Diagnosis of Diabetes using a Weight-Adjusted Voting Approach

Lin Li

Department of Computer Science & Software Engineering
Seattle University
Seattle, WA 98122, USA
lil@seattleu.edu

Abstract—**Diabetes is a worldwide public health challenge with a yearly increasing incidence. Many approaches using different machine learning classifiers have been developed for automatic diagnosis of diabetes. However, they mostly rely on a single classifier or a hybrid model to make the diagnosis decision, which might be weaker than a voted decision of multiple classifiers. In this study, we present an approach that combines three classifiers (i.e. support vector machine, artificial neural network, and naïve bayes) to diagnose diabetes. The approach can adjust each classifier's weight based on their ability and history of making correct predictions. A rule that mixes majority voting and weights of classifiers was proposed and applied for the final diagnosis decision. The Pima Indians diabetes data set (268 diabetes patients and 500 normal subjects) was used in the work. A wrapper method was adopted to select features for classification. An experimental comparison of our method with other voting strategies and each single classifier used in our study demonstrated that our approach performed better in sensitivity.**

Keywords—*diabetes diagnosis, feature selection, mixture of classifiers, support vector machine, artificial neural network, naïve bayes, weight adjusting, voting*

I. INTRODUCTION

Diabetes, the seventh leading cause of death in the United States in 2010, is a prevalent disease with an increasing incidence worldwide. In 2012, 29.1 million Americans had diabetes, of which 27.8% were undiagnosed [1]. There are three types of diabetes: type 1, type 2, and gestational diabetes. In adults, type 2 diabetes, either the body does not use insulin properly or the pancreas does not produce enough insulin, accounts for about 90% - 95% of all diagnosed cases [1]. Diabetes patients might suffer multiple complications such as blindness, hypertension, cardiovascular disease, hypoglycemia, hyperglycemia, kidney disease, etc. It has been shown that early identification of type 2 diabetes can help prevent or delay 80% of complications [2] and reduce the number of deaths caused by the complications. Therefore, it is desirable to correctly identify people at risk of diabetes in a timely manner.

To diagnose diabetes, a physician has to explore patient's data and consider many factors (e.g. family history, age, body mass index, etc.). A physician's diagnosis can be subjective and is highly dependent on the experiences. Hence, many automated classification systems that use machine learning approaches have been developed to help physicians obtain an objective second opinion for diagnosis decision. A variety of classifiers have been utilized for diagnosis, such as artificial neural network [3 - 5], support vector machine [6], naïve bayes [7], decision tree [7], nearest-neighbor [7, 8], etc. In addition, hybrid models that harness the power of different classifiers have also been proposed [9 - 11]. However, the diagnosis decision based on the classification result of a single classifier or a hybrid model only might be weak. Different classifiers probably offer contradictory classification results while providing complementary information. Therefore, it is helpful to combine the decisions of multiple classifiers. If the decision-making is based on a group of classifiers which takes individual opinion of each classifier into consideration, the misclassified data - especially the patients who are undiagnosed by a certain classifier might be correctly diagnosed due to the correct decisions of other classifiers.

There are a number of methods for combining classifiers, including mixture of experts [12 - 14], voting [15, 16], boosting [17], bagging [18], etc. A few of them have been adapted for diagnosis of diabetes [16, 19]. Some of these methods do not consider the weight of classifiers or each classifier has equal weight. But in fact, the weights of classifiers should be different and should be counted in the final decision. It makes more sense to give larger weights to classifiers which often make correct decisions and smaller weights to classifiers which usually make wrong decisions. On the other hand, some other methods adjust the weights of classifiers based on their power of prediction. In the meanwhile, they iteratively adjust the weights of instances, meaning that hard-to-classify instances get higher weights, which again influence the predictions of classifiers. The iterative interference between classifiers and instances makes the decision-making procedure complicated and time-consuming.

In this work, we present an approach that combines an ensemble of three classifiers - support vector machine (SVM), artificial neural network (ANN), and naïve bayes (NB) for diagnosis of diabetes by using the Pima Indian diabetes (PID) data set [20]. The contributions of this study are twofold. First, an algorithm was devised to dynamically adjust each classifier's weight according to their power and history of making correct predictions of diabetes. Second, a new voting rule was proposed to mix the majority voting and the classifiers' weight for the final diagnosis decision of the ensemble. Compared with other voting approaches, our method

effectively improves the sensitivity of the ensemble, which is very important in disease diagnosis. Improved sensitivity indicates that more diabetes patients can get in-time alarm and take appropriate actions such as changing life-style to prevent or delay the complications of diabetes.

