

Identification of Diabetes Disease Using Committees of Neural Network-Based Classifiers

Ali Hassan El-Baz, Aboul Ella Hassanien and Gerald Schaefer

Abstract Diabetes mellitus is one of the most serious health challenges in both developing and developed countries. In this paper, we present a design of a classifier committee for the detection of diabetes disease based on the Pima Indian diabetic database from the UCI machine learning repository. The proposed method uses multi-layer perceptron (MLP) and cascade-forward back propagation network (CFBN) predictors as base classifiers. The combined committee is based on varying the parameters related to both the design and the training of the neural network classifiers. Our experimental evaluation confirms that the derived approach provides a robust classification system, and yields classification accuracies of 95.31 and 96.88 % based on using combined MLP and combined CFBN classifiers respectively. The experimental results obtained thus show that the proposed classifier committee can form as useful basis for automatic diagnosis of diabetes.

1 Introduction

Diabetes mellitus is one of the most serious health challenges in both developing and developed countries. According to the International Diabetes Federation (IDF), 382 million people suffered from diabetes in 2013 while this is expected to rise to 592 million by 2035. Diabetes mellitus is a complex metabolic disorder which is characterised by persistent hyperglycemia and results from defects in insulin secretion, insulin action or both [1]. Diabetes can be either Type-1 (insulin dependent) or Type-2 (non-insulin dependent). The dataset we are focussing on in this paper originates from a population of Pima Indians and comprises Type-2 positive and negative

A. Hassan El-Baz (✉)
Department of Mathematics, Damietta University, Damietta, Egypt
e-mail: ali_elbaz@yahoo.com

A. Ella Hassanien
Faculty of Computers and Information, Cairo University, Giza, Egypt

G. Schaefer
Department of Computer Science, Loughborough University, Loughborough, UK
e-mail: gerald.schaefer@ieee.org

instances. According to [2], the Pima Indians of Arizona have the highest reported incidence of diabetes in the world, while the predominant Type-2 diabetes observed there is typically characterised by slow and gradual commencement. Consequently, the traditional diagnostic method which is partially based on the plasma glucose test may be delayed by up to ten years [1].

Biomedical signal classification can be defined as a categorisation of the recorded input data, which is often based on extracting significant features into distinct classes [3]. The objective is to group the observed signals, thus generating decision boundaries to separate the involved categories. Typical examples include diagnostic checks so as to identify whether a patient suffers from a certain disease such as diabetes or not.

While the traditional approach used in designing pattern classification systems is based on experimental comparison of several classification algorithms in order to select the best one [4], more recently combinations of multiple classifier decisions are investigated [4–12]. A combined classifier can perform better than any of its base constituent classifiers due to the complementary performance of the base classifiers which leads to a more stable combination. A diverse set of base classifiers can be generated by parameter variation of the base classifier and/or variations of the training sets as in bagging and boosting approaches [13–15]. Ensembles of neural network classifiers have been introduced in [16–18]. The base classifiers are different from each other in one or more aspect such as the training parameters, initial weights, network architecture or training algorithm.

In this paper, we present an effective approach to diabetes classification based on a committee of neural network classifiers. We enforce diversity in the ensemble by varying the architectures of the neural networks, with the involved parameters chosen adaptively to generate a pool of classifiers. The final decisions are combined using a majority voting rule, and the resulting ensemble is shown to give excellent classification performance on the Pima Indian diabetic database from the UCI machine learning repository.

2 Diabetes Dataset

The dataset used for our study is the Pima Indians Diabetes Database (PIDD) from the UCI machine learning repository.¹ The dataset is a collection of medical diagnostic reports from 768 records of female patients of Pima Indian heritage in Arizona. The following attributes form each record: number of times pregnant, plasma glucose concentration, diastolic blood pressure, triceps skin fold thickness, 2-hour serum insulin, body mass index, diabetes pedigree function, and age. Each subject was tested as either positive or negative with 500 non-diabetic patients and 268 diabetic ones.

