

A Prediction Model to Diabetes Using Artificial Metaplasticity

Alexis Marcano-Cedeño, Joaquín Torres, and Diego Andina

Group for Automation in Signals and Communications
Technical University of Madrid

Abstract. Diabetes is the most common disease nowadays in all populations and in all age groups. Different techniques of artificial intelligence has been applied to diabetes problem. This research proposed the artificial metaplasticity on multilayer perceptron (AMMLP) as prediction model for prediction of diabetes. The Pima Indians diabetes was used to test the proposed model AMMLP. The results obtained by AMMLP were compared with other algorithms, recently proposed by other researchers, that were applied to the same database. The best result obtained so far with the AMMLP algorithm is 89.93%.

1 Introduction

Diabetes is a major health problem in both industrial and developing countries, and its incidence is rising. It is a disease in which the body does not produce or properly use insulin, the hormone that “unlocks” the cells of the body, allowing glucose to enter and fuel them [1]. Diabetes increases the risks of developing kidney disease, blindness, nerve damage, blood vessel damage and it contributes to heart disease. The World Health Organization in 2000 indicated there were ~ 170 million people with diabetes, and estimated that the number of cases of the disease worldwide will be more than doubled to 366 million by 2030 [2]. Diabetes occurs in two major forms: type 1, or insulindependent diabetes, and type 2, or non-insulin-dependent diabetes. The type 1 diabetes, is characterized by an absolute deficiency of insulin secretion. Individuals at increased risk of developing this type of diabetes can often be identified by serological evidence of an autoimmune pathologic process occurring in the pancreatic islets and by genetic markers [3, 4]. The most common form of diabetes is Type 2 [3, 4]. It is believed that some of the causes of type II diabetes are associated with: diet changes, aging, urbanization, and increasing prevalence of obesity and physical inactivity [3, 4]. This type diabetes results from insulin resistance (a condition in which the body fails to properly use insulin), combined with relative insulin deficiency. In Type II diabetes, either the body does not produce enough insulin or the cells ignore the insulin [3, 4]. Although detection of diabetes is improving, about half of the patients with Type II diabetes are undiagnosed and the delay from disease onset to diagnosis may exceed 10 years. Thus, earlier detection of Type II diabetes and treatment of hyperglycaemia and related to metabolic abnormalities is of vital importance [3, 4, 5].

There are many factors to analyze to diagnose the diabetes of a patient, and this makes the physician's job difficult. There is no doubt that evaluation of data taken from patient and decisions of experts are the most important factors in diagnosis. But, this is not easy considering the number of factors that has to evaluate [6]. To help the experts and helping possible errors that can be done because of fatigued or inexperienced expert to be minimized, classification systems provide medical data to be examined in shorter time and more detailed. Expert systems and different artificial intelligence techniques for classification systems in medical diagnosis is increasing gradually. As for other clinical diagnosis problems, classification systems have been used for diabetes diagnosis problem [7].

The main objective of the present research was to apply the artificial metaplasticity on multilayer perceptron (AMMLP) as prediction model for prediction of diabetes. The Pima Indian diabetes data set [8] was used to test the proposed model AMMLP. The results obtained by AMMLPT were compared with other algorithms, recently proposed by other researchers, that were applied to the same database.

The remainder of this paper is organized as follows. Section 2 presents a brief data mining, artificial metaplasticity. We present the data base and artificial metaplasticity. The Section 4 gives the experimental results obtained by proposed model. Section 5 compares our results to other methods from the literature. Finally, Section 6 presents the summarized conclusions.

2 Methods and Materials

2.1 Pima Indian Diabetes Data Set

The reason for using this data set is because it is very commonly used among the other classification systems that we have used to compare this study for Pima Indian diabetes diagnosis problem [8]. The data set which consists of Pima Indian diabetes disease measurements contains two classes and 768 samples. The class distribution is:

- Class 1: normal (500) (65.10%)
- Class 2: Pima Indian diabetes (268) (39.89%)

All samples have eight features. These features are:

