ISBN: 978-1-4799-6526-7

Diagnosis of Diabetes Mellitus Using Extreme Learning Machine

Jefri Junifer Pangaribuan
Informatics Technology Department
Faculty of Informatics Technology
STMIK Mikroskil
Medan, Indonesia
jefrijunifer@yahoo.co.id

Suharjito
Magister of Information Technology
Binus Graduate Program
Bina Nusantara University
Jakarta, Indonesia
harjito@yahoo.com

***Abstract*—In 2010, Global Status Report on NCD World Health Organization (WHO) reported that 60 percent of deaths in the world caused by the non-communicable diseases, and one of the non-communicable diseases that consumed a lot of attention was diabetes mellitus. Diabetes is a serious threat to the health development, because diabetes is a disease that caused most other diseases (complications), such as blindness, kidney failure, heart disease, diabetic foot (gangrene) so it had to be amputated, up to the most serious is strokes. Estimated in 2030, Indonesia will have 21.3 million people with diabetes mellitus. An increasing number of diabetes caused by the late diagnosis of this disease. Therefore, we need a new forecast which could be an invaluable tool in determining whether someone has diabetes or not. So many methods used to generate accurate predictions; one of them is artificial neural network. This study will implement a new method of neural network, namely the Extreme Learning Machine (ELM). Extreme learning machine is a feed-forward artificial neural network with one or more hidden layers known as single hidden layer feed-forward neural. Based on the results of experiments conducted, it appears that Extreme Learning Machine is able to provide a good prediction accuracy results with a very good prediction rate.**

Keywords— diagnose; diabetes mellitus; computer science; artificial neural network; extreme learning machine

I. INTRODUCTION

Diabetes is the world's oldest disease. Diabetes is associated with glucose metabolism in the blood. Medically, the definition of diabetes mellitus extends to a series of symptoms that occur in a person caused by the increase in blood sugar levels (hyperglycemia) due to lack of insulin [1].

Diabetes mellitus is closely associated with regulation mechanism of normal sugar. Increased blood sugar levels will trigger the insulin hormone production by the pancreas gland. Diabetes mellitus is a disease that causes other diseases (complications). A complication could cause deadly diseases, such as heart attack and stroke. It is associated with continuously elevated blood sugar level which caused collapsed veins, nerves and other internal structures. Complex substance that consisted of sugar in the blood vessel wall caused the blood vessels to be thickened. As a result of this thickening, the blood's flow will be reduced, leads to the skin and nerves [1].

According to Prof. Dr. Sidartawan Soegondo, Indonesia became the fourth country in the world that has the highest diabetes rate and increased up to 14 million people. It is based on the report of World Health Organization (WHO), in which the number of diabetics in Indonesia in 2000 was 8.4 million people followed by India (31.7 million), China (20.8 million) and the United States (17.7 million). WHO reported that there are more than 143 million people who suffered diabetes. This number projected the prevalence that will double in the 2030 and as much as 77% of which occur in developing countries [2].

The increasing number of diabetes caused by delayed diagnosis. Patients have died from complications before the existence of the diagnosis. There are many factors that caused the delay that influence the choices available or the variety of variables. Therefore, we need a prediction that could be an invaluable tool in determining whether a person is suffering from diabetes mellitus or not.

Many methods are used to obtain accurate predicted results. One of them is Artificial Neural Network (ANN) that adopts the learning system in the human brain. ANN is applied intensively particular in sales forecasting, because of its advantages in the control area, prediction and pattern recognition. Many studies conclude that the neural network method is better than the conventional forecasting methods.

This study implemented a new method of ANN, namely Extreme Learning Machine (ELM). ELM is a feed-forward neural network with one hidden layer, or better known as the single hidden layer feed-forward neural network [3]. ELM has advantages in learning speed, and has a better accuracy rate than conventional methods such as Moving Average and Exponential Smoothing [4]. Besides being a relatively new method, the implementation of ELM on the diagnoses of diabetes mellitus has not been done.