¹Available at <http://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes>.

3 Proposed Classifier Committee

A variety of classifiers has been used for diagnosis of diabetes including nearest neighbour, multi-layer perceptron, decision tree, Bayes, and fuzzy classification algorithms to name a few. However, rather than relying on a single predictor, combining multiple classifiers can lead to improved and more robust classification performance. In this paper, we design such a combined classifier, or classifier ensemble, which makes use of parameter based diversity enforcement, for the purpose of diabetes diagnosis.

3.1 *Multi-layer Perceptron (MLP) Neural Network*

Multi-layer neural networks (MLNNs) have been successfully used in replacing conventional pattern recognition methods for a variety of applications including disease diagnosis [19, 20]. In the first stage of this study, a multi-layer neural network structure was used for diabetes identification on the Pima Indian dataset. The network structure consists of input layer, hidden layers, and output layer, where the hidden layer output layer neurons use non-linear sigmoid activation functions. There are eight inputs features, while the two outputs relate to the two classes. We used the Levenberg-Marquardt algorithm [21] for training the neural networks.

3.2 *Cascade-Forward Back Propagation Network (CFBN)*

A cascade type of feed-forward neural networks consists of a layer of input, a layer of output neurons, and one or more hidden layers. Similar to a general feed-forward networks, the first layer has weights coming from the input. However, each subsequent layer has weights coming from the input and all previous layers, while all layers have biases. The last layer is the network output. It was found [22] that CFBN can provide a more effective algorithm compared to feed-forward back propagation methods.

3.3 *Proposed Classifier Ensemble*

Combining multiple classifiers has been intensively studied and is widely regarded as an effective technique to overcome the limitations of individual classifiers [31, 32] with many well-known models having been proposed [33, 34]. Importantly, these classifier ensemble methods, such as neural network ensembles, have been applied successfully in many real-world applications [24, 25].

In general, there are two main categories of classifier ensembles. The first one aims at learning multiple classifiers at the feature level, and usually trains and combines multiple classifiers in the learning process, such as boosting [26] and bagging [14]. The second group tries to combine classifiers at the output level, where the results of multiple available classifiers are combined, e.g., multiple classifier systems or mixtures of experts [27]. In this paper, we focus on the second approach. In particular, given multiple classifiers (available or sequently learned, homogeneous or heterogeneous), the classifier ensemble is learned by combining these component classifiers in an intelligent way.

Classifiers differing in feature representation, architecture, learning algorithm or training data exhibit complementary classification behaviour and the fusion of their decisions can yield higher performance than the best individual classifier. It had been observed that although one classifier would yield the best performance, the sets of patterns misclassified by different classifiers would not necessarily overlap. This suggests that different classifiers potentially offer complementary information about the patterns to be classified, which can be exploited in order to improve the overall classification performance. By combining classifiers, the idea is not to rely on a single predictor but instead combine the individual opinions of multiple classifiers to arrive at a consensus decision.

In [28], three approaches to building multiple classifier systems were presented: divide and conquer approach, sequential approach and parallel approach. In the divide and conquer approach, the types of input are isolated on which each specific classifier performs well, and a new input is directed accordingly. In the sequential approach, one classifier is used at the beginning and other classifiers are invoked only if the first classifier fails, while in the parallel approach, all individual classifiers are applied in parallel to the same input data, and a combination method is applied to aggregate their decisions. It is this last technique that we employ in this paper.

In particular, we train—randomly—many individual neural network classifiers. Each classifier is different from the others in terms of number of epochs, learning rate, and/or number of hidden neurons of the network. At the end, we combine the classifier results by using the majority voting technique (also known as plurality voting technique). Both multi-layer neural network and cascade-forward back propagation network are used as base classifier in our proposed ensemble systems.

