



Predicting Cardiovascular Disease as a Long-Term Diabetes Complication using SOM

K. Rajathi, R. Asmetha Jeyarani, K. Blessing Christiana, T. Vijaya Vahini

Abstract: Cardiovascular disease (CVD) is the most common serious of long term type2 diabetic's mellitus. It is estimated that most of the T2DM patients causes death due to CVD. Around 90% of CVD can be prevented with proper prediction of diabetes. Type 2 diabetes mellitus begins with insulin resistance, a condition in which it fails to respond to insulin properly. This paper explores Hybrid Wavelet Neural Network to train the system to learn the pattern to predict the disease and Self Organized map method is used for information clustering and visualization of excessive dimensional records to predict the disease with less parameter high accuracy which can help to prevent the disease. Modified Teaching Learning Based Optimization algorithm achieves the optimized learning from the pertained network. Teaching and learning based optimized technique yield better accuracy with a dataset of 770 patients. The measure of accuracy is compared with other algorithms and it is analyzed for further ratification.

Keywords: Cardiovascular (CVD), Type2 diabetes mellitus (T2DM), Self-organized mapping (SOM), Hybrid wavelet neural network, Modified teaching learning based optimization.

I. INTRODUCTION

CVD is a first-rate cause of demise and incapacity among people with diabetes. CVD consists of stroke, coronary artery ailment and peripheral artery sickness. People with diabetes are at improved chance of CVD, and these activities commonly arise at an in advance age in comparison to people without diabetes [1]. As the quantity of humans with diabetes is predicted to boom, the outlook for CVD will become even greater alarming. The worldwide impact of the sickness it

people of these because of cardiovascular headaches. Populous countries have the largest absolute quantity of humans demise from CVD, become estimated that over 37.9 million human beings international died from non-communicable sicknesses, of which 17.5 million have been because of CVD. In 2015, about five million had been predicted and died due to diabetes most encompass in china, India and Russian Federation excessive

In 2015, about five million had been predicted to have died from diabetes, most and encompass China, India and the Russian Federation. Excessive-income countries are more likely to have low CVD mortality prices, notwithstanding having a high occurrence of CVD chance factors including diabetes, obesity and raised cholesterol. The general public of the chance prediction fashions, check with the calculation of the 5 year CVD hazard for patients with T2DM. The use of the most unusual threat predictors including the age, the intercourse, the period of identified diabetes, the glycosylated hemoglobin (HbA1c) concentration and the smoking dependency [2].

Here using gadget learning strategies tactics the pleasant prediction and accuracy goes to be performed. While the existing device learnings strategies used in the paper before in the classification of strategies. Class of diabetes and CVD using synthetic Neural Networks (ANNs) and Bayesian Networks which indicates that better possibility to obtain greater correct results in diabetes and/or CVD type is while it is carried out to ANN. Some of the proposed procedures in the existing literature are 1. Branch and sure processes, 2. Sequential forward/backward seek algorithms 3. Stochastic algorithms, which include simulated annealing, ant colony optimization, neural networks and genetic algorithm. All these is implemented using a k-Nearest Neighbors (KNN) classifier method. Neural networks Hybrid with a Genetic algorithm (GA) have been used for the identity of important subsets of risk factors for CVD headaches.

A. Dataset

This proposed method for threat prediction models deals with seven hundred T2DM sufferers which consider variety of factors. In this dataset, one hundred fifty five T2DM sufferers out of the 700 T2DM sufferers, around 7.32% evolved deadly or non-deadly CVD. Out of the a hundred and fifty five patients with CVD incidents, 4 patients skilled stroke and the rest experienced CHD. The prevalence of fatal or non-deadly CVD (high-quality instances) changed into en- coded to one and the non-occurrence of CVD (poor instances) to 0 [3].

Manuscript received on February 10, 2020.

Revised Manuscript received on February 20, 2020.

Manuscript published on March 30, 2020.

* Correspondence Author

K. Rajathi*, Associate Professor, Department of CSE, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, India. Email:K.rajathimtech@gmail.com

R. Asmetha Jeyarani, Assistant Professor, Department of CSE, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, India. Email: asmetha.r@gmail.com

K. Blessing Christiana, Assistant Professor, Department Department of IT, New Prince Shri Bhavani College of Engineering & Technology, Chennai, India.

Email:hblesingchristiana@gmail.com.

Ms Vijaya Vahini, Department of CSE, Vel Tech Rangarajan Dr Sagunthala R&D Institute of Science and Technology, Chennai, India. Email:vijivahini28@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Predicting Cardiovascular Disease as a Long-Term Diabetes Complication using SOM

Risk Factor - Categorical variables		Number of patients out of 700	Number of patients in (%)
Sex	Male	358	59.96
	Female	402	60.04
	Hyper tension	260	46.42
Smoking Habit	Non-smokers	146	26.61
	Current smokers	253	42.07
	Ex- smokers	125	22.32
Chest pain	Asympt	252	36.02
	Non_anginal	292	41.71
	atyp_angina	145	20.71

Table 2: Risk Factors For The Incidence Of Fatal Or Non-Fatal Cvd In T2dm

Risk Factor - Continuous Variable	Average	Standard Deviation
Age	58.56	10.7
Diabetes duration	7.67	7.37
Body Mass Index (BMI)	9.49	5.54
Glycosylated Haemoglobin	7.43	1.81
Pulse Pressure (mmHg)	56.75	15.8
Fasting Glucose (mg/dL)	165.15	56.15
Total Cholesterol (mg/dL)	226.64	50.04
Triglycerides (mg/dL)	167.39	110.81