The issue that rose in this study was how to diagnose the disease using ELM with accurate results than backpropagation. The purpose of this study is to implement the ELM to diagnose diabetes mellitus and analyze the accuracy of the results of

978-1-4799-6527-4/14/$31.00 ©2014 IEEE

ELM. We collect the data that will be trained and tested from UC Irvine Machine Learning Repository and create the diagnoses using ELM with MATLAB tools.

The next section of this paper contains a brief description of the ELM and the problems with ANN. The third section describes the methodology used in this study. The fourth contains results and discussion, and then closed by the conclusions and suggestions in the last section.

II. LITERATURE STUDY

A. Diabetes Mellitus

According to American Diabetes Association (ADA) in 2010, diabetes mellitus is a group of metabolic diseases with hyperglycemia characteristic that occurred because of abnormalities of insulin secretion, insulin action, or both [5].

A variety of complaints can be found in person with diabetes. Suspicion of diabetes should be considered if there are complaints like the following [5]:

- Classic complaints of diabetes mellitus include: polyuria (overload of urination), polydipsia (overload of thirst), and weight loss that cannot be explained.
- Other complaints include: weak body, tingling, itching, blurred eyes, erection dysfunction in men, and pruritus vulvae in women.

B. Expert System

Expert system is a computerized system which uses knowledge of a specific field to reach the solution of a problem from that field. To build a good expert system, there are some components needed, such as: user interface, knowledge base, inference machine, and working memory [6].

C. Data Mining

The term data mining refers therefore to the overall process consisting of data gathering and analysis, development of inductive learning models and adoption of practical decisions and consequent actions based on the knowledge [7].

Data mining activities can be subdivided into two major investigation streams, according to the main purpose of the analysis: interpretation and prediction. Seven basic data mining tasks can be identified: characterization and discrimination, classification, regression, time series, analysis, association rules [7].

D. Artificial Neural Network (ANN)

(ANN) has been extensively applied for pattern classification and regression problems. The major reason for the success of ANNs is their ability in obtaining a non-linear approximation model function describing the association between the dependent and independent variables using the given input samples [8].

Though ANNs have many advantages such as better approximation capabilities and simple network structures, however, it suffers from several problems such as presence of local minima's, imprecise learning rate, selection of the number of hidden neurons and over fitting.

The popular learning algorithm used in feed-forward neural networks is the backpropagation learning algorithm where gradients can be computed efficiently by propagation from the output to the input. There are several issues on backpropagation learning algorithms:

- When the learning rate η is too small, the learning algorithm converges very slowly. However, when η is too large, the algorithm becomes unstable and diverges.
- Another peculiarity of the error surface that impacts the performance of the back-propagation learning algorithm is the presence of local minima. It is undesirable that the learning algorithm stops at a local minimum if it is located far above a global minimum.
- Neural network may be over-trained by using backpropagation algorithms and obtain worse generalization performance. Thus, validation and suitable stopping methods are required in the cost function minimization procedure.
- Gradient-based learning is very time-consuming in most applications

E. Extreme Learning Machine (ELM)

ELM is feed-forward neural network with one hidden layer, or better known as single hidden layer feed-forward neural network. ELM has several interesting and significant features different from traditional popular gradient-based learning algorithms for feed-forward neural networks [9]:

- The learning speed of ELM is extremely fast. In simulations [9], the learning phase of ELM can be completed in seconds or less than seconds for many applications. Previously, it seems that a virtual speed barrier existed, which most (if not all) classic learning algorithms cannot break through and it is not unusual to take very long time to train a feed-forward network using classic learning algorithms even for simple applications.
- ELM has better generalization performance than the gradient-based learning such as backpropagation in most cases.
- The traditional classic gradient-based learning algorithms may face several issues like local minima, improper learning rate and over fitting, etc. In order to avoid these issues, some methods such as weight decay and early stopping methods may need to be used often in these classical learning algorithms. The ELM tends to reach the solutions straightforward without such trivial issues. The ELM learning algorithm looks much simpler than most learning algorithms for feed-forward neural networks.
- Unlike the traditional classic gradient-based learning algorithms which only work for differentiable activation functions, as easily observed the ELM learning algorithm could be used to train SLFNs with many non-differentiable activation functions.

