Using the Extreme Learning Machine (ELM) Technique for Heart Disease Diagnosis

Salam Ismaeel, Ali Miri and Dharmendra Chourishi

Department Computer Science, Ryerson University, Toronto, Canada
salam.ismaeel, ali.miri, dharm@ryerson.ca

***Abstract*—One of the most important applications of machine learning systems is the diagnosis of heart disease which affect the lives of millions of people. Patients suffering from heart disease have lot of independent factors such as age, sex, serum cholesterol, blood sugar, etc. in common which can be used very effectively for diagnosis. In this paper an Extreme Learning Machine (ELM) algorithm is used to model these factors. The proposed system can replace a costly medical checkups with a warning system for patients of the probable presence of heart disease. The system is implemented on real data collected by the Cleveland Clinic Foundation where around 300 patients information has been collected. Simulation results show this architecture has about 80% accuracy in determining heart disease.**

***Index Terms*—Extreme learning machine (ELM), Neural Networks, Heart Disease, Prediction and Diagnosis Systems, Pattern Classification**

I. Introduction

Heart disease is the major cause of death in today's world. The need to predict heart-attacks is a necessity for improving the healthcare sector. Accurate and precise prediction of heart disease mainly depends on prior knowledge and available information from similar pathological cases [1]. Heart disease patients have lot of factors such as blood pressure, cigarette smoking, cholesterol, diabetes, age, etc. that can be used very effectively for diagnosis [2].These factors are independent of each other, thus using artificial intelligence (AI) and machine learning systems will be a suitable choice to model them.

Intelligence systems have the ability to learn or adapt, and to modify the functional dependencies in response to new experiences or due to changes in functional relationships [3]. Extreme Learning Machines (ELMs) have all of these abilities. They all display excellent generalization performance, require little human intervention, and low computational time for training [4]. These features make the ELM more favourable than recently used techniques for the prevention of heart disease risk such as Feed-Forward Neural Network (FFNN) with back propagation training [5]–[7]. Genetic neural networks [2], [8], fuzzy-neural networks [9], and Learning vector quantization [10]. The main contribution in this work is that it improves detection of heart-disease risk by utilizing the ELM scheme described by Fathurachman *et. al.* [11]. Specifically:

- The output consists of 5 values, each of which represents a stroke level between 0-4 as described by the standard Cleveland collection databases [12] instead of having one target class.
- The proposed architecture uses all previous data to predict a new patient's state, and it is not necessary to separate the real data into training and testing phases, making this technique more suitable for upgrading of the information as each new patient represents an updated data set.

The rest of the paper is organized as follows. Section II provides background about the ELM technique. In section III, we describe the proposed ELM algorithm for heart disease with classification. Section IV includes the implementation and results discussion of the proposed architecture on the Cleveland database and Section V concludes the work.

II. The Extreme Learning Machine Technique (ELM)

ELM was first developed by Huang in 2004, and was published in 2006 [13]. It consists of a three layer neural network with sigmoid activation function. The last modification of the original algorithm was done by the same author in 2011 [14]. The main advantages of ELM compared with conventional gradient based learning methods are [15]–[17]:

- It avoids many problems such as stopping criteria, learning rate, learning epochs, and local minimum that are difficult to cope with in classical methods.
- It is a fast learning algorithm because the learning of ELM is one pass without re-iteration.
- It can obtain better generalized performance than back propagation (BP) in most cases.
- It is suitable for almost all nonlinear activation functions.

The mathematical model of the ELM is described as follow [4], [18]:

Suppose there are N samples of data (x_i, t_i), where $x_i = [x_{i1}, ..., x_{im}] \in \Re^m$ and $t_i = [t_{i1}, ..., t_{iq}] \in \Re^q$. The extreme learning machine (ELM) algorithm consists of a single hidden layer feed-forward neural network with $\tilde{N}$ hidden nodes and the activation function $g(x)$ is shown as:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_j) = \sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i.x_j + bi) = o_j, \quad j = 1, ..., N \tag{1}$$

where $w_i = [w_{i1}, w_{i2}, ..., w_{im}]^T$ is the weight vector of the connectors from the input node to the i^{th} hidden node, and $\beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{iq}]^T$ is the weight vector of the connectors between the i^{th} hidden node and the output nodes. The variable b_i is the threshold of the ith hidden node. By approximating the samples with zero error, i.e.:

$$\sum_{i=1}^{\tilde{N}} \parallel o_j - t_j \parallel = 0 \tag{2}$$

we see that there exist w_i, β_i, and b_i such that:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(w_i.x_j + bi) = t_j, \quad j = 1, ..., N \tag{3}$$

By using the following substitutions: $H(w_i, ..., w_{\tilde{N}}, b_i, ..., b_{\tilde{N}}, x_1, ..., x_N) =$

$$\begin{bmatrix} g(w_1.x_1 + b_1) & \cdots & g(w_{\tilde{N}}.x_1 + b_{\tilde{N}}) \\ \vdots & \vdots & \vdots \\ g(w_1.x_N + b_1) & \cdots & g(w_{\tilde{N}}.x_N + b_{\tilde{N}}) \end{bmatrix} \tag{4}$$

where $\beta = \left[\beta_1^T, ..., \beta_{\tilde{N}}^T\right]^T$ and $T = \left[t_1^T, ..., t_{\tilde{N}}^T\right]^T$, all N equations of Eq. 3 can be written as:

$$H\beta = T \tag{5}$$

From the previous description of the ELM set of equations we can notice that:

1) Input weights and bias as w and b can be randomly generated and used without tuning.
2) The output weights β can be tuned by a simple and fast operation.
3) Only two parameters need to be selected by the user, i.e. number of hidden neurons (L) and regularization (C), which will be presented in the proposed algorithm.

