ISSN (Online): 2230-7893
www.IJCEM.org

Diagnosis of Diabetes in Female Population of Pima Indian Heritage with Ensemble of BP Neural Network and SVM

Rahmat Zolfaghari

Islamic Azad University, Hashtgerd Branch, Department of Computer Engineering, Tehran, Iran
zolfaghari@alum.sharif.edu

Abstract

Diabetes mellitus is one of the most serious health challenges facing American Natives in the United States today. The publicly available Pima Indian diabetic database (PIDD) at the UCI Machine Learning Lab has become a standard for testing data mining algorithms to see their accuracy in predicting diabetic status from the 8 variables given. In this study we will try to predict the presence of diabetes based on ensemble of SVM and BP NN. The predictive accuracy was 88.04 which was the best accuracy and it was very promising with regard to the other classification systems in the literature for this problem.

Keywords: *Diabetes disease diagnosis, Machine learning approach, Data Mining, Neural Network, SVM, C4.5.*

1. Introduction

Diabetes mellitus is a kind of metabolic diseases in which a person has high blood sugar. There are two general reasons for diabetes: 1- the body does not produce enough insulin**.** Only 5-10% of people with diabetes have this form of the disease(type1)[1]. With the help of insulin therapy and other treatments, even young children with type1 diabetes can learn to manage their condition and live healthy. 2-cells do not respond to the insulin that is produced (type2).

Insulin is the principal hormone that regulates uptake of glucose from the blood into most cells (primarily muscle and fat cells, but not central nervous system cells). Therefore deficiency of insulin or the insensitivity of its receptors plays a central role in all forms of diabetes mellitus.

Humans are capable of digesting some carbohydrates, in particular those most common in food; starch, and some disaccharides such as sucrose, are converted within a few hours to simpler forms most notably the monosaccharide glucose, the principal carbohydrate energy source used by the body.

If the amount of insulin available is insufficient, if cells respond poorly to the effects of insulin (insulin insensitivity or resistance), or if the insulin itself is defective, then glucose will not have its usual effect so that glucose will not be absorbed properly by those body cells that require it nor will it be stored appropriately in the liver and muscles. The net effect is persistent high levels of blood glucose, poor protein synthesis, and other metabolic derangements, such as acidosis.

Hyperglycemia precipitates micro and/or macrovascular complications including: retinopathy, nephropathy, neuropathy, cardiovascular diseases, peripheral vascular diseases and stroke[2]. Insulin and oral hypoglycemic agents are used in the management of type 1 and type 2 diabetes mellitus respectively [1]. This high blood sugar produces the classical symptoms of polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger).

Type 2 diabetes mellitus (T2D) is a complex disease of major public health importance. Its incidence is rapidly increasing in the developed countries. It is estimated that by the year 2030, there will be ~366 million people affected by Type 2 diabetes (T2D) worldwide [3]

Medical diagnostics is quite difficult and visual task which is mostly done by expert doctors. An expert doctor commonly takes decisions by evaluating the current test results of a patient or the expert doctor compares the patient with other patients with the same condition by referring to the previous decisions. Therefore, it is very difficult for a physician to diagnose hepatitis. For this reason, in recent times, many machine learning and data mining techniques have been considered to design automatic diagnosis system for diabetes.

Machine learning is a burgeoning new technology with a wide range of applications. It has the potential to become one of the key components of intelligent information systems, enabling compact generalizations, inferred from large databases of recorded information, to be applied as *knowledge* in various practical ways—such as being embedded in automatic processes like expert systems, or used directly for communicating with human experts and for educational purposes. Presently, however, the field is not well placed to do this. Most research effort is directed towards the invention of new algorithms for learning, rather than towards gaining experience in applying existing techniques to real problems.

In this project we will try to predict the presence of diabetes based on some relevant covariates. We apply ensemble method to increase the accuracy. We used two machine learning method neural network and SVM. The development of ensembles of classifiers is a topic of great activity in the field of supervised learning [4].

An ensemble of classifiers categorizes new examples by combining the decisions made by its components. The individual decisions of the base classifiers are combined by either weighted or unweighted voting. An ensemble can be much more accurate than any of the classifiers of which it is composed. For instance, if in a two classes classification problem, the error of the individual classifiers is under 0.5, i.e. lower than a random classifier, the ensemble error should tend to 0 when the number of classifiers increases [5].