II. MATERIALS AND METHODS

A. Subjects

The Pima Indian diabetes data set, one of the mostly used datasets for prediction of type 2 diabetes, is from the University of California, Irvine (UCI) machine learning repository [21]. The data set has 268 diabetes patients and 500 normal subjects. All subjects are females who are at least 21 years old and of Pima Indian heritage. Each subject has eight attributes, including:

- Number of times pregnant
- Plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGTT)
- Diastolic blood pressure (mm Hg)
- Triceps skin fold thickness (mm)
- 2-hour serum insulin (mu U/ml)
- Body mass index (weight in kg/(height in m)^2)
- Diabetes pedigree function
- Age (years)

We assumed zero values that do not biologically make sense indicate missing data. So we removed subjects with zeros in places where they are biologically impossible, such as the diastolic blood pressure, and body mass index. After data removal, the data set used in our experiment included 250 diabetes patients and 478 normal subjects. Mean and standard deviation (S.D.) of attributes are presented in TABLE I.

TABLE I MEAN AND STANDARD DEVIATION OF ATTRIBUTES FOR THE DIABETES PATIENTS AND NORMAL SUBJECTS

Group	Diabetes mean (S.D.)	Normal mean (S.D.)
Number of times pregnant	4.92 (3.71)	3.31 (3.03)
Plasma glucose (2 hours OGTT)	142.54 (30.02)	111.06 (24.96)
Diastolic blood pressure	75.33 (12.34)	70.92 (12.15)
Triceps skin fold thickness	23.35 (17.38)	20.36 (14.71)
2-hour serum insulin	107.56 (140.87)	71.91 (100.01)
Body mass index	35.17 (6.95)	30.78 (6.99)
Diabetes pedigree function	0.56 (0.38)	0.43 (0.30)
Age	37.41 (11.15)	31.29 (11.67)

B. Feature Selection and Classification

We used Weka [22] to conduct feature selection on the data set to find the subset with the smallest number of feature achieving the best classification performance. Both filter methods and wrapper methods were tested. Compared with the feature sets selected by filter methods, the five features - plasma glucose (2 hours OGTT), 2-hour serum insulin, body mass index, diabetes pedigree function, and age selected by a wrapper method could provide better classification outcome. Hence, these five features were used in our experiments to train learning models.

We also used Weka to perform classification, using three classifiers – SVM, ANN, and NB. The data set was partitioned into three disjoint sets. The first part (83 diabetes patients, 159 normal) – dataset 1, was used in training models for each single classifier. The second part (83 diabetes patients, 159 normal) – dataset 2, was used for testing the performance of each classifier; In the meanwhile, it was used, together with the predictions of each classifier, to adjust classifiers' weight in the ensemble for making the final diagnosis decision (details will be discussed in the next sub-section). The third part (84 diabetes patients, 160 normal) – dataset 3, was used to test the performance of the ensemble of classifiers making decisions based on the proposed weight-adjusted voting approach. Figure 1 demonstrates the procedure of training and testing.
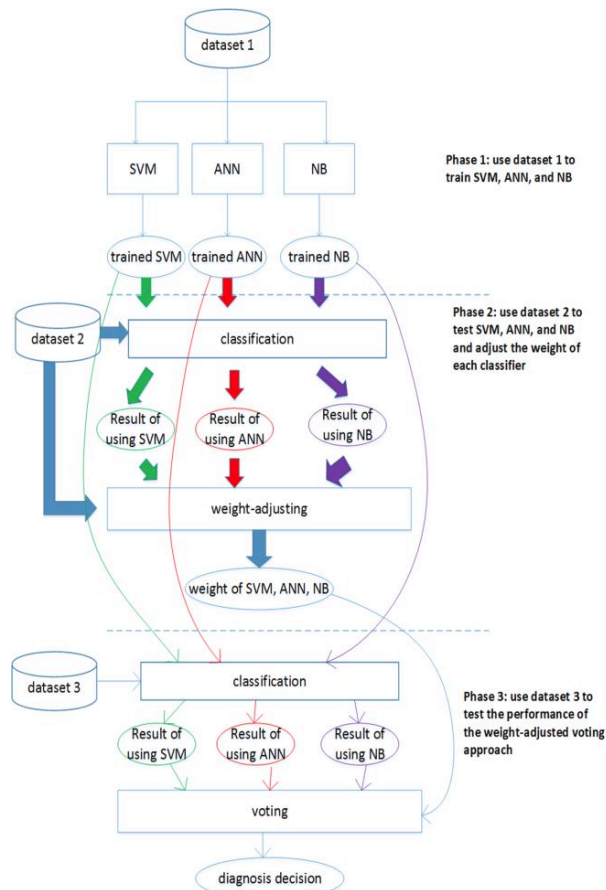