- Feature 1: Number of times pregnant (Pregn).
- Feature 2: Plasma glucose concentration a 2 h in an oral glucose tolerance test (Gluko).
- Feature 3: Diastolic blood pressure (mm Hg) (Diasp).
- Feature 4: Triceps skin fold thickness (mm) (Thick).
- Feature 5: 2-h serum insulin (IU/ml)(Insul).
- Feature 6: Body mass index (weight in $kg/(height\ in\ m)^2$) (Massi).
- Feature 7: Diabetes pedigree function (Predf).
- Feature 9: Age (years) (Age).
- Class: Diabetes onset within five years (0 or 1)(Yes or No)

2.2 Data Preparation

The quality of the data is the most important aspect as it influences the quality of the results from the analysis. The data should be carefully collected, integrated, characterized, and prepared for analysis. In this study, we applied the techniques of data preprocessing in order to improve the quality of the mining result and the efficiency of the mining process. In this study we analyze a data set composed of 768 data instances. A preliminary analysis of the data indicates the usage of zero for missing data. Since, it does not make sense to have the value of a variable such as plasma-glucose concentration 0 in living people; all the observations with zero entries are removed. After removing all the above said values and variables, only 763 instances remain from the data in our study.

2.3 Artificial Metaplasticity Neural Network

The concept of biological metaplasticity was defined in 1996 by Abraham [9] and now widely applied in the fields of biology, neuroscience, physiology, neurology and others [9, 10].

Recently, Ropero-Peláez [10], Andina [11] and Marcano-Cedeño [12] have introduced and modeled the biological property metaplasticity in the field of artificial neural networks (ANN), obtaining excellent results.

There are different artificial metaplasticity (AMP) models [12]. However, of all AMP models tested by the authors, the most efficient model (as a function of learning time and performance) is the approach that connects metaplasticity and Shannon's information theory, which establishes that less frequent patterns carry more information than frequent patterns [13]. This model defines artificial metaplasticity as a learning procedure that produces greater modifications in the synaptic weights with less frequent patterns than frequent patterns, as a way of extracting more information from the former than from the latter. Biological metaplasticity favors synaptic strengthening for low-level synaptic activity, while the opposite occurs for high level activity. The model is applicable to general ANNs [11, 12], although in this paper it has been implemented for a multilayer perceptron (MLP).

MLP has been used for the solution of many classification problems in pattern recognition applications [14]. The functionality of the topology of the MLP is in most cases determined by a learning algorithm, the Backpropagation (BP), based on the method of steepest descent. In the process of upgrading the connection weights, is the most commonly used algorithm by the ANN scientific community. The BP algorithm presents some limitations and problems during the MLP training [15]. Many researchers have centered their work in improving and developing algorithms for reduce those problems and increasing their advantages [15, 11, 12].

The artificial metaplasticity on multilayer perceptron algorithm (AMMLP) is included in the training algorithm by affecting the weights in each iteration step using a weight function that assumes an estimation or an hypothesis of the

real distribution of training patterns. In this paper, has been used the following function to weight the weights updates in the learning phase [11, 12]:

$$f_X^*(x) = \frac{A}{\sqrt{(2\pi)^N} \cdot e^{-B \sum_{i=1}^N x_i^2}} \quad (1)$$

where N is the number of neurons in the MLP input layer, and parameters A and $B \in R^+$ are algorithm optimization values which depend on the specific application of the AMMLP algorithm. Values for A and B have been empirically determined. Eq. (1) is a gaussian distribution, so we have assumed that X pdf is Gaussian (if it is not the case, we should use the real X pdf instead) [11, 12]. Then, $f_X^*(x)$ has high values for un-frequent x values and close to 1 for the frequent ones and can therefore be straightforwardly applied in weights updating procedure to model the biological metaplasticity during learning [12].

2.4 The AMMLP Algorithm

MLP exhibits a sigmoidal activation function with scalar output usually in the range (0,1). This property is true for all neurons of the network. All AMMLP were trained with the same training data set and tested with the same evaluation data set. The network was trained with 60% of the data (455 samples), of which 158 were diabetes and 297 were not diabetes. The testing set, composed of the remaining 40% of the data, consisted of 108 diabetes sample and 200 not diabetes sample.

100 AMMLPs with different initial weights, sampled from random values of a normal distribution (mean of 0 and a variance of 1) have been generated. In each experiment, 100 networks were trained to achieve an average result that is independent of the initial random value of the ANN values. Two different criteria were applied to stop training: a) training was stopped when the error reached 0.01 (error decreases but cannot reach to 0); b) training was performed with a fixed number of 2.000 epochs.