Artificial Neural Networks (ANNs) are well-established tools used with success in many problems such as pattern recognition, classification problems, regression problems differential equations, etc [6].

Another of the most popular algorithm in disease diagnosis is SVM. Remarked for their high predictive power, support vector machines (SVMs) have also been widely used for mining liver fibrosis [7], as well as other medical data [8].

The publicly available Pima Indian diabetic database (PIDD) at the UCI Machine Learning Lab has become a standard for testing data mining algorithms to see their accuracy in predicting diabetic status from the 8 variables given. In this study, a diabetes disease diagnosis was realized by using the ensemble of SVM and NN and tested on Pima Indian dataset. The obtained results of ensemble SVM and NN approach proved that this method is more accurate than the other methods.

2. Related Work

There have been extensive studies of this dataset in the Machine Learning Literature. Various classification algorithms have been applied to the data set, and no algorithm performs exceptionally well.

There have been many studies applying data mining techniques to the PIDD. The independent or target variable is diabetes status within 5 years, represented by the 9^{th} variable (class=1). Smith et al. used a neural network ADAP algorithm using Hebbian learning to build associative models. They used 576 randomly selected cases for training and the remaining 192 test cases showed an accuracy of 76% [9]. Quinlan applied C4.5 and it was 71.1% accurate [10].

Wahba's group at the University of Wisconsin applied penalized log likelihood smoothing spline analysis of variance (PSA). They eliminated patients with glucose and BMIs of zero leaving n=752. They used 500 for the training set, and the remaining 252 as the evaluation set which showed an accuracy of 72% for the PSA model and 74% for a GLIM model [11]

Michie et al. used 22 algorithms with 12-fold cross validation and reported the following accuracy rates on the test set: Discrim 77.5%, Quaddisc 73.8%, Logdisc 77.7%, SMART 76.8%, ALLOC80 69.9%, k-NN 67.6%, CASTLE 74.2%, CART 74.5%, IndCART 72.9%, NewID 71.1%, AC2 72.4%, Baytree 72.9%, NaiveBay 73.8%, CN2 71.1%, C4.5 73%, Itrule 75.5%, Cal5 75%, Kohonen 72.7%, DIPOL92 77.6%, Backprop 75.2%, RBF 75.7%, and LVQ 72.8% [12]

The Multi-Stream Dependency Detection (MSDD) algorithm was used on two-thirds of the dataset for training. No mention is made of deleting any missing values. Accuracy on the one-third for evaluation was 71.33% [13].

Turney applied algorithms (ICET, EG2, CS-ID3, IDX) to the dataset with cost information to determine which algorithm is best when including the costs of classification, tests and classification errors. ICET performed the best [14].

Bayesian neural nets were applied to the dataset using the same deletions and training sample as Wahba. The standard neural network had an accuracy of 75.4%, the Bayesian approach 79.5% [15].

Ripley used the 532 cases that excluded missing insulin levels, with a training set of 200 and a test set of 332. When methods could deal with missing values, he added 100 of the missing insulin cases to the training set. Accuracy rate for logistic regression was 80.2%, MARS and PPR models 77.4%, neural network 77%, k-NN for k=9 75.3%, OLVQ 78.9%, CART 75.6% increasing to 77.7% if 100 incomplete cases were added to the training set [16].

The ARTMAP-IC neural network adds distributed prediction and category instance counting to the basic fuzzy ARTMAP system. This was applied to the database with the same training and test sets as Smith. It achieved an accuracy of 81%, when ARTMAP was 66%, logistic regression 77%, and KNN 77% [17].

Khan used multiplier-free feedforward networks (MFN), and correctly noted that 49% of the patients had zero values for variables that cannot be zero. Nevertheless, he selected all of the diabetic patients and an equal number of non-diabetics to attain a balanced set but with missing variables. It is not clear whether the non-diabetics were randomly selected or selected to minimize missing variables. He took half of these in the form of a balanced subgroup (n=268) as a training set and standardized the 8 variables to zero mean and unit variance. He computed accuracy for the MFN and also discrete weight networks (DWN) and continuous-weight networks (CWN) using the

n=268 evaluation set and results were 78.0%, 76.9%, and 78.4% respectively.