Even though ELM was classified as new algorithm, its application to some predictions have been made, such as predicting material properties [10], demand forecasting [4], predict in retail industry [11], and bank client classification [12].

The results of implementation of ELM for demand forecasting shows that ELM has better accuracy than conventional methods such as Moving Average and Exponential Smoothing (ES). The error rate (Mean Square Error and Mean Absolute Error Percentage) of ELM is smaller than the two conventional methods [4].

In one of simulations [11], Grey Relation Analysis (GRA) and ELM is combined to predict retail industry. This combination called GELM. GELM performance is compared with prediction model Generalized ARCH (GARCH), Generalized BackPropagation Network (GBPN), and GRA Multilayer Functional Link Network (GMFLN). The result is shown as follow:

TABLE I. THE COMPARED RESULTS OF DIFFERENT FORECASTING MODELS

Model Type		**MAD**	**Training Time**
Model time series statistics	GARCH	0.13876	-
Model ANN	GBPN	0.09837	11.573
	GMFLN	0.08911	4.216
	GELM	**0.07039**	**0.3750**

As a learning technique, ELM has demonstrated good potentials to resolving regression and classification problems. Until recently, ELM techniques have received considerable attention in computational intelligence and machine learning communities, in both theoretic study and applications [13].

III. METHODOLOGY

A. Research Step

The first step of this study started by determining the research background, research purpose and defines the scope of research. Literature study conducted to deepen the understanding of ELM method and what steps are needed for the diagnosis of diabetes using ELM.

The second step of this study is the data collection. The third step is the implementation of ELM method through distribution of the data sets, training and testing ELM. The fourth step is analyzing the results of ELM prediction, and then the last step is to draw conclusions. The following is the outline of this study:

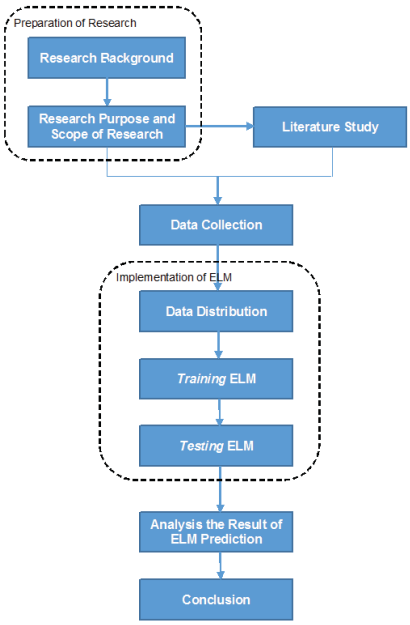

Fig. 1. Research Steps

B. Data Collection

Data used in this study are classification data taken from UCI Repository. Data are provided by Vincent Sigillito from The Johns Hopkins University and given to the National Institute of Diabetes and Kidney Diseases on May 9, 1990.

Data set consisted from women population aged 21 years and lived in Phoenix, Arizona. Data are taken according to the criteria of the World Health Organization (WHO), which 768 people will be taken to be the data set and will be divided into training data and testing data.

C. Prediction Using ELM

This study use MATLAB as diagnosis tool. The network model of ELM as shown below [14]:

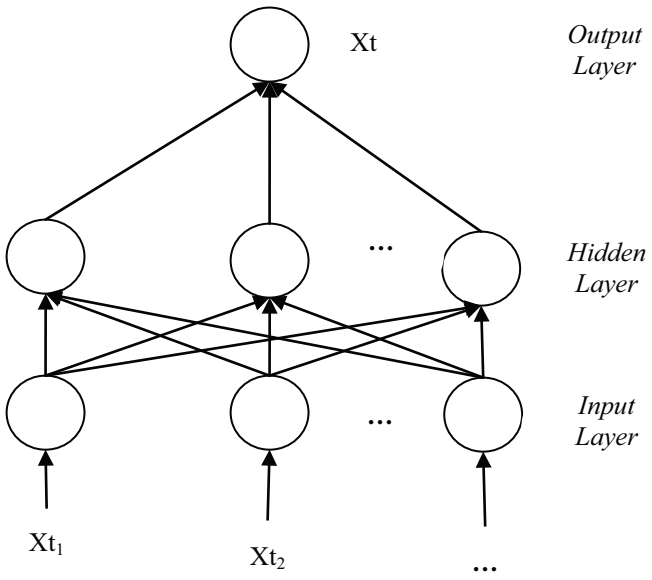

Fig. 2. Network Model of ELM

Step 1: Training and Testing Data Sets

Training and testing process is absolutely necessary in the diagnosis process using ELM. Training process used to developthe ELM while testing is used to evaluate the ability of ELM asa diagnosis tool. Distribution of training data and testing will beshared with the following composition:

- Training data as much as 75% of the total data
- Testing data as much as 25% of the total data

Step 2: Training ELM

- Training Data Normalization
  The data that will be used in ELM process is normalized so they have a certain range of values. This is necessary because the activation function produce output with a range of data [0,1] or [-1,1]. The following is formula that is used for normalization:

$$x = 2 \times \frac{(x_p - min\, x_p)}{(max\, x_p - min\, x_p)} - 1$$

  Where:
  x = value after normalization in the range between -1 to 1
  x_p = original data value which is unnormalized
  min x_p = minimum value in the data set
  max x_p = maximum value in the data set

- Determine Activation Function and Number of Hidden Neurons
  In the training process, activation function of ELM and the number of hidden neuron is determined in advance. This study uses sigmoid activation function.ELM produce a stable output forecast with the number of hidden neurons from 0 to 30 [3]. However, if the output that is produced not optimal, then number of hidden neurons will be changed.

- Calculate Input Weights, Bias of Hidden Neuron, and Output Weights
  Output from ELM training process are input weight, output weight, and bias of hidden neuron with a low error rate measured by MSE. Input weights are defined by random, while output weights are inverse of hidden layer and output matrix. Mathematically can be written as follows:

$$\beta = H^T T$$

$$H = (w_i, \dots, w_N, b_i, \dots, b_N, x_i, \dots, x_N) = \begin{matrix} g(w_i.x_i + b_i) & \cdots & g(w_N.x_i + b_N) \\ \vdots & \vdots & \vdots \\ g(w_i.x_N + b_i) & \cdots & g(w_N.x_N + b_N) \end{matrix}$$

$$\beta = \begin{matrix} \beta_1^T \\ \vdots \\ \beta_N^T \end{matrix} \qquad T = \begin{matrix} t_1^T \\ \vdots \\ t_N^T \end{matrix}$$

- Training Data Denormalization
  Output from ELM training process will be denormalized using the following formula:

$$x = 0.5 \times (x_p + 1) \times (\max x_p - min\, x_p) + min\, x_p$$

  Where:
  x = value after denormalization
  x_p = original data value which is normalized
  min x_p = minimum value in the data set before normalization
  max x_p = maximum value in the data set before normalization

Step 3: Testing ELM

Based on the input weights and output weights obtained from the training process, the next step is to diagnose using the ELM. At this stage, input data is normalized and denormalized using the same range and formula with the training data.

D. ELM Performance Analysis

Accuracy analysis of the ELM performance will be done by seeing if the error rate of Mean Square Error (MSE) is small. Here is the formula of the MSE:

$$mse = \frac{1}{N}\sum_{i=1}^{N}(y_i - t_i)^2$$

Where:
N = number of data
y_i = output data
t_i = actual data

To demonstrate the speed and accuracy of ELM learning process, the ELM performance will be compared with backpropagation on the same data.

IV. RESULT AND DISCUSSION

A. Data Collection

By utilizing search facility on the site UCI Machine Learning Repository, the classification of the data sets can be obtained. The data consisted by 769 women population which lives in Phoenix, Arizona, and taken according the criteria of WHO. The value of the data sets is provided in text file.

