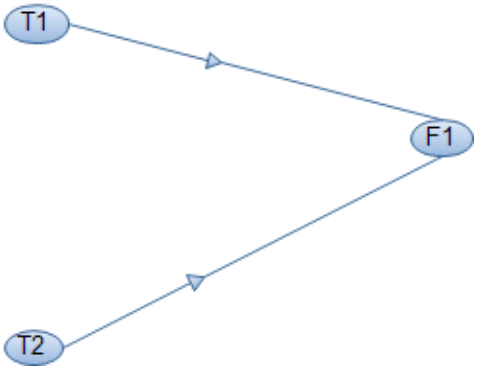
Example 1: Consider the database shown in Fig. 1a. It can be equivalently represented as a bipartite graph, as shown in Fig. 1b. The graph representation of the transaction database is inspiring. It gives us the idea of applying link-based ranking models to the evaluation of transactions. In this bipartite graph, the support of an item i is proportional to its degree, which shows again that the classical support does not consider the difference between transactions. However, it is crucial to have different weights for different transactions in order to reflect their different importance. The evaluation of item sets should be derived from these weights. Here comes the question of how to acquire weights in a database with only binary attributes. Intuitively, a good transaction, which is highly weighted, should contain many good items; at the same time, a good item should be contained by many good transactions. The reinforcing relationship of transactions and items is just like the relationship between hubs and authorities in the HITS model [3]. Regarding the transactions as "pure" hubs and the items as "pure" authorities, we can apply HITS to this bipartite graph. The following equations are used in iterations:

$$auth(i) = \sum_{T:i \in T} hub(T),\ hub(T) = \sum_{i:i \in T} auth(i) ....(1)$$

When the HITS model eventually converges, the hub weights of all transactions are obtained. These weights represent the potential of transactions to contain high-value items. A transaction with few items may still be a good hub if all component items are top ranked. Conversely, a transaction with many ordinary items may have a low hub weight.

The weight measurement described with an example below:

$T1$



Graph1: example relation graph between feature F1 and transactions T1, T2

Let's consider an adjacency matrix of the graph(graph1) created from two transactions and one feature is

$$A = \begin{bmatrix} (T1,T1)\ (T1,T2)\ (T1,F1) \\ (T2,T1)\ (T2,T2)\ (T2,F1) \\ (F1,T1)\ (F1,T2)\ (F1,F1) \end{bmatrix} \rightarrow \begin{bmatrix} 001 \\ 001 \\ 000 \end{bmatrix}$$ with

$$A^t = \begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 0 \\ 1\ 1\ 0 \end{bmatrix}.$$

transpose

$$u = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Assume the initial hub weight vector is:

We compute the authority weight vector by:

$$v = A^t.u = \begin{bmatrix} 0\ 0\ 0 \\ 0\ 0\ 0 \\ 1\ 1\ 0 \end{bmatrix}.\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Then, the updated hub weight is:

$$u = A.v = \begin{bmatrix} 001 \\ 001 \\ 000 \end{bmatrix}.\begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}.$$

This already corresponds to our intuition that feature is the most authoritative, since as of the example it is the only one with incoming edges, and transactions T1 and T2 are equally important hubs.

DFWQ-LSSVM

According to the Conventional LS-SVMs, the given function is performed by the equal contributions from all the characteristics. But, actually, the various characteristics play different roles with various weights. Thus, different contributions from different characteristics can be performed by using the theory proposed by Zhichao Wang [9].

Given,

$\{x_i, y_i\}_{i=1}^{N}$

Represents coaching group and

$\alpha \in R^d$, where α represents the weighted vector.

$$\sum_{i=1}^{d} \alpha_i = 1, \quad \alpha_i \geq 0 \qquad \text{-----(8)}$$

Now, the equation (3) can be used to provide optimal solution to the problem (4), which is as follows:

$$\min \frac{1}{2} \| w \|^2$$

$$s.t. y_i (w.diag(\alpha))$$

Where,

$$diag(\alpha) = \begin{pmatrix} \alpha_1 & 0\text{K} & 0 \\ 0 & \alpha_2\text{K} & 0 \\ \text{K} & \text{K} & \text{K} \\ 0 & 0\text{K} & \alpha_d \end{pmatrix}$$