III. Proposed ELM algorithm for Heart disease classification

Our proposed classifier architecture, used for classification of heart stroke level with the help of 13 inputs, is derived from the standard database and 5 target values of category of disease, each of which represents a class. The resulting ELM classifier has the structure depicted in figure 1.

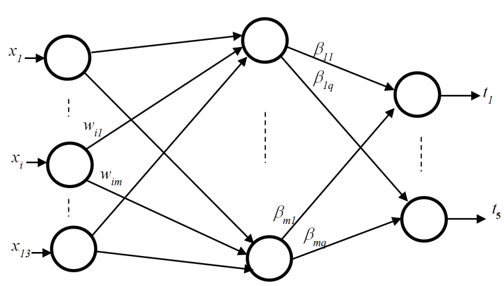

Figure 1: Proposed ELM architecture

Algorithm 1, describes the steps of the proposed diagnostic system. As we can see from the algorithm steps, the output binary values are changed to [-0.9 +0.9] to make the output more distinguishable; the two selected parameters that effect the performance of the network are the constant value C, which is chosen here to be 10000, and number of hidden neurons L, which is set to 100 for most of the experiments.

The biggest calculation effort will be done in the inversion of the matrix G. Whenever the dimensions are big, the computer needs more time and even specific inverse calculation algorithms, like the Moore-Penrose generalized inverse which has been used by Samet and Miri [18] to increase the speed of computation.

Algrithm1: Proposed ELM Algorithm

1: Detect and remove outliers from the dataset
2: Normalize the data to $[-0.9\ 0.9]$
3: Convert the required output, single value between from 0-4 to sets of 5 values $[-0.9\ 0.9]$
4: Choose the values of: the regulation parameter $C = 10000$, and the number of hidden neurons L
5: Randomly initialize weight w_i and bias value b_i
6: Calculate the hidden layer output matrix $H(x, w, \beta)$ using Eq. 4
7: Calculate G^{-1} where:

$$G = \frac{1}{C} + H^T H \tag{6}$$

8: Calculate the output of the hidden layer respective to new unknown input $h(z, w, \beta)$ where z is the new input.
9: Calculate the output of ELM for the unknown input:

$$y_{ELM} = h\beta = h\left(\frac{I}{C} + H^T H\right) H^T T = hG^{-1} H^T T \tag{7}$$

IV. Simulation results and discussion

The proposed technique is used to model the Cleveland Clinic dataset. This database contains 76 attributes, but all published experiments refer to using a subset of 13 numeric attributes as input and the attribute 14 represents the output. The output refers to the presence of heart disease in the patient. It is integer value from 0 (not present) to 4. Experiments with the Cleveland database have concentrated on simply attempting to distinguish presence (values 1,2,3,4) from absence (value 0) [12].

This data has been used extensively for testing in various machines learning techniques. Table I shows the data attributes. As we can see, all the inputs are totally independent. In other words, there is no mathematical model which can represent these variables.

Figure 2 shows the relation between the number of neurons and the percentage error. As expected, the error decrease when the number of neurons increases until it reaches zeros at 190 neurons. The performance of the proposed ELM architecture

Table I: Data atributes

Attribute	Description	Data type
age	years	Continuous
sex	1:male, 0:female	Binary
Chest Pain	1:typical angina, 2:atypical angina, 3:non-anginal pain, 4:asymptomatic	$1-4$
Rbp	Resting blood pressure in $mmHg$	Continuous
Chol	Serum cholestoral in mg/dl	Binary
Fbs	fasting blood sugar $> 120mg/dl$	Binary
Resteeg	Resting electrocardiographic results, 0:normal, 1:having ST-T wave abnormality, 2:showing probable or definite left ventricular hypertrophy by Estes' criteria	$0-2$
Thalach	Maximum heart rate achieved	Continuous
Exang	Exercise induced angina	Binary
Oldpeak	ST depression induced by exercise relative to rest	Continuous
Slop	Slope of the peak exercise ST segment 1:upsloping, 2:flat, 3:downsloping	1-3
Ca	Number of Major Vessels $(0-3)$ colored by flourosopy	Continuous
Thal	3: ormal, 6:fixed defect, 7: reversible defect	$3, 6, 7$
Output	Class Target the predicted attribute	$0-4$