Eklund & Hoang used a number of algorithms on the dataset with 80% training/20% evaluation sets. The problem of missing variables set to zero was ignored. The accuracy of the algorithms tested was C4.5 71.02%, C4.5 rule 71.55%, ITI 73.16%, LMDT 73.51%, and CN2 72.19% .

Liu integrated classification and association rule mining in class association rules (CAR) that were applied to the data set. The best of the 4 CAR models had an accuracy of 73.1% compared to 75.5% for C4.5rules. In many other datasets it tended to outperform C4.5. King et al. used 14 algorithms on the PIDD. They discarded the insulin variable with the most missing cases, leaving n=532. The accuracy of the data mining tools used was CART 76%, Scenario 30%, See5 73%, SPlus 79%, WizWhy 74%, DataMind 69%, DMSK 67%, NeuroShell2--Neural 77%, PcOLPARS 81%, PRW 80%, MQ Expert 77%, NeuroShell2—PolyNet 78%, Gnosis 81%, and KnowledgeMiner 78% .

Classification by aggregating emerging patterns (CAEP) applied to the PIDD was initially 72% accurate, but could only identify 30% of the diabetic patients. After modifications it was 75% accurate.

Although the cited articles use somewhat different subgroups of the PIDD, accuracy for predicting diabetic status ranges from 66% to 81%. While some of these are means of a larger group of randomizations, most are simply one randomization into a training set and test set arriving at accuracy. This run the risk of a particularly good or bad accuracy being a quirk of that particular randomization rather than the method used.

Looking at the 392 complete cases, guessing all are non diabetic gives an accuracy of 65.1%. Since 1988, many dozens of publications using various algorithms have resulted in accuracy rates of 66% to 81%. Rough sets as a data mining predictive tool has been used in medical areas since the late 1980s, but not applied to the PIDD to our knowledge.

We designed ensemble methods: one consisted of three multilayer feedforward neural networks and the other was made up of neural networks, Support Vector Machine (SVM). By using the majority voting technique, the outcomes from individual base classifiers were then combined together. The experimental result showed that the ensemble approach produced a better result than that of each base classifier.

3. Dataset Description

The Pima Indian diabetes database, donated by Vincent Sigillito, is a collection of medical diagnostic reports from 768 records of female patients at least 21 years old of Pima Indian heritage, a population living near Phoenix, Arizona, USA.

The database has n=768 patients each with 9 numeric variables: (1) number of times pregnant, (2) 2-hour OGTT plasma glucose, (3) diastolic blood pressure, (4) triceps skin fold thickness, (5) 2-hour serum insulin, (6) BMI, (7) diabetes pedigree function, (8) age, (9) diabetes onset within 5 years (0, 1). The goal is to use the first 8 variables to predict #9.

There are 500 non-diabetic patients (class = 0) and 268 diabetic ones (class = 1) for an incidence rate of 34.9%. Thus if you simply guess that all are non-diabetic, your accuracy rate is 65.1% (or error rate of 34.9%). We expect a useful data mining or prediction tool to do much better than this. There are a few errors in the data. Although the database is labeled as having no missing values, someone liberally added zeros where there were missing values. Five patients had a glucose of 0, 11 more had a body mass index of 0, 28 others had a diastolic blood pressure of 0, 192 others had skin fold thickness readings of 0, and 140 others had serum insulin levels of 0. None of these are physically possible, and after they were deleted there were 392 cases with no missing values. Studies that did not realize the previous zeros were in fact missing variables essentially used a rule of substituting zero for the missing variables. Ages range from 21 to 81 and all are female. Table 1 shows the characteristics of the selected data sets.

Table 1. Data sets characteristics.

	#Negative	#Positive	#Classes	Numeric Features
Pima	500	268	2	8

There are eight numerical attributes representing physiological measurements and medical test results that are shown in following table:

Table 2. Statistical information about Pima Dataset.