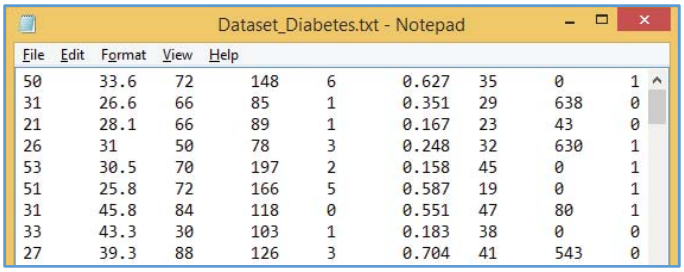

Fig. 3. Extraction of Data Set

B. Experiment Results

Source code [15] is modified to suit the input of this study. Numbers of selected inputs are 8 pieces such as: age, body mass index, the diastolic blood pressure, plasma glucose concentrate in an oral glucose tolerance test, the number of pregnancies, diabetes pedigree function, triceps skin fold thickness, and serum insulin.

Determination number of hidden neuron will affect the optimal prediction of ELM. Prediction of ELM produces a stable output forecasting with 0 to 30 hidden neurons range. Figure 3 shows the accurate data of predictions that generated using the ELM method based on the value number of hidden neurons. The accuracy could be seen from the value of output

that has been analyzed by looking at the error rate of MSE. The error rate of MSE approaching 0 is the best result.

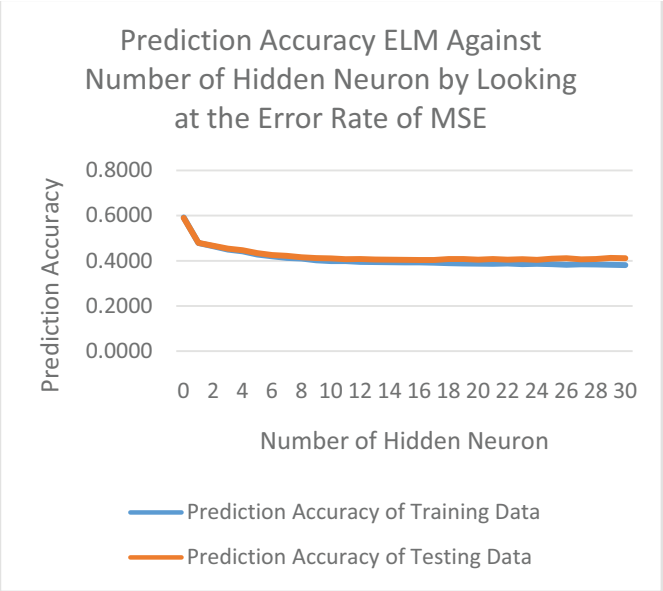

Fig. 4. Graph of Prediction Accuracy ELM Against the Change Number of Hidden Neurons by Looking at the Error Rate of MSE

Table II shows comparison performance in speed between ELM and backpropagation. ELM is excellent in terms of speed of training and testing which it could do in a split of second with the same data. Backpropagation takes slower than ELM. ELM's advantages in speed may not be too visible in this paper because the small number of data. But if the number of data is large, the use of ELM will be very influential.

TABLE II. COMPARISON OF ELM AND BACKPROPAGATION IN TRAINING AND TESTING SPEED.

Data (person)	ELM		Backpropagation	
	Training Speed (s)	Testing Speed (s)	Training Speed (s)	Testing Speed (s)
1	0.5156	0.0085	0.9040	0.0116
2	0.0937	0.0057	0.5002	0.0501
3	0.0938	0.0009	0.6810	0.0140
4	0.0625	0.0064	0.4225	0.0204
5	0.0781	0.0006	0.1091	0.0170
Average	**0.1687**	**0.0044**	0.5234	0.0226

Table II shows that the speed of ELM in predicting the training data is 3.102 times faster and 5.136 times faster than backpropagation. Results of the speed approaching 0 seconds are the best result.

Table III shows comparison in accuracy for training and testing accuracy between ELM and backpropagation. It shows that ELM also has good level accuracy than backpropagation.