Figure 1. The procedure of training and testing in our experiment

C. Weight-Adjusted Voting

The idea behind our weight-adjusted voting approach is as following. We note that this approach is for an ensemble of three classifiers. For a given instance, if a classifier, say A, makes a correct prediction while the other two classifiers, say B and C, make wrong predictions, then the classifier A's weight is increased and the weights of B and C are equally decreased. In this case, classifier A gets more credit as it is the only one that correctly classifies the given instance. On the other hand, if classifier A makes a wrong prediction while B and C make correct predictions, then A's weight drops and other two classifiers' weights equally step up. In this situation, however, B and C get less credit as there are two classifiers which make correct predictions. Moreover, if three classifiers all make correct predictions or all make wrong predictions, then each of them does not change weight. The algorithm for adjusting the weight of classifier is described as follows.

ALGORITHM 1: Adjusting the Weight of Classifier.

Input:
- N: the number of instances in dataset 2
- Array A of size N which saves the prediction results of dataset 2 using classifier A
- Array B of size N which saves the prediction results of dataset 2 using classifier B
- Array C of size N which saves the prediction results of dataset 2 using classifier C
- Array L of size N which saves the actual class membership of each instance in dataset 2
- $step1$: the step size used to change weights when 2 classifiers make correct decisions and 1 classifier makes wrong decision
- $step2$: the step size used to change weights when 1 classifier makes correct decision and 2 classifiers make wrong decisions

Output:
- W: a vector $\{W_A, W_B, W_C\}$ that stores the weight of classifier A, B, and C respectively

1   compare A and L, B and L, C and L to build a matrix R of N rows and 3 columns such that $R[i][j]$ records whether the jth classifier correctly diagnoses the ith instance (0 for wrong decision, 1 for correct decision).
2   $W_A = W_B = W_C := \frac{1}{3}$
3   **for** $i := 0$ **to** $N - 1$ **do**
4   **if** there are two 0s and one 1 in the vector $R[i]$
5   weight of correct classifier increases $2 \times step2$
6   weight of each wrong classifier decreases $step2$
7   **if** there are two 1s and one 0 in the vector $R[i]$
8   weight of wrong classifier decreases $step1$
9   weight of each correct classifier increases $\frac{step1}{2}$
10   **return** W

In this algorithm, it is important to determine the values of two variables *step1* and *step2* as well as how the correct classifier(s) and wrong classifier(s) change weight. By saying correct/wrong classifier, we mean the classifier which makes correct/wrong decision. Improper values of *step1* and *step2* or ways of changing weights might lead to final weights that do not make sense after adjustment, for example, negative weights, or weights that are larger than 1. We tried various sets of values and ways to adjust weights. Although the solution was not unique, it turned out that the current way, as listed in lines 5,6,8,9 of the algorithm, *step1* = 0.01, and *step2* = 0.02 can provide the best classification performance. In our experiment, initially, the weight of each classifier was 0.3333. After weight adjusting, the weight of SVM, ANN and NB was 0.0833, 0.5633, and 0.3533, respectively.

To combine the decisions of each classifier, we propose a new rule for voting. The idea of this strategy is in favor of positive votes (i.e. diagnosing a subject as being ill) than negative votes (i.e. diagnosing a subject as being healthy). That is to say, if the number of positive votes exceeds the number of negative votes, the decision is based on majority voting. However, if the number of positive votes is not zero but it is smaller than the number of negative votes, then the majority voting does not apply here. Instead, we compare the weight of the positive voter with the average weight of other two negative voters for the final decision. The details of the rule are in TABLE II.