Substituting (8) and (9) into (5), yields the following new optimization problem:

$$\min_{w,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} \xi_i$$

$$sty_i(w.diag(\alpha)x_i + b) \geq 1 - \xi_i, \geq 0, i = 1, 2, ..., n$$

$$\sum_{i=1}^{d} \alpha_i = 1, \alpha_i \geq 0$$

Thus, we can write the categorization decision method is:

$$f(x) = sign(\sum_{i=1}^{N} L_i y_i K'(X_i, X_j) + b)$$

$K'(X_i, X_j)$ represents the weighted characteristic of the RBF kernel as

$$K'(X_i, X_j) = \exp(-\gamma \sqrt{\sum_{k=1}^{d} \alpha_k (x_{ik} - x_{jk})^2})$$

Intensified Tabu search (ITS) for characteristic selection

As we know, BDF chooses the characteristics based on the betterment of the recognition rate. After considering we came to know that BDF increases the count of support vectors according to the size of the problem [26]. This feature seems to be interesting in producing quick and better decision method, but it applies only if it is connected to the betterment in the recognition rate.

The support vectors and some more characteristics care mainly used to provide a quick and better SVM BDF. Due to this reason, in order to solve the conflict between the complexity and performance, Decision Function Quality (DFQ) criterion is used in association with regularization theory. Thus, SVM makes sure to coach right from the basic i.e. tiny dataset St', where it stands for the primary coaching group St. It will reduce the ambiguity related to the BDF. Even the primary set is also further optimized by using LBG algorithm based on certain assumptions. The basic assumption is to consider parameter k as a variable of problem in choosing model. It is so, because k may not be able to handle all kinds of prototypes generated by LBG algorithm during the process.

Hence, the value of k (i.e. the range of optimization), the characteristic subgroup β, the regularization constant C along with attributes of the kernel such as (σ with gaussian kernel) must be selected for every kernel method K using the model selection method. If we consider θ as a model, k_θ, β_θ, C_θ and σ_θ respectively will be the representatives of the attributes discussed so far. Moreover, q(θ) represents the DFQ criterion for a model θ (c.f. Section 3.1).

The Section 3.3 deals with the presumption of DFQ criterion along with a learning set S_l showing q(θ) ≡ SVM-DFQ(θ, S_l) which is to be optimized for model θ. The optimizing θ^* for q(θ) not being tractable, we decide to define a TS function for choosing a model with optimal intensification and diversification methodologies.

V. DECISION FUNCTION QUALITY(DFQ):

For smooth calculation of the equation, we need DFQ for the theta we have. It can be known by the recognition rate RR with the help of complexity CP of decision function hu. Here comes the q(θ) = $R_R(h_\theta) - C_p(h_\theta)$ be the DFQ[25].

Here, the correct and accurate result from equation can be calculated by using smoothness term and fitting term in terms of recognition rate (RR).Cp indicates the smoothness term. The model complexity of a SVM BDF[25] is given by

$$C_p(h_\theta) = C_{p1} \log_2(\eta sv) + C_{p2} \log_2(\cos t(\beta)) \qquad (5)$$

To discuss the parameters of the function, cp1 and cp2 are trade off between classification rate improvement and complexity reduction. Beta is a Boolean vector with n size of represented features. Ki is to represent cost for ith feature cost(beta) combined to the subset of selected features is:

cost(β) = $\sum \beta_i k_i$. When those costs are unknown, k_i = 1 is used for all features.

Simplification Step:

Reducing training set size is the simplest way to reduce complexity of SVM. This LBG algorithm[25] is used to simplify the dataset . The simplification details are in the below table and can be used in the further discussion:

Table 1: Synopsis of simplification step

Simplification(S,k)
S′ ⇐ ∅ FOR c ∈ {−1, +1} T = {x | (x, c) ∈ S} IF 2^k < |T| THEN T′ ⇐ LBG(T, k) ELSE T′ ⇐ T S′ ⇐ S′ ∪ {(x, c) | x ∈ T′} ENDFOR RETURN S′