The AMMLP algorithm was developed in MATLAB (MATLAB version 7.6.0.324, R2008a) and on a 3.4 GHz Pentium IV computer with 2 GB of RAM.

2.5 Schema Using

The proposed model AMMLP follow the iterative phases that show Fig. 1.

1. In this step we take the data provided by the medical institute. The possible factors which affect Type-2 diabetes
2. During data preprocessing step, some inappropriate and inconsistent data are deleted.
3. The AMMLP is used to predict accuracy
4. The performance evaluation is measured using the Accuracy, Specificity, and Sensitivity.
5. Compared with existing models and algorithms

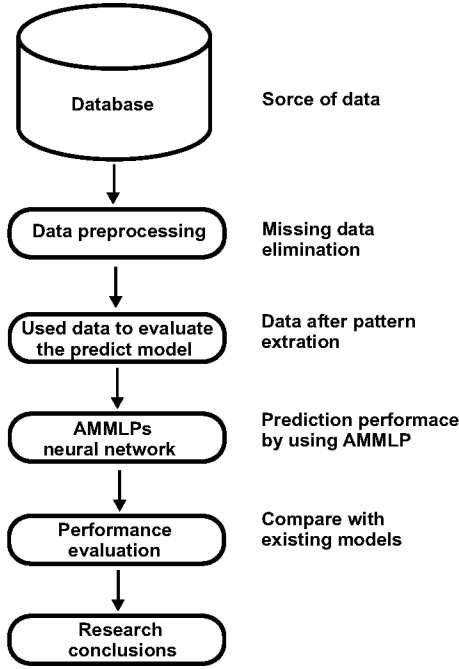


Fig. 1. Scheme using in this research

3 Results

This section present the results obtained by prediction model AMMLP.

To measure the performance of the proposed model AMMLP for the prediction of diabetic, the evaluation was performed as follows:

We determine the accuracy of prediction model using different structures of AMMLP. To evaluate this part, three criteria were applied: analysis of sensitivity, specificity and total classification accuracy prediction, confusion matrix for the best result obtained by AMMLP. These criteria are defined as:

- *Specificity*: number of correctly classified diabetes / number of total diabetes.
- *Sensitivity*: number of correctly classified not diabetes records /number of total not diabetes.
- *Total classification accuracy prediction*: number of correctly classified records/ number of total records.
- *Confusion matrix*: a confusion matrix contains information about actual and predicted classifications performed by a classifier. Performance of the classifier is commonly evaluated using the data in the matrix.

After testing different learning rate, hidden layer nodes and metaplasticity parameters, we find best networks structure and metaplasticity parameters for this case (see Table 1).

Table 1. AMMLP results obtained for different network structures and parameters of metaplasticity algorithm

Network Structure			Metaplasticity Parameters		Mean Squared Error	Clustering Accuracy (%)	
I	HL	O	A	B		Training	Testing
8	4	1	37	0.2	0.01	82.14 %	89.93 %
8	7	1	38	0.3	0.01	69.01 %	87.66 %
8	8	1	39	0.5	0.01	74.06 %	81.49 %

Table 2 and 3 show the confusion matrix and prediction accuracy obtained by AMMLP in the diabetes prediction in terms of specificity, sensitivity and total accuracy prediction.

Table 2. Confusion matrix obtained by AMMLP in the diabetes prediction

Classifier	Desired Result	Output Results	
		Diabetes	No diabetes
AMMLPs	Diabetes	82	26
	No diabetes	5	195

Table 3. Prediction accuracies obtained by AMMLP in the diabetes prediction

Classifier	Prediction Accuracies (%)		
	Specificity	Sensitivity	Total prediction accuracy
AMMLPs	75.92%	97.50%	89.93%

4 Comparison

The prediction accuracy obtained by AMMLP was compared with the best results obtained by other researchers using the same database, are summarized in Table 4.

5 Conclusion

In this paper, was applied the artificial metaplasticity on multilayer perceptron (AMMLP) as prediction model for prediction of diabetes. The proposed model AMMLP obtained an excellent accuracy in the prediction of 89.93%, the highest obtained until now. We think that the proposed model AMMLP-DT can serve as a second opinion for physicians when making their final diagnostic decisions.