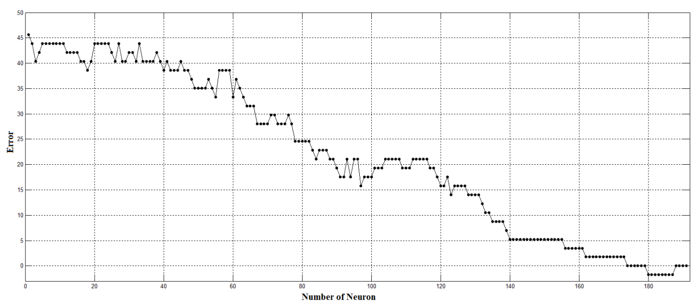

Figure 2: Number of Neuron and Persantage Error

for 100 neurons shows that this model is about 86.5% accurate. This output gives exactly the presence value of the heart disease $[0 \ -4]$ instead of giving Yes/No presence as in Fathurachman et al [11]. Although Fathurachman's model's accuracy was 84% but it was just for one output (0 and 1), i.e. two classes only.

The ELM model with five outputs is more accurate than the back propagation neural network (BPNN) proposed by Patel and Joshi [6]. The accuracy of their model was 67% for unknown or new data sets for 4 classes $(0-3)$, instead of the 5 classes in our work.

V. Conclusion

In this work, the ELM intensity is used to model heart disease from real datasets provided by the Cleveland Clinic Foundation. This model has five outputs (0-4), gives better performance when it is compared with other models of BPNN with four classes or ELM with a single output.

With the ELM technique, we can use all previous knowledge to analyze a new patient situation. The ELM solves the problems of learning time, making it more suitable for implementation on Big Data. It can also be developed further to fix the problem of missing attributes by estimating the effect of these attributes on the overall decision.

References

[1] D. Ravish, N. Shenoy, K. Shanthi, and S. Nisargh, "Heart function monitoring, prediction and prevention of heart attacks: Using artificial neural networks," in *Contemporary Computing and Informatics (IC3I), 2014 International Conference on*, Nov 2014, pp. 1–6.

[2] H. Murthy and M. Meenakshi, "Dimensionality reduction using neuro-genetic approach for early prediction of coronary heart disease," in *Circuits, Communication, Control and Computing (I4C), 2014 International Conference on*, Nov 2014, pp. 329–332.

[3] S. Ismaeel and K. Al-Jebory, "Adaptive fuzzy system modeling," *Engineering Technology*, vol. 20, no. 4, pp. 201–212, 2001.

[4] J. Xin, Z. Wang, L. Qu, and G. Wang, "Elastic extreme learning machine for big data classification," *Neurocomputing*, vol. 149, Part A, no. 0, pp. 464 – 471, 2015.

[5] R. R. Reader and S. V. Dudul, "Intelligent diagnosis of heart diseases using neural network approach," *International Journal of Computer Applications*, vol. 1, no. 2, pp. 97–102, 2010.

[6] A. Patel and M. Joshi, "Heart diseases diagnosis using neural network," in *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on*, July 2013, pp. 1–5.

[7] J. Sonawane and D. Patil, "Prediction of heart disease using multilayer perceptron neural network," in *Information Communication and Embedded Systems (ICICES), 2014 International Conference on*, Feb 2014, pp. 1–6.

[8] S. Amin, K. Agarwal, and R. Beg, "Genetic neural network based data mining in prediction of heart disease using risk factors," in *Information Communication Technologies (ICT), 2013 IEEE Conference on*, April 2013, pp. 1227–1231.

[9] B. Rajakumar and A. George, "On hybridizing fuzzy min max neural network and firefly algorithm for automated heart disease diagnosis," in *Computing, Communications and Networking Technologies (ICCCNT),2013 Fourth International Conference on*, July 2013, pp. 1–5.

[10] J. S. Sonawane and D. Patil, "Prediction of heart disease using learning vector quantization algorithm," in *IT in Business, Industry and Government (CSIBIG), 2014 Conference on*, March 2014, pp. 1–5.

[11] M. Fathurachman, U. Kalsum, N. Safitri, and C. Utomo, "Heart disease diagnosis using extreme learning based neural networks," in *Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of*, Aug 2014, pp. 23–27.

[12] "Cleveland dataset." [Online]. Available: https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.cleveland.data

[13] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1â3, pp. 489 – 501, 2006.

[14] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 42, no. 2, pp. 513–529, April 2012.

[15] X. Wang and M. Han, "Improved extreme learning machine for multivariate time series online sequential prediction," *Engineering Applications of Artificial Intelligence*, vol. 40, no. 0, pp. 28 – 36, 2015.

[16] M. M. Ma and B. He, "Study on deep structure of extreme learning machine (ds-elm) for datasets with noise," in *Advanced Materials Research*, vol. 989. Trans Tech Publ, 2014, pp. 3679–3682.

[17] G.-G. Wang, M. Lu, Y.-Q. Dong, and X.-J. Zhao, "Self-adaptive extreme learning machine," *Neural Computing and Applications*, pp. 1–13, 2015.

[18] S. Samet and A. Miri, "Privacy-preserving back-propagation and extreme learning machine algorithms," *Data Knowledge Engineering*, vol. 79,80, no. 0, pp. 40 – 61, 2012.