Attribute	Type	Min	Median	Mean	Max	Standard Deviation
1	**continuous**	**0.000**	**2.000**	**3.517**	**17.000**	**3.400**
2	**continuous**	**56.00**	**115.000**	**121.030**	**199.00**	**32.000**
3	**continuous**	**24.000**	**72.000**	**71.500**	**110.0**	**19.400**
4	**continuous**	**7.000**	**29.000**	**29.180**	**99.00**	**16.000**
5	**continuous**	**0.000**	**63.480**	**79.800**	**846.000**	**115.200**
6	**continuous**	**18.200**	**32.800**	**32.890**	**67.10**	**7.900**
7	**continuous**	**0.085**	**0.416**	**0.503**	**2.420**	**0.300**
8	**continuous**	**21.000**	**28.000**	**31.610**	**81.000**	**11.800**

4. Ensemble BP Neural Networks and SVM

For a successful ensemble diabetes disease design, whether each classifier is independent and dissimilar to others play an important role. A good design of base classifiers is their outputs should be divergent to each other as much as possible.

Hence, we propose a three-layer hierarchy multi-classifier Diabetes disease architecture as illustrates in Fig. 1. In the first layer, two groups are constructed and each consists of a set of three base feature selecting classifiers. In order to promote the diversity, different soft computing techniques as well as different feature spaces are applied to the base feature selecting classifiers. In the second layer, the inferences derived from three base feature selecting classifiers in the same group are integrated. Then the outputs from three groups are fused together to produce a final conclusion of the ensemble in the third layer.

But before start ensemble of methods we scale the data that discuss about it in next section.

A. Scaling

The main advantage of scaling is to avoid attributes in greater numeric ranges dominate those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculations. Because kernel values usually depend on the inner products of feature vectors, e.g. the linear kernel and the polynomial kernel, large attribute values might cause numerical problems. Scaling is defined by a linear transformation according to equation (1) where x is the original data, $X_{normalized}$ is the normalized data, X_{max} and X_{min} are the maximum and minimum values of x. At this step 41 numeric features are constructed and normalized to the interval [lowerbound, upperbound].

$$X_{Normalized} = \frac{X - X_{min}}{X_{max} - X_{min}}(upperbound - lowerbound) \quad (1)$$

B. Back Propagation Neural Network Classifiers

A back propagation neural network uses a feed forward structure to solve classification problems by its supervised learning algorithm. It consists of a collection of processing units that are highly interconnected. The network weights are updated by using gradient-based optimization algorithm during the training period. When the network converges to the local minima of error, the output layer of the network will show the result when data is fed into the input layer.

Based on the data given for training, neural networks have the ability to learn how to process intrusion detection tasks. It acts as a computational model to process the network traffic information.

By the use of training procedure, the neural network gains the knowledge to extract the normal and attack signatures from the provided data automatically. With its ability to generalize from learned data, the neural network performs generalization of attacks and fault tolerance to imprecise and uncertain information. At the end of the training procedure, the future network traffic are then identified as whether malicious attacks or normal usage behavior.

C. Support Vector Machine Classifiers

Linear discrimination problem can be stated as a set of N entities in the feature space $X=\{x_0, x_1, x_2, \ldots, x_p\}$ is partitioned in two classes, sometime referred to as patterns, a "yes" class and a "no" class, such as for instance a set of banking customers in which a, typically very small, subset of fraudsters constitutes the "yes" class and that of the others the "no" class. The problem is to find a function $u=f(x_0, x_1, x_2, \ldots, x_p)$ that would discriminate the two classes in such a way that u is positive for all entities in the "yes" class and negative for all the entities in the "no" class. When the discriminant function is assumed to be linear so that $u=w_1*x_1+w_2*x_2+\ldots+w_p*x_p+w_0$ at constant $w_0, w_1, \ldots, w_p$, the problem is of linear discrimination. It differs from that of the linear regression in only that aspect that the target values ui here are binary, either "yes" or "no", so that this is a classification rather than regression, problem.

To make it quantitative, define $u_i=1$ if i belongs to the "yes" class and $u_i= -1$ if i belongs to the "no" class. The intercept w_0 is referred to, in the context of the discrimination/classification problem, as bias.

A linear classifier is defined by a vector *w* so that:

If $\hat{u}_i = \langle w,x_i \rangle > 0$, *predict* $\hat{u}_i=1$;
If $\hat{u}_i = \langle w,x_i \rangle < 0$, *predict* $\hat{u}_i= -1$;
that is, $\hat{u}_i = sign(\langle w,x_i \rangle)$.

Algorithm 1. Linear classifier Equation.

D. Combination Methods

Besides the notability of multiplicity among the base classifiers, the right choice of a combination method is also an important issue in creating a supreme performance. A variety of combination methods have been reported for combining the outputs of the base classifiers into an ensemble result.