Table 2: Synopsis of DFQ calculating for a defined model θ

SVM-DFQ(θ,Sl)
(s_t, s_v) ⇐ Split(S_l) $S_t^{'}$ ⇐ Simplification(s_t,, k_θ) h_θ ⇐ Training SVM($S_t^{'}$, $k_{\beta\theta}, c_\theta, \sigma_\theta$) ($m_{-1}^{correct}$, $m_{+1}^{correct}$) ⇐ Testing BDF(h_θ, S_v) $R_R \Leftarrow \frac{m_{-1}^{correct}}{2_{m-1}} + \frac{m_{+1}^{correct}}{2_{m+1}}$ c_p ⇐ Complexity(h_θ) q(θ) ⇐ R_R _ c_p

VI. DFQ ESTIMATION

The Decision Function Quality (DFQ)[25] criterion of a particular model θ is calculated from a attained dataset Sl. we can observer the elocation of values from the details given in the Table 3 .Let s_t, s_v represents the datasets produced in a random split (Split function in synopsis SVM-DFQ) with $|s_t| = \frac{2}{3}S_l$, $|s_v| = \frac{1}{3}S_l$. s_t, s_v will be signify the databases utilized to train SVM (training dataset) and to identify rate consideration (validation dataset). This dissociation is important in order to overcome the risk of over fitting when empirical estimation is used. The SMO algorithm version of the Torch library [31] is used to realize SVM training step. When SVM training is per-formed with unbalanced class datasets, it is more suitable to use Balanced Error Rate (BER) instead of classical Error Rate for the estimation of recognition rate. Recognition rate formulation (noted R_R) in Table 2 corresponds to BER estimation where m_y represents the number of examples in each class(y ∈ {+1,−1}) and $m_y^{correct}$ the number of examples correctly identified. . The kernel functions k_β utilized for training SVM are decided from a distance $d_\beta : d_\beta(x_i, x_j) = \sqrt{\sum_{l=1}^{n} \beta_l (X_i^l - x_j^l)^2}$. Utilizing d_β in the kernel function, the feature selection problem is embedded in the model selection problem. In the present study Gaussian kernels $K_\beta^G = \exp(\frac{-d_\beta^2}{\lambda_1^2})$ are utilized.

VII. PROPOSED METHOD OF WEIGHTED CLASSIFICATION BY LSSVM

This part explains the desired method (DFWQ-EMAQPSO-LSSVM) for the identifying of diabetes diseases (see figure2). Especially the system works in three stages automatically

1. PCA is applied for feature reduction

2. Best feature weights are estimated using DFWQ

3. EMA-QPSO is employed for finding the optimal values for C and γ.

At first, PCA method is used to identify four features from diabetes dataset. Thus, in feature choosing stage, only large principal components will be utilized. Then, the OFW-LSSVM is used to classify patients, the feature weights which are received by OFW and at last , the MCS algorithm is used to detect the best value for C and γ parameters of OFW-LSSVM. The description of training procedure is :

- Set up parameters of EMA-QPSO and initialize the population of n nests (Algorithm 1)
- Compute the corresponding fitness function $\frac{classified}{total}$ formulated by (total denotes the number of training samples, and classified denotes the number of correct classified samples) for each particle.
- 3. Find the best solution using EMA-QPSO