3.4 *Combination Rules*

Fixed combiners are widely studied in the literature on combining classifiers [29, 30]. The confidence $q_j(x)$ for class j is computed by

$$q_j(x) = \frac{h_j(x)}{\sum_i h_j(x)} \quad (1)$$

with

$$h_j(x) = rule_i(p_{ij}(x)). \quad (2)$$

The final classification can be made by

$$w(x) = \arg_j \max(q_j(x)). \quad (3)$$

This Maximum rule selects the classifier producing the highest estimated confidence.

In contrast, in ensemble majority voting, which is a popular way of combining classifiers, the votes for each class over the input classifiers are counted and the majority class selected. In terms of the rule of Eq. 2 this equates to

$$h_j(x) = \sum_i I(\arg_i \max(q_j(x)) = i), \quad (4)$$

where $I()$ is the indicator function defined as

$$I(y) = \begin{cases} 1, & \text{if } y = \text{true}; \\ 0, & \text{otherwise} \end{cases}. \quad (5)$$

4 Experimental Results

4.1 MLP Ensemble

In the first experiment, MLP is used as base classifier for our proposed ensemble. In order to generate diversity in the classifier ensemble, a pool of 16 different classifiers are generated and the final hypothesis is combined using majority voting. The classification accuracy of the best individual classifier on the test data is 80.21% while the proposed combined classifier achieved a classification accuracy of 95.31%. We note that the proposed combined classifier compares favourably with results from the literature. Figure 1 shows the accuracy of base MLP versus the number of individual classifiers, while Figs. 2 and 3 show the accuracy of base MLP versus the number of hidden neurons and number of training epochs, respectively.

4.2 CFBN Ensemble

In the second experiment, a cascade-forward back propagation network (CFBN) is used as the base classifier for the combined classifier. The best individual classifier gave a classification accuracy of 81.77% on the test data. In contrast, the proposed combined classifier achieved a performance of 96.88%. The CFBN approach thus gave slightly better compared to the MLP, both for the best individual classifier and

Fig. 1 Accuracy of base MLP versus the number of individual classifiers

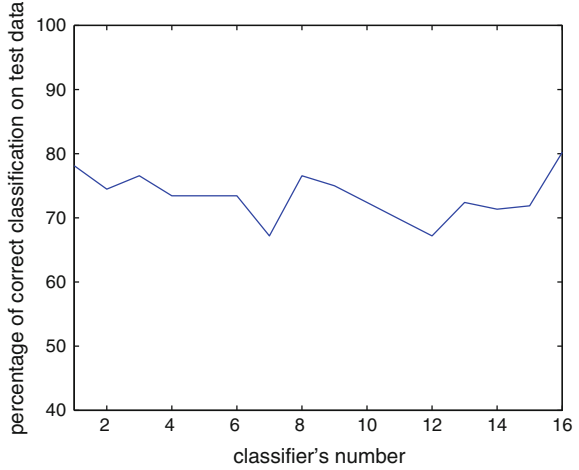
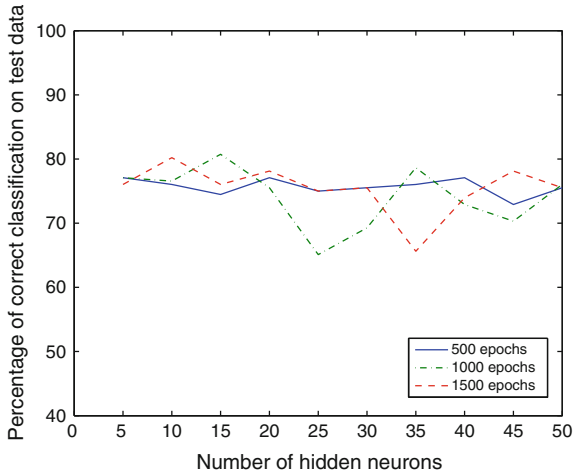


Fig. 2 Accuracy of base MLP versus the number of hidden neurons



for the generated ensemble. Figure 4 shows the accuracy of base CFBN versus the number of individual classifier, while Figs. 5 and 6 show the accuracy of base CFBN versus the number of hidden neurons and number of training epochs, respectively.