TABLE II THE VOTING RULE IN OUR APPROACH

# of positive votes	# of negative votes	decision of the ensemble
3	0	Diabetes
2	1	Diabetes
1	2	Diabetes, if the weight of positive voter is larger than the average weight of two negative voters. Healthy, otherwise.
0	3	Healthy

III. RESULTS

As pointed out in Section 2, we adjusted the weight of classifiers using dataset 2. Then we tested the performance of the ensemble of weighted classifiers on dataset 2, and also on a completely new dataset - dataset 3. We compared the performance of using SVM, ANN or NB alone with the performance of the ensemble of these three classifiers using the proposed weight-adjusted voting approach. In addition, we tested the performance of the ensemble using some existing rules of voting and compared them with our approach. There are five rules of voting involved in the comparison, including:

- Rule 1: average of probabilities
- Rule 2: product of probabilities
- Rule 3: majority voting
- Rule 4: minimum probability
- Rule 5: maximum probability

TABLE III and TABLE IV present the performance of each single classifier and the ensemble of classifiers using different voting approaches on dataset 2 and dataset 3, respectively.

From the comparison results, it is easy to tell that NB performed well in terms of sensitivity and the ensemble of classifiers using our approach achieved better sensitivity than using SVM, ANN or NB alone. Also, in the five existing rules for voting, the rule 3 – majority voting provided the highest accuracy as well as sensitivity. However, the sensitivity of using majority voting was lower than that of our approach. These conclusions hold for both dataset 2 and dataset 3.

TABLE III COMPARISON OF PERFORMANCE OF CLASSIFIER(S) ON DATASET 2 (ALL VALUES ARE IN %)

classifier	sensitivity	specificity	accuracy
SVM	44.6	91.2	75.2
ANN	44.6	94.3	77.3
NB	54.2	84.3	74.0
the ensemble (using our approach)	**57.8**	**83.0**	**74.4**
the ensemble (rule 1)	45.8	91.2	75.6
the ensemble (rule 2)	44.6	91.2	75.2
the ensemble (rule 3)	**47.0**	**91.2**	**76.0**
the ensemble (rule 4)	44.6	91.2	75.2
the ensemble (rule 5)	44.6	91.2	75.2

TABLE IV COMPARISON OF PERFORMANCE OF CLASSIFIER(S) ON DATASET 3 (ALL VALUES ARE IN %)

classifier	sensitivity	specificity	accuracy
SVM	46.4	95.6	78.7
ANN	45.2	94.4	77.5
NB	57.1	87.5	77.0
the ensemble (using our approach)	**58.3**	**86.8**	**77.0**
the ensemble (rule 1)	47.6	95.0	78.7
the ensemble (rule 2)	46.4	95.6	78.7
the ensemble (rule 3)	**53.6**	**94.4**	**80.0**
the ensemble (rule 4)	46.4	95.6	78.7
the ensemble (rule 5)	46.4	95.6	78.7

IV. DISCUSSION AND CONCLUSION

In this paper, we propose a weight-adjusted voting approach to automatically diagnose diabetes based on the decisions of an ensemble of three classifiers – SVM, ANN and NB. Our approach beats each single classifier in the ensemble from the perspective of sensitivity while maintaining reasonably high specificity and accuracy. Although the rule 3 – majority voting could achieve better performance in terms of accuracy and sensitivity than other existing voting approaches, its sensitivity was actually lower than a single classifier – NB in the ensemble. This voting strategy increased the specificity at the cost of decreasing the sensitivity. On the contrary, our approach is biased towards sensitivity, based on the consideration that it is more important to be able to diagnose a diabetes patient as illness. Timely diagnosis decisions will provide alarms and enough time ahead for patients to take proper actions for their health.

To sum, our approach for diagnosis of diabetes has three advantages. First, the decision is made on an ensemble of classifiers rather than a single classifier which might dogmatize. Second, the weight of each classifier in the ensemble is not equal. The more powerful classifier with good diagnosis history gets higher weight. Third, the approach only adjusts the weight of classifier but does not go back to iteratively change the weight of instances, which makes it a simple and efficient approach while providing good performance.