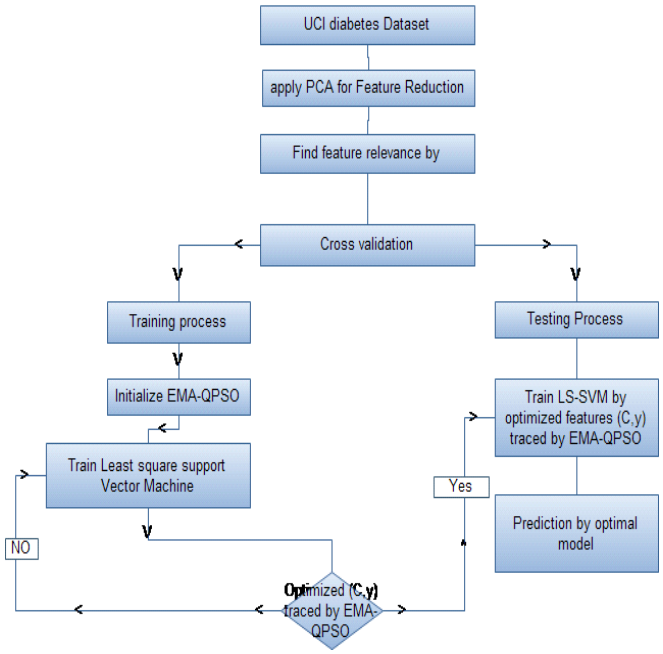


Figure 2: Flowchart of the LSSVM supported by DFWQ-EMAQPSO

VIII. EXPERIMENTAL RESULTS

The proposed model was compared with other popular models like LS-SVM, PCA-LS-SVM, PCA-MI-LS-SVM, MI-CS-SVM, PCA-PSO-LS-SVM classifiers and our earlier model OFW-ITS-LSSVM [34]. We utilized fold cross validation develop the holdout method. The data set was divided into k subsets, and the holdout method was iterated k times. Every time, one of the k subsets is utilized as the test set and rest are put together to form a training set. Then the average error across all k trials is computed (Polat and Günes, 2007). This method was used as 10 -fold cross validation in our experiments. we considered the related parameters of PSO in PCA-PSO-LS-SVM classifier as follows: swarm size was set to 50 ; the parameters C and γ were arbitrary taken from the intervals $[10^{-3},200] and [10^{-3},2]$, respectively.

The inertia weight was 0.9, acceleration constants C_1 and C_2 was fixed to 2, and maximum number of redundancy was fixed to 70. Classification results of classifiers were shown in a confusion matrix. Like displayed in table 4, each cell has the raw number of examples categories to correspond integration of real system results.

Table 4: Confusion matrix

Output/desired	Non-diabetic	Diabetic	Method
Non-diabetics	44	6	LS-SVM
Diabetics	11	17	
Non-diabetics	45	5	PCA-LS-SVM
Diabetics	9	19	
Non-diabetics	44	6	PCA-MI-LS-SVM
Diabetics	4	24	
Non-diabetics	45	5	PCA-PSO-LS-SVM
Diabetics	4	24	
Non-diabetics	48	2	MI-MCS-SVM
Diabetics	3	25	
Non-diabetics	49	1	OFW-ITS-LS-SVM
Diabetics	1	27	
Non-diabetics	50	0	LS-SVM with DFWQ-EMA-QPSO
Diabetics	1	27	

Thus the frequency of disease and a patient is misclassified. Furthermore, Table 5 displays the categories accuracies of OFW-ITS-LSSVM. The present model gets the correct categories accuracy of 95.78% among classifiers on the test set. Determining the test performance of the classifiers is done by addition of specificity and sensitivity that are classified as : Specificity: number of true negative decisions / number of real negative case sensitivity: number of true positive decisions / number of real positive cases.

A true positive decision happens only if the positive expectation of the network mingles with a positive expectation of the physician. A true negative decision happens if the two i.e. network and the physician advice negative expectation.

Table 5: The values of the statistical parameters of the classifiers

Methods	Sensitivity (%)	Specificity (%)	Classification accuracy
LSSVM [32]	73.91	80	78.21
LSSVM with PCA [33]	79.16	83.33	82.05
LSSVM with MI and PCA[2]	80	91.66	87.17
LSSVM with PCA and PSO[2]	82.75	91.83	88.46
SVM with PCA, IM and MCS[2]	92.59	94.11	93.58
OFW-ITS-LSSVM	94.96	97.76	95.78

As per the Table 6, it is observed that utilizing the LSSVM classifier with OFW and ITS, it is easy to get the correct classification accuracy compared to other methods.. Hence It is apt to say that this method gives a high rate of accuracy in identifying of Diabetes disease. The method can also combined with software to help the physicians to take final decision confidently.