According to their characteristics, they can be classified as linear combination methods, non-linear methods, statistical based methods, and computationally intelligent methods.

Linear combination method is the simplest method to fuse base classifiers' outputs together. Summation and average are the popular ways for the combination. Non linear method such as majority voting is used when the output of classifier is a ranked list of classes in accordance with the degrees

of belief on classes the input pattern belongs to. Statistical-based methods are Dempster-Shafer techniques and Bayesian combination methods.
The computationally intelligent method is based on computational intelligence techniques such as fuzzy logic, neural networks, and genetic algorithms.

E. Ensemble BP NN with SVM

Ensemble methods have received considerable attention in the machine learning community to increase the effectiveness of classifiers. In order to construct a good ensemble classifier, the ensemble needs to construct accurate and diverse classifiers and to combine outputs from the classifiers effectively [18,19]. There exist several methods to obtain and combine the diverse classifiers.
In bagging [20], a training dataset is divided into several different subsets that may be overlapping. After that, a machine learning algorithm is trained on each subset. Then, the majority voting scheme is used to combine the class-votes of the different classifiers. If the outputs of the different classifiers are strongly uncorrelated, the ensemble may correct for independent mistakes by single classifiers and this improves the classification accuracy. Constructing and combining a set of classifiers is more complicated in boosting. Boosting methods construct a set of classifiers in a sequential way. First one classifier is trained on all data, and then examples that are misclassified by the first classifier get higher weights in the training process of the next classifier. This is repeated until the whole set of classifiers has been trained. The final ensemble uses a weighted majority voting scheme where the weight of a classifier is dependent on the accuracy of the classifier.
Another ensemble method is the hierarchical mixtures of experts (HME) architecture. In the HME there is a gating network that learns to partition the input space in different regions where different classifiers are used to learn and predict the examples falling in their different regions. The HME exploits the divide and conquer principle, but it is more complicated to use together with SVMs.
Stacking is another ensemble method that learns to combine the outputs of different classifiers. First different
classifiers are trained, and then another classifier receives as inputs all the predictions of the different classifiers and is trained to optimally combine the different classifier outputs.

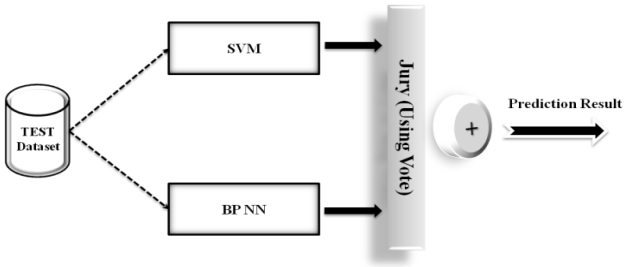

Figure 3. Ensemble of SVM and BP NN.

In our proposed method we used stacking classifiers to combine different SVM and BP NN trained with different diabetes diagnosis (Fig. 3), and this led to better results than using a single SVM or BP NN.

5. Experimental Results

The best value of kernel function's parameters is unknown for us so we use grid search technique to obtain these parameters. To receive this goal we computed accuracy for combination of (C, γ) and the value that can attain the best accuracy, consider as values for function's parameters. by increasing the number of iterations the accuracy is increased. Also increase of iteration effects on running time and lead to growth of accuracy (Fig. 4).

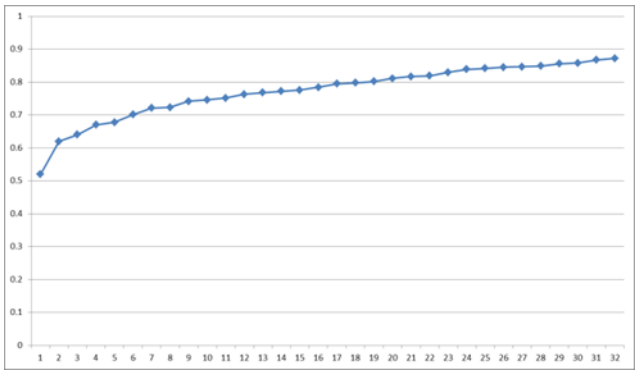

Figure 4. Accuracy Versus #Iterations.