TABLE III. COMPARISON OF ELM AND BACKPROPAGATION IN TRAINING AND TESTING ACCURACY

Data (person)	ELM		Backpropagation	
	Training Accuracy (mse)	Testing Accuracy (mse)	Training Accuracy (mse)	Testing Accuracy (mse)
1	0.370	0.387	0.483	0.894
2	0.368	0.382	0.425	0.901
3	0.361	0.392	0.372	0.854
4	0.362	0.405	0.390	0.763
5	0.373	0.399	0.386	0.798
Average	**0.367**	**0.393**	0.411	0.842

The accuracy is obtained from the value of output that has been analyzed by looking at the error rate of MSE. Table III shows that the accuracy of ELM's performance in predicting training data is 1.12 times and 2.14 times better than backpropagation.

V. CONCLUSION

The results of this study showed that ELM has a high accuracy and a very good speed in diagnosed diabetes mellitus. In terms of speed, ELM's performance is better than the performance of backpropagation, where the speed obtained when the prediction process start per each data. ELM's performance in terms of accuracy is also better than the performance of backpropagation, where the accuracy could be seen from the value of output that has been analyzed by looking at the error rate of MSE. The error rate of MSE for the overall ELM testing data is 0.4036 and the error rate of MSE for the overall backpropagation testing data is 0.9425.

REFERENCES

[1] N.L. Sumadewi, "Isolasi Senyawa Tanin dan Uji Efek Hipoglemik Ekstrak Kulit Batang Bungur (Lagerstroemia Speciosa Pers.) terhadap Darah Mencit yang Diinduksi Aloksan," Unud Library, 2011.

[2] Darmono, "Pengaturan Pola Hidup Penderita Diabetes Untuk Mencegah Komplikasi Kerusakan Organ-Organ Tubuh," Semarang, Diponegoro University, 2005.

[3] Z.L. Sun, T.M. Choi, K.F. Au, and Y. Yu, "Sales Forecasting using Extreme Learning Machine with Application in Fashion Retailing," Elsevier Decision Support Systems 46, pp. 411-419, 2008.

[4] I.D. Agustina, W. Anggraeni, and A. Mukhlason, "Penerapan Metode Extreme Learning Machine untuk Peramalan," ITS Library, 2010

[5] PERKENI, "Konsensus Pengendalian dan Pencegahan Diabetes Melitus Tipe 2 di Indonesia," Jakarta, 2011.

[6] F. Masykur, "Implementasi Sistem Pakar Diagnosis Penyakit Diabetes Melitus Menggunakan Metode Fuzzy Logic Berbasis Web," Undip Library, 2012.

[7] C. Vercellis, "Business Intelligence: Data Mining and Optimization for Decision Making," Chennai: John Wiley & Sons, 2009.

[8] R. Singh and S. Balasundaram, "Application of Extreme Learning Machine Method for Time Series Analysis," World Academy of Science, Enginering and Technology, 2007.

[9] G.B. Huang, Q.Y. Zhu, and C.K. Siew, "Extreme Learning Machine: Theory and Applications," Elsevier Science: Neurocomputing 70, pp. 489-501, 2006.

[10] E.A. El-Sebakhy, "Extreme Learning Machine as a New Framework in Predicting Material Properties: Methodology and Comparison," in the 12th International Conference of International Association for Computer Methods and Advances in Geomechanics, Goa, India, 2008.

[11] F.L. Chen and T.Y. Ou, “Constructing a Sales Forecasting Model by Integrating GRA and ELM: A Case Study for Retail Industry,” International Journal of Electronic Business Management, pp. 107-121, 2011.

[12] G. Duan, Z. Huang, and J. Lan, “Extreme Learning Machine for Bank Clients Classification,” IEEE, 2009.

[13] G.B. Huang, D.H. Wang, and Y. Lan, “Extreme Learning Machine: A Survey,” Springer, pp. 107-122, 2011.

[14] J.D. Balkina, “Automatic Neural Network Modeling for Univariate Time Series,” Elsevier, no. International Journal of Forecasting, 2000.

[15] [Online] http://www.ntu.edu.sg/home/egbhuang/