Table 6: Classification accuracy: comparing DFWQ-EMAQPSO- LSSVM with OFW-ITS-LSSVM and other methods from literature

Method	Classification accuracy
QDA	59.5
C4.5 rules	67
RBF	68.23
C4.5 (5xCV)	72
Bayes	72.2
Kohonen	72.8
ASR	74.3
DB-CART	74.4
Naïve Bayes	74.5
CART DT	74.7
BP	75.2
SNB	75.4
NB	75.5
kNN	75.5
MML	75.5
RBF	75.7
LVQ	75.8
Semi-Naïve Bayes (5xCV)	76
MLP + BP	76.4
FDA	76.5
ASI	76.6
SMART	76.8
GTO DT (5xCV)	76.8
BFGS quasi Newton	77.08
LM	77.08
LDA	77.5
GD	77.6
SVM (5xCV)	77.6
GDA-LS-SVM	79.16
GRNN	80.21
LDA-MWSVM	89.74
MI-MCS-SVM	93.58
OFW-ITS-LSSVM	95.78
DFWQ-EMAQPSO-LSSVM	96.1

From figure 3, 4, 5 and 6, it is clearly evident that the DFWQ-EMAQPSO-LSSVM model is efficient than our earlier model OFW-ITS-LSSVM [34] in memory usage execution time and scalability.

IX. CONCLUSIONS

Over all the work propose a new automatic method to diagnose Diabetes disease depend on dynamic Feature Weight quantization based Vector Machines and Escalated Mediocre Agent based QPSO. For discarding the other features, Principal Component Analysis was utilized. Later Mutual Information was used to the chose features to weight them depend on their related task of classification. Outcome proves that it devises the accuracy of the method.

And also evident that proposed EMA-QPSO is scalable when compared to intensified tabu search [34] that we proposed in our earlier model. This proposed model also featured a new technique called Dynamic Feature Weights Quantization that helps measure the feature weights without prior ontology. The outcome has proved that the present model is faster and significantly more reliable than OFW-ITS-LSSVM [34] and other models.. The method can also combine with software to help the physicians to take final decision confidently in order to diagnose Diabetic disease.

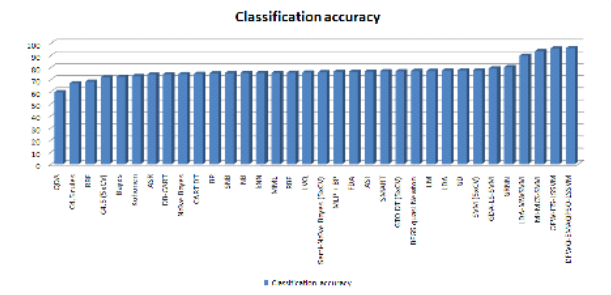


Figure 3: Bar chart representation of comparing DFWQ-EMAQPSO-LSSVM with OFW-ITS-LSSVM and other methods from literature