There are different metrics to evaluate the Diabetes disease diagnosis system performance. These metrics are as follows: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). TP is the number of correct predictions that an instance is negative; FN is the number of incorrect predictions that an instance is negative; FP is the number of incorrect of predictions that an instance is positive; TN is the number of correct predictions that an instance is positive.
A confusion matrix contains information about actual and predicted classifications done by a classification system. Performance of such a system is commonly evaluated

ISSN (Online): 2230-7893
www.IJCEM.org

using the data in the matrix. Table 5 shows the confusion matrix for a two class classifier:

Table 3. Representation of confusion matrix.

	Actual	
Predicted	*Negative*	*Positive*
Negative	TN	FN
Positive	FP	TP

False Alarm Rate (FAR), Detection Rate (DR), Recall and F-Measure are some performance evaluation metrics that compute as follows:

$$FAR = \frac{FP}{FN+TP} \quad (2)$$

$$DR = Precision = \frac{TP}{FP+TP} \quad (3)$$

$$Recall = \frac{TP}{FN+TP} \quad (4)$$

$$F - Measure = \frac{2*DR*Recall}{DR+Recall} \quad (5)$$

In Fig. 5 we indicate the distribution of records among two classifiers and upon a weight (the accuracy of each classifiers for each class) we decide the final label of them. Table 4 shows the confusion matrix of our method.

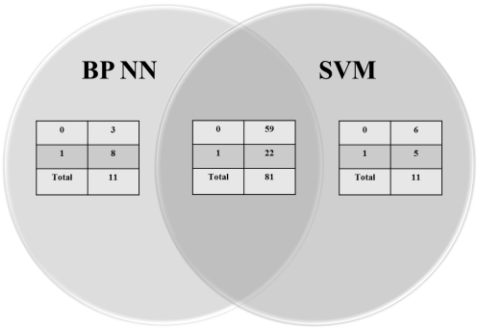

Figure 5. Distribution of patterns between two classifieres.

Table 4. Confusion Matrix Of Ensemble of SVM and BP NN.

	Negative Class	Positive Class
Predict Negative	58	5
Predict Positive	6	23
FAR	FAR_N=0.0781	FAR_P =0.2142
DR	DR_N=0.9206	DR_P=0.7931
Recall	$Recall_N$ =0.9062	$Recall_P$ =0.8214
F-Measure	$F\text{-}Measure_N$=0.9133	$F\text{-}Measure_P$=0.8070
Time (per Sec.)	~ 30	
Accuracy	0.8804	

The prediction method was compared with other competing prediction methods for Pima dataset. The results are summarized in table 5 and indicate the Error Rate of our method obviously less than other methods.

Table 5. Confusion Matrix Of Ensemble of SVM and BP NN

Algorithm	Max Storage	Time (sec.)		Error Rate		Rank
		Train	*Test*	*Train*	*Test*	
Discrim	338	27.4	6.5	0.220	0.225	4
Quadisc	327	24.4	6.6	0.237	0.262	12
Logdisc	311	30.8	6.6	0.219	0.223	2
SMART	780	3762.0	*	0.177	0.232	5
ALLOC80	152	1374.1	*	0.288	0.301	22
K-NN	226	1.0	2.2	**0.000**	0.324	23
CASTLE	82	35.3	4.7	0.260	0.258	11
CART	144	29.6	0.8	0.227	0.225	10
IndCART	596	215.6	209.4	0.079	0.271	15
NewID	87	9.6	10.2	0.000	0.289	20
AC	373	4377.0	241.0	0.000	0.276	19
Baytree	68	10.4	0.3	0.008	0.271	15
NaiveBay	431	25.0	7.2	0.239	0.262	12
CN2	190	38.4	2.8	0.010	0.289	20
C4.5	61	11.5	0.9	0.131	0.270	14
ITrule	60	31.2	1.5	0.223	0.245	7
Cal5	137	236.7	0.1	0.232	0.273	9
Kohonen	62	1966.4	2.5	0.134	0.273	18
DIPOl92	52	35.8	0.8	0.220	0.224	3
Backprop	147	7171.0	0.1	0.198	0.248	8
RBF	179	4.8	0.1	0.218	0.243	6
LVQ	69	129.5	1.2	0.101	0.272	17
Our method	96	26.8	3.5	0.094	**0.2196**	1

6. Conclusion

The present paper proposes an effective biological machine learning algorithm. Proposed ensemble SVM and BP NN testes on one of the widely studied problems in bioinformatics, which is the Diabetes diagnosis. For this problem, our results show 88.04% accuracy that is the best accuracy up to now for this dataset. To conclude, the prediction method described in this paper can be contributed to experiments.