5 Performance Measurements

There are a number of metrics to measure the performance of classification methods which are commonly used in medical diagnosis tasks, which are based on the definitions of true positives (TP, the number of correct predictions in which an instance is positive), true negatives (TN, the number of incorrect predictions in which an in-

Fig. 3 Accuracy of base MLP versus the number of training epochs

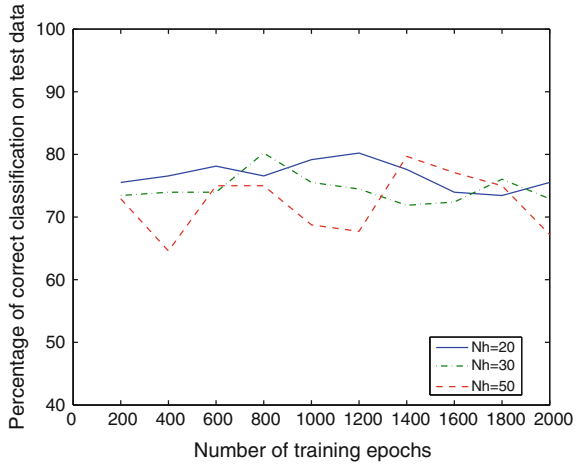
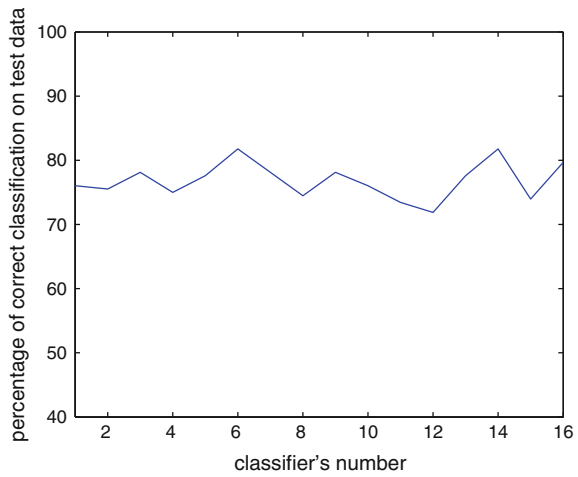


Fig. 4 Accuracy of base CFBN versus the number of individual classifiers



stance is negative), false positives (FP, the number of incorrect predictions in which an instance is positive), and false negatives (FN, the number of correct predictions in which an instance is negative).

Recall (sensitivity) is the percentage of real positive cases that are correctly predicted positive. Conversely, precision (confidence) indicates the percentage of predicted positive cases that are correctly real positives. The F-measure combines precision and recall as their harmonic mean. Specificity indicates the percentage of samples that were classified as normal and which were labeled as normal. In Fig. 7, we summarise the experimental results, in terms of these measures. As we can see, the performance of combining CFBNs outperforms an ensemble of MLPs.