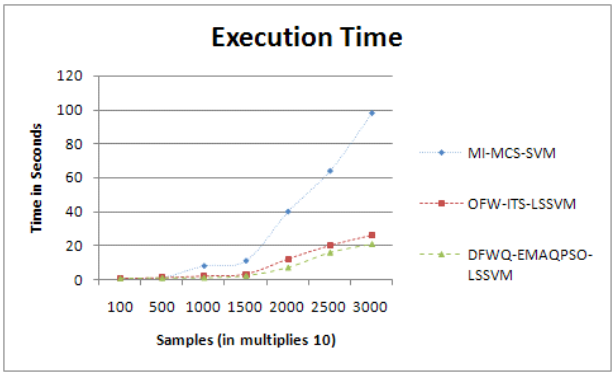


Figure 4: Line chart representation of the execution time comparison

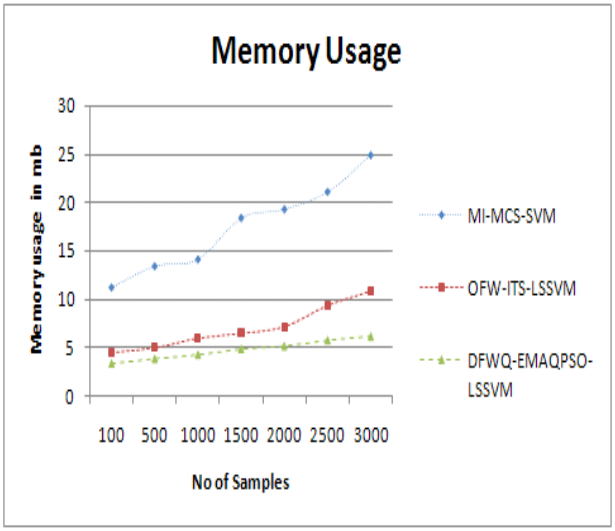


Figure 5: Line chart Representation of the Memory usage comparison

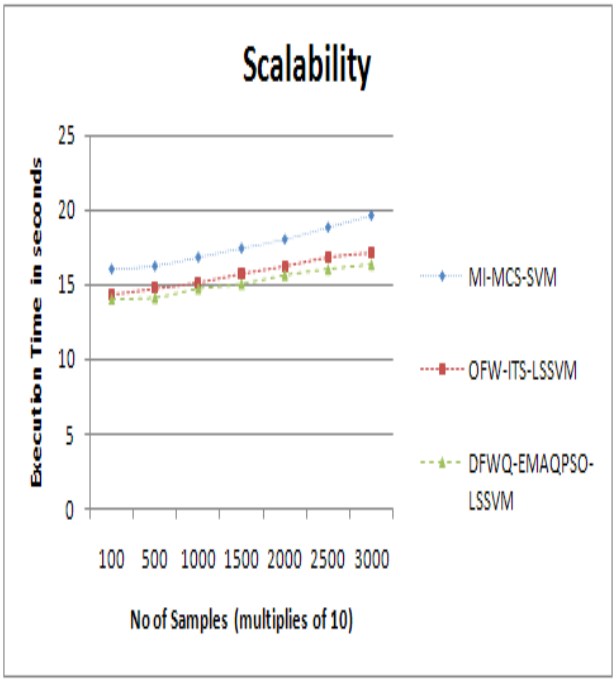


Figure 6: Line chart representation of the scalability comparison

REFERENCES

1. Polat, K., Güneş, S., 2007. An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. Digital Signal Processing 17, 702-710.
2. Vapnik, V., 1995. The Nature of Statistical Learning Theory, New York.
3. Çalişir, D., Doğantekin, E., 2011. An automatic diabetes diagnosis system based on LDA-Wavelet Support Vector
4. Übeyli, E.D., 2007. Comparison of different classification algorithms in clinical decision-making. Expert Systems 24, 17-31.
5. Acır, N., Özdamar, Ö., Güzeliş, C., 2006. Automatic classification of auditory brainstem responses using SVM-based feature selection algorithm for threshold detection. Engineering Applications of Artificial Intelligence 19, 209-218
6. Lin, M., Oki, T., Holloway, T., Streets, D.G., Bengtsson, M., Kanae, S., 2008. Long-range transport of acidifying substances in East Asia—Part I: model evaluation and sensitivity studies. Atmospheric Environment, in press, doi:10.1016/j.atmosenv.2008.04.008
7. Valentini, G., Muselli, M., Ruffino, F., 2004. Cancer recognition with bagged ensembles of support vector machines. Neurocomputing 56, 461-466.
8. Zhang, Y.L., Guo, N., Du, H., Li, W.H., 2005. Automated defect recognition of C- SAM images in IC packaging using Support Vector Machines. The International Journal of Advanced Manufacturing Technology 25, 1191-1196.
9. Lei Zhang , Zhichao Wang "Ontology-based clustering algorithm with feature weights",2010Journal of Computational Information Systems 6:9 (2010) 2959-2966.
10. Karabatak, M., Ince, M.C., 2009. An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications 36, 3465-3469.
11. Mehmet Fatih, A., 2009. Support vector machines combined with feature selection for breast cancer diagnosis. Expert Systems with Applications 36, 3240-3247.
12. Polat, K., Güneş, S., Arslan, A., 2008. A cascade learning system for classification of diabetes disease: Generalized Discriminant Analysis and Least Square Support Vector Machine. Expert Systems with Applications 34, 482-487.
13. Pardo, M., Sberveglieri, G., 2005. Classification of electronic nose data with support vector machines. Sensors and Actuators B: Chemical 107, 730-737.
14. Fred Glover, Tabu search fundamentals and uses, http://leeds-faculty.colorado.edu/glover/TS%20-%20Fundamentals&Uses.pdf, 1995
15. Xing, H.-j., Ha, M.-h., Hu, B.-g., Tian, D.-z., 2009. Linear feature-weighted support vector machine. Fuzzy Information and Engineering 1, 289-305.
16. Asuncion, A., Newman, D. J. (2007) Pima Indians Diabetes Data Set, UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabets, Irvine, CA: University of California, School of Information and Computer Science.
17. Cios, K. J., Pedrycz, W., Swiniarski, R.W., Kurgan, L. A. (2007) Data Mining: A Knowledge Discovery Approach, New York: Springer.
18. Vapnik, V.; Statistical Learning Theory, John Wiley: New York, 1998.