References

[1]. American Diabetes Association Reports of the Experts Committee on the Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*. 2001;23:S4–19

[2]. Danny Meetoo, Peter McGovern, Reema Safadi "An epidemiological overview of diabetes across the world" British Journal of Nursing ,2007,pp. 1002 - 1007 .

[3]. Derosa G "Efficacy and tolerability of pioglitazone in patients with type 2 diabetes mellitus: comparison with other oral antihyperglycaemic agents" Drugs. 2010 . - pp 1945-1961
[4]. Dietterich, T. G., 'An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization", Machine Learning, 2000, pp. 139-158.
[5]. Dietterich, T.G., "Machine Learning Research: Four Current Directions", Artificial Intelligence Magazine, 1997, pp. 97-136.
[6]. I.Tsoulos, D.Gavrilis, E.Glavas, "Neural network construction and training using grammatical evolution", Science Direct Neurocomputing Journal, Vol.72, Issues 1-3, December 2008, pp. 269-277.
[7]. Jiang Z, Yamauchi K, Yoshioka K, Aoki K, Kuroyanagi S, Iwata A, et al, "Support vector machine-based feature selection for classification of liver fibrosis grade in chronic hepatitis C", J Med Syst, 2006, pp. 20:389–94.
[8]. Michie, D., D. J. Spiegelhalter, et al., "Machine learning, neural and statistical classification". New York, Ellis Horwood. 1994.
[9]. Smith, J. W., J. E. Everhart, et al. "Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. Proceedings of the Symposium on Computer Applications and Medical Care (Washington, DC). R. A. Greenes. Los Angeles, CA, IEEE Computer Society Press, 1988, pp. 261-265.
[10]. Quinlan, J. R. "C4.5: programs for machine learning", San Mateo, Calif., Morgan Kaufmann Publishers, 1993.
[11]. Wahba, G., C. Gu, et al., "Soft Classification, a.k.a. Risk Estimation, via Penalized Log Likelihood and Smoothing Spline Analysis of Variance", The mathematics of generalization: the proceedings of the SFI/CNLS Workshop on Formal Approaches to Supervised Learning, Santa Fe, Addison-Wesley Pub , pp. 331-360, 1992.
[12]. Michie, D., D. J. Spiegelhalter, et al., "Machine learning, neural and statistical classification", New York, Ellis Horwood, 1994.
[13]. Oates, T., "MSDD as a Tool for Classification", EKSL Memorandum, Department of Computer Science, University of Massachusetts at Amherst, 1994.
[14]. Turney, P. D., "Cost-Sensitive Classification: Empirical Evaluation of a Hybrid Genetic Decision Tree Induction Algorithm." Journal of Artificial Intelligence Research **2**: 369-409, 1995.
[15]. Bioch, J. C., O. van der Meer, et al., "Classification using Bayesian neural nets",The 1996 IEEE International Conference on Neural Networks, 1488-1493,1996.
[16]. Ripley, B. D., "Pattern recognition and neural networks", Cambridge ; New York, Cambridge University Press,1996.
[17]. Carpenter, G. A. and N. Markuzon., "ARTMAP-IC and medical diagnosis: instance counting and inconsistent cases", Neural Networks, **11**(2): 323-336,1998.
[18]. T. G. Dietterich. "Ensemble methods in machine learning", In Multiple Classifier Systems, Springer-Verlag, 1–15, 2000.
[19]. L. Breiman. "Bagging predictors. Machine Learning", Vol. 24(2):123–140, 1996.
[20]. Y. Freund and R. E. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting", In European Conference on Computational Learning Theory, 23–37, 1995.

Mr Rahmat Zolfaghari is presently working as faculty in Islamic Azad University Hashtgerd Branch, Department of Computer Engineering, Tehran Iran, He is having 12 years experience both in industry and academia, He received his Software Engineering Bachelor (BS) of Shahid Beheshti University in Iran and Software Engineering Master (MS) of Sharif University of technology in Iran, His research interests are Database, Software Design and Modelling