Table 4. Classification accuracies of proposed Model AMMLP and other classifiers for the Pima Indians diabetes

Author (year)	Method	Accuracy (%)
Carpenter and Markuzon, (1998)	ARTMAP-IC	81.0
Deng and Kasabov, 2001	ESOM	78.4 ±1.6
Kayaer and Yildirim, 2003	GRNN	80.21
Abdel-Aal, (2005)	T-MC	77.6
Luukka and Leppälampi (2006)	PCA-Entropy	80.47
Polat and Gunes, (2007)	PCA-ANFIS	89.47
Srinivasa,Venugopal and Patnaik, (2007)	SAMGA	74.6
Ji and Carin, (2007)	POMDP	71.43
Polat, Gunes, and Aslan (2008)	LS-SVM	82.05
Kahramanli and Allahverdi (2008)	FNN	84.2
Ghazavi and Liao, (2008)	FUZZY MODELS	77.65
Termutas <i>et al.</i> (2009)	MLNN-LM	82.37
Lekkas and Mikhailov (2010)	eClass	79.37
Aibinu <i>et al.</i> , (2010)	CVNN-GDA	81.00
Dogantekin <i>et al.</i> , (2010)	LDA-ANFIS	84.61
Aibinu, Salami and Shafie (2011)	CAR	81.28
Castro, Nebot and Mugica (2011)	LR-FIR	75.39
In this study, 2011	AMMLP	89.93

References

1. Mohamed, E.I., Linderm, R., Perriello, G., Di Daniele, N., Poppl, S.J., De Lorenzo, A.: Predicting type 2 diabetes using an electronic nose-base artificial neural network analysis. *Diabetes Nutrition & Metabolism* 15(4), 215–221 (2002)
2. Shaw, J.E., Sicree, R.A., Zimmet, P.Z.: Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice* 87, 4–14 (2010), doi:10.1016/j.diabres.2009.10.007
3. American diabetes association, <http://www.diabetes.org/diabetes-basics/>
4. International Diabetes Federation, <http://www.idf.org>
5. Temurtas, H., Yumusak, N., Temurtas, F.: A comparative study on diabetes disease diagnosis using neural networks. *Expert Systems with Applications* 36(4), 8610–8615 (2009), doi:10.1016/j.eswa.2008.10.032
6. Polat, K., Gunes, S., Aslan, A.: A cascade learning system for classification of diabetes disease: Generalized discriminant analysis and least square support vector machine. *Expert Systems with Applications* 34(1), 214–221 (2008), doi:10.1016/j.eswa.2006.09.012
7. 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 Processing* 17(4), 702–710 (2007)
8. UCI machine learning respiratory, <http://archive.ics.uci.edu/ml/datasets.html>.
9. Abraham, W.C.: Activity-dependent regulation of synaptic plasticity(metaplasticity) in the hippocampus. In: Kato, N. (ed.) *The Hippocampus: Functions and Clinical Relevance*, pp. 15–26. Elsevier Science, Amsterdam (1996)
10. Kinto, E., Del-Moral-Hernandez, E., Marcano-Cedeño, A., Ropero-Peláez, J.: A preliminary neural model for movement direction recognition based on biologically plausible plasticity rules. In: Mira, J., Álvarez, J.R. (eds.) *IWINAC 2007*. LNCS, vol. 4528, pp. 628–636. Springer, Heidelberg (2007), doi:10.1007/978-3-540-73055-2_65

11. Andina, D., Alvarez-Vellisco, A., Jevtić, A., Fombellida, J.: Artificial metaplasticity can improve artificial neural network learning. *Intelligent Automation and Soft Computing; Special Issue in Signal Processing and Soft Computing*, 15(4), 681-694 (2009); ISSN: 1079-8587
12. Marcano-Cedeño, A., Quintanilla-Domínguez, J., Andina, D.: Breast cancer classification applying artificial metaplasticity algorithm. *Neurocomputing*, doi:10.1016/j.neucom.2010.07.019
13. Shannon, C.E.: A mathematical theory of communication. *The Bell System Technical Journal* 27, 379–423 (1948), doi:10.1145/584091.584093
14. Hagan, M.T., Demuth, H.B., Beale, M.: *Neural network design*. PWS Pub. Co., Boston (1996)
15. Leung, H., Haykin, S.: The complex backpropagation algorithm. *IEEE Transactions on Signal Processing* 39, 2101–2104 (1991)