Our future work will concentrate on expanding the size of ensemble and further developing the approach to make it work for an ensemble of big size. That is to say, we will include more classifiers in the ensemble for diagnosing diabetes. Also, we will involve the classifiers with improved performance in the ensemble. In addition, the approach will be applied to diagnose other diseases.

REFERENCES

[1] National Diabetes Statistics Report, 2014, [Online], Available: http://www.cdc.gov/diabetes/pubs/statsreport14/national-diabetes-report-web.pdf

[2] International Diabetes Federation, Diabetes Atlas, 6th ed. Brussels, Belgium: International Diabetes Federation, 2013.

[3] H. Temurtas, N. Yumusak, and F. Temurtas, "A comparative study on diabetes diagnosis using neural networks", Expert Systems with Applications, vol. 36, pp. 8610 – 8615, 2009.

[4] O. Karan, C. Bayraktar, H. Gumuskaya, and B. Karlik, "Diagnosing diabetes using neural networks on small mobile devices", Expert Systems with Applications, vol. 39, pp. 54 – 60, 2012.

[5] T. Jayalakshmi, and A. Santhakumaran, "A novel classification method for diagnosis of diabetes mellitus using artificial neural networks", the Proceedings of 2010 International Conference on Data Storage and Data Engineering, Bangalore, pp. 159 – 163, 2010.

[6] N. H. Barakat, A. P. Bradley, and M. N. H. Barakat, "Intelligible support vector machines for diagnosis of diabetes mellitus", IEEE Transactions on Information Technology In Biomedicine, vol. 14 (4), pp. 1114 – 1120, 2010.

[7] Y. Huang, P. McCullagh, N. Black, and R. Harper, "Feature selection and classification model construction on type 2 diabetic patients' data", Artificial Intelligence in Medicine, vol. 41, pp. 251 – 262, 2007.

[8] M. A. Chikh, M. Saidi, and N. Settouti, "Diagnosis of diabetes diseases using an artificial immue recognition systems2 (AIRS2) with fuzzy k-nearest neighbor", Journal of Medical Systems, vol. 36 (5) , 2721 – 2729, 2012.

[9] E. D. Ubeyli, "Automatic diagnosis of diabetes using adaptive neuro-fuzzy inference systems", Expert Systems, vol. 27 (4), pp. 259 – 266, 2010.

[10] D. Calisir, and E. Dogantekin, "An automatic diabetes diagnosis system based on LDA-wavelet support vector machine classifier", Expert Systems with Applications, vol. 38, pp. 8311 – 8315, 2011.

[11] M. Khashei, S. Eftekhari, J. Parvizian, "Diagnosing diabetes type 2 using a soft intelligent binary classification model", Review of Bioinformatics and Biometrics, vol. 1 (1), pp. 9 – 23, 2012.

[12] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, "Adaptive mixtures of local experts", Neural Computing, vol. 3, pp. 179 – 187, 1991.

[13] K. Chen, L. Xu, and H. Chi, "Improved learning algorithms for mixture of experts in multiclass classification", Neural Netw., vol. 12 (9), pp. 1229 – 1252, 1999.

[14] M. I. Jordan, and R. A. Jacobs, "Hierarchical mixture of experts and the EM algorithm", Neural Computing, vol. 6, pp. 2181 – 2214, 1994.

[15] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20 (3), pp. 226 – 239, 1998.

[16] H. Kim, H. Kim, H. Moon, and H. Ahn, "A weight-adjusted voting algorithm for ensembles of classifiers", Journal of the Korean Statistical Society, vol. 40 (4), pp. 437 – 449, 2011.

[17] R. E. Schapire, "The strength of weak learnability", Machine Learning, vol. 5, pp. 197 – 227, 1990.

[18] L. Breiman, "Bagging predictors", Machine Learning, vol. 24, pp. 123 – 140, 1996.

[19] E. D. Ubeyli, "Modified mixture of experts for diabetes diagnosis", Journal of Medical System, vol. 33, pp. 299 – 305, 2009.

[20] Pima Indians Diabetes Data Set, [Online], Available: http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes

[21] K. Bache and M. Lichman, UCI Machine Learning Repository, [Online], Available: http://archive.ics.uci.edu/ml

[22] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Written, "The WEKA Data Mining Software: An Update", SIGKDD Explorations, vol.11 (1), pp. 10 – 18, 2009.