Fig. 5 Accuracy of base CFBN versus the number of hidden neurons

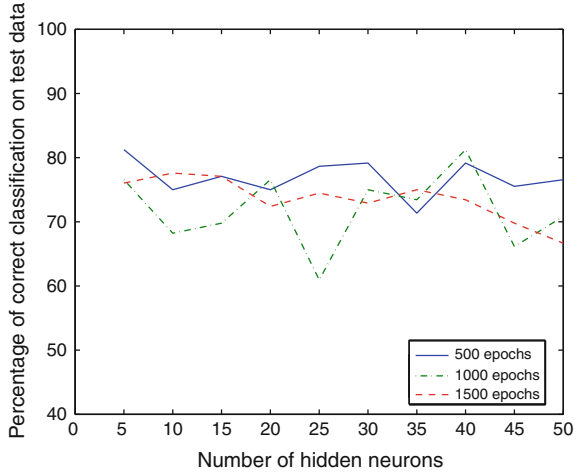


Fig. 6 Accuracy of base CFBN versus the number of training epochs

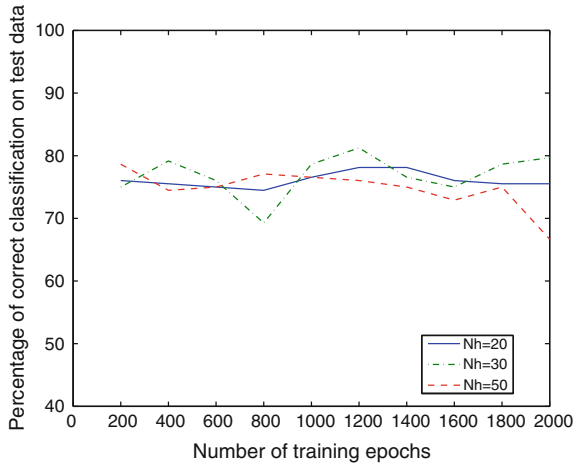
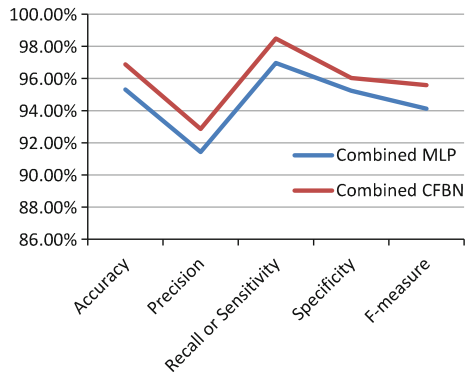


Fig. 7 Performance measures of the proposed methods on test data



6 Conclusions

Automatic identification of diabetes is a challenging task in pattern recognition. Combinations of weak classifiers have shown the potential to achieve good generalisation performance. In this paper, we have introduced a classifier ensemble to solve the instability problem of individual pattern classifiers in the context of diabetes diagnosis. The proposed classifier consists of multiple individual classifiers which are different in both architecture and training parameters. As confirmed in the experimental evaluation, the achieved classification accuracies of the combined classifiers clearly outperform those of the best individual classifiers. The performance is further found to outperform that of the reported methods in the literature. Of the two investigated neural network models that were employed as base classifiers, cascade-forward back propagation networks were found to outperform multi-layer perceptron networks, both as individual classifiers and in terms of the resulting ensembles.

References

1. Holt, R., Hanley, N.: *Essential Endocrinology and Diabetes*. Blackwell Publishing, Malden (2006)
2. Knowler, W.C., Bennett, P.H., Bottazzo, G.F., Doniach, D.: Islet cell antibodies and diabetes mellitus in Pima Indians. *Diabetologia* **17**(3), 161–174 (1979). Springer, Heidelberg, Berlin
3. Rangayyan, R.M.: *Biomedical Signal Analysis, A Case-study Approach*. IEEE Press series, Canada (2002)
4. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **3**, 226–239 (1998)
5. Rogova, G.: Combining the results of several neural network classifiers. *Neural Netw.* **7**(5), 777–781 (1994)
6. Zhou, Z.H., Wu, J., Tang, W.: Ensembling neural networks: Many could be better than all. *Artif. Intel. J.* **137**(1–2), 239–263 (2002)
7. Kuncheva, L.I.: *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, New York (2004)
8. Garca-Osorio, Csar, de Haro-Garca, Aida, Garcia-Pedrajas, Nicols: Democratic instance selection: A linear complexity instanceselection algorithm based on classifier ensemble concepts. *Artif. Intell. J.* **174**(5–6), 410–441 (2010)
9. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F.: An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. *Pattern Recogn.* **44**(8), 1761–1776 (2011)
10. Tabassian, M., Ghaderi, R., Ebrahimpour, R.: Combination of multiple diverse classifiers using belief functions for handling data with imperfect labels. *Expert Syst. Appl.* **39**(2), 1698–1707 (2012)
11. Boln-Canedo, V., Snchez-Maroo, N., Alonso-Betanzos, A.: An ensemble of filters and classifiers for microarray data classification. *Pattern Recogn.* **45**(1), 531–539 (2012)
12. Podolak, I.T., Roman, A.: Cores: Fusion of supervised and unsupervised training methods for a multi-class classification problem. *Pattern Anal. Appl.* 1–19 (2011)
13. Diettrich, T.G.: Ensemble methods in machine learning. In: *Proceedings of the first international workshop on multiple classifier systems*. Lecture Notes in Computer Science, pp. 1–15. June, Springer, Berlin (2000) (Cagliari, Italy)
14. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1997)

15. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**(1), 119–139 (1997)
16. Giacinto, G., Roli, F.: Design of effective neural network ensembles for image classification. *J. Image Vis. Comput.* **19**(9–10), 697–705 (2001)
17. Sharkey, A.J.C.: On combining artificial neural nets. *Connection Sci.* **8**, 299–314 (1996)
18. Sharkey, A.J.C., Sharkey, N.E., Gerecke, U., Chandroth, G.O.: The test and select approach to ensemble combination. In: Kittler, J., Roli, F. (Eds.) *Proceedings of the first International Workshop on Multiple Classifier Systems (MCS2000)*. LNCS 1857, pp 30–44. Springer, New York (2000)
19. Delen, D., Walker, G., Kadam, A.: Predicting breast cancer survivability: A comparison of three data mining methods. *Artif. Intell. Med.* **34**(2), 113–127 (2005)
20. Polat, K., Gunes, S.: An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease. *Digital Signal Process.* **17**(4), 702–710 (2007)
21. Hagan, M.T., Menhaj, M.: Training feed forward networks with the Marquardt algorithm. *IEEE Trans. Neural Netw.* **5**, 989–993 (1994)
22. Filik, U.B., Kurban, M.: A new approach for the short-term load forecasting with autoregressive and artificial neural network models. *Int. J. Comput. Intell. Res.* **3**(1), 66–71 (2007)
23. Sokolova, M., Japkowicz, N., Szpakowicz, S.: Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. *AI 2006: Adv. Artif. Intell.* 1015–1021 (2006)
24. Dasarathy, B.V.: A special issue on applications of ensemble methods. *Inf. Fus.* **9**(1)(2008)
25. Oza, C., Tumer, K.: Applications of ensemble methods. *Inf. Fus.* **9**(1), 2–3 (2008)
26. Schapire, R.: The strength of weak learnability. *Mach. Learn.* **5**(2), 197–227 (1990)
27. Yuksel, S.E., Wilson, J.N., Gader, P.D.: Twenty years of mixture of experts. *IEEE Trans. Neural Netw. Learn. Syst.* **23**(8), 1177–1193 (2012)
28. Ho, T.K.: Data complexity analysis for classifier combination. In: *Proceedings of the 2nd International Workshop of Multiple Classifier System 2006*. Lecture Notes in Computer Science, pp. 53–67. Springer, Cambridge (2001)
29. Kittler, J., Hatef, M., Duin, R.P.W., Matas, J.: On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(3), 226–239 (1998)
30. Xu, L., Krzyzak, A., Suen, C.Y.: Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans. Systems, Man, Cybern.* **22**(3), 418–435 (1992)
31. Ho, T., Hull, J., Srihari, S.: Decision combination in multiple classifier systems. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**(1), 66–75 (1994)
32. Dietterich, T.: Ensemble method in machine learning. In: *Proceedings of International Workshop on Multiple Classifier Systems*, pp. 1–15 (2000)
33. Krogh, A., Vedelsby, J.: Neural network ensembles, cross validation, and active learning. In: *Advances in Neural Information Processing Systems*, vol. 7, pp. 231–238 (1995)
34. Hansen, L.K., Salamon, P.: Neural network ensembles. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(10), 993–1001 (1990)