19. Sun J, Xu W, Feng B, A Global Search Strategy of Quantum-Behaved Particle Swarm Optimization. In Proc. of the 2004 IEEE Conf. on Cybernetics and Intelligent Systems, Singapore: 291 – 294, 2004.
20. Suykens, J. A. K.; Vandewalle, J.; Neural Process. Lett. 1999, 9, 293.
21. Suykens, J. A. K.; van Gestel, T.; de Brabanter, J.; de Moor, B.; Vandewalle, J.; Least-Squares Support Vector Machines, World Scientifics: Singapore, 2002.
22. Zou, T.; Dou, Y.; Mi, H.; Zou, J.; Ren, Y.; Anal. Biochem. 2006, 355, 1.
23. Ke, Y.; Yiyu, C.; Chinese J. Anal. Chem. 2006, 34, 561.
24. Niazi, A.; Ghasemi, J.; Yazdanipour, A.; Spectrochim. Acta Part A 2007, 68, 523.
25. Varewyck, M.; Martens, J.-P.; , "A Practical Approach to Model Selection for Support Vector Machines With a Gaussian Kernel," Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on , vol.41, no.2, pp.330-340, April 2011 doi: 10.1109/TSMCB.2010.2053026
26. I. Steinwart. Sparseness of support vector machines - some asymptotically sharp bounds. In NIPS, pages 169–184, 2004.
27. A. Tikhonov and V. Arsenin. Solution of Ill-posed Problems. Winston & Sons, 1977.
28. A. Tikhonov and V. Arsenin. Solution of Ill-posed Problems. Winston & Sons, 1977.
29. A. Tikhonov and V. Arsenin. Ill-Posed Pproblems: Theory and Applications. Kluwer Academic Publishers, 1994.
30. F. Glover and M. Laguna. Tabu search. Kluwer Academic Publishers, 1997.
31. R. Collobert and S. Bengio. SVMTorch: Support vector machines for large-scale regression problems. In Journal of Machine Learning Research, volume 1, pages 143–160, 2001.
32. Least Squares Support Vector Machines for Classification and nonlinear modelling PASE 2000 (2000) by J. A. K. Suykens posted to classification lssvm pattern_recognition regression svm by Borelli on 2006-01-18
33. Davar Giveki, Hamid Salimi, GholamReza Bahmanyar, Younes Khademian, Automatic Detection of Diabetes Diagnosis using Feature Weighted Support Vector Machines based on Mutual Information and Modified Cuckoo Search, arXiv:1201.2173v1, ARXIV, 01/2012
34. Fawzi Elias Bekri, Dr. A. Govardhan "OFW-ITS-LSSVM: Weighted Classification by LS-SVM for Diabetes diagnosis", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 3, 2012

AUTHORS PROFILE

Fawzi Elias Bekri He studied B.Sc IT and M.Sc IT at sikkim manipal University, Manglore. He did his M.Phil at JNTU, Hyderabad. Now he is doing Ph.D at Jawaharlal Nehru Technological University (JNTU), Hyderabad, A. P., India. His areas of interest include Data mining, KDD in healthcare sector, Software Engineering, Databases and Object Oriented Technologies.

Dr. A. Govardhan, Received Ph.D. degree in Computer Science and Engineering from Jawaharlal Nehru Technological University in 2003, M.Tech. from Jawaharlal Nehru University in 1994 and B.E. from Osmania University in 1992. He is working as Professor in CSE & Director of Evaluation of Jawaharlal Nehru Technological University, . He has published around 108 papers in various national and international Journals/conferences. His research of interest includes Databases, Data Warehousing & Mining, Information Retrieval, Computer Networks, Image Processing, Software Engineering, Search Engines and Object Oriented Technologies