The above considered danger factors composing the enter space are summarized in desk alongside their descriptive statistic baseline facts related to the medical popularity of a T2DM affected person. The use of information accrued from screening helps to build the proposed danger prediction model renders every consideration issue prompted the CVD danger as evidenced with the aid of several research. In particular, age constitutes a critical risk thing by way of about tripling the CVD hazard with every decade. Period of diabetes and extended BMI, growth the hazard. As a trademark of the average blood glucose concentrations over the previous 2–three months, glycosylated hemoglobin (HbA1c) level has been demonstrated to be an independent chance element for CVD activities. The relation between Pulse stress (PP) and CHD is nonlinear in patients with T2DM. Patients with PP outside the variety of 45 and 55 mmHg are at expanded threat of future CHD. Multiplied fasting glucose ranges had been additionally appreciably related to CVD and all-motive mortality. A typical cholesterol levels, including excessive LDL and HDL cholesterol, in conjunction with excessive triglycerides stages indicate terrible lipid counts which

regularly arise in sufferers with premature CHD. Certainly, active smoking is related to increased risks, up to 50%, of CVD events in T2DM. Despite the fact that men are at greater risk of coronary heart disease than pre-menopausal ladies, ladies with diabetes are twice as probable as adult males with diabetes to expand coronary heart disease [7]. High blood pressure is quantitatively the maximum crucial hazard component for untimely CVD, accounting for a predicted 54 percentage of all strokes and forty seven percent of all ischemic coronary heart disorder events globally. Lipid-lowering therapy and anti-thrombotic marketers (e.g., aspirin) are considered to be protective elements for the incidence of CVD [8]. Several big trials have proven that insulin does no longer growth the CVD danger in spite of proof of more suitable atherosclerosis obtained from in vitro research. Parental history of diabetes has been related to the prevalence of CVD in T2DM.

II. METHODS

In present study the machine learning techniques used are SOM, HWNN, and MTLBO. Comparing these three techniques we conclude which method gives the better accuracy for the imbalanced dataset of 700 sufferers of CVD and T2DM.

A. Hybrid Wavelet Neural Network

Wavelet Neural community (WNN) belong to new class of neural network with unique skills in gadget identity and type. The concept of WNN is in- spired by both the technology of wavelet decomposition and Neural Networks [4]. Especially, wavelets are obtained from a single prototype wavelet $\psi(t)$ referred to as mother wavelet. Inside the gift observe, a shape of a feed forward Hybrid WNN (HWNN) turned into wavelet and the sigmoid characteristic. As a way to handle propagation's problem of local minima, two momentum phrases have been utilized in its gaining knowledge of algorithm.

$$J = \Gamma_i N \cdot (y_i^{HWNN} - Y_{real})$$

$$\Gamma_i = 1$$

$$\text{If } Y_{real} = 0$$

in which HWNN and Y_{real} are the HWNN output and the goal output, respectively, and N is the variety of the sample of the schooling data set. Through applying the sub-sampling technique, the HWNN changed into educated with representation of positive instance same to 33.33% given that the CVD prevalence fee in the general T2DM [3] population is lower, education the HWNN with this dataset could keep away from over-becoming but might reason the prediction of high hazard ratings which would, in flip, result in bad calibration. To overcome this problem, a weighting coefficient $t(\Gamma)$ become applied to the cost function so one can decrease the contribution of the advantageous instances to the weights adjustment.

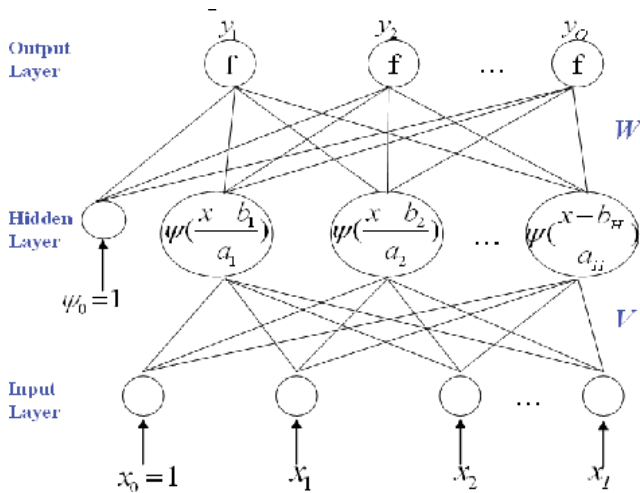


Fig 1: Wavelet Neural Network

B. Self Organized Mapping

In popular, SOMs belong to the category of aggressive examining networks and are extensively used for information clustering and visualization of excessive dimensional records [5]. Their philosophy is based totally at the notion of unsupervised getting to know, in line with which a system can learn, how to constitute input records in the absence of any facts related to the goal outputs. But, SOMs can be skilled to analyze enter-output mappings and can be, for that reason, effectively applied for characteristic approximation, via an aggressive-cooperative getting to know scheme. The neurons of the SOM are driven to capture the spatial relationships of input information and ultimately carry out a vector quantization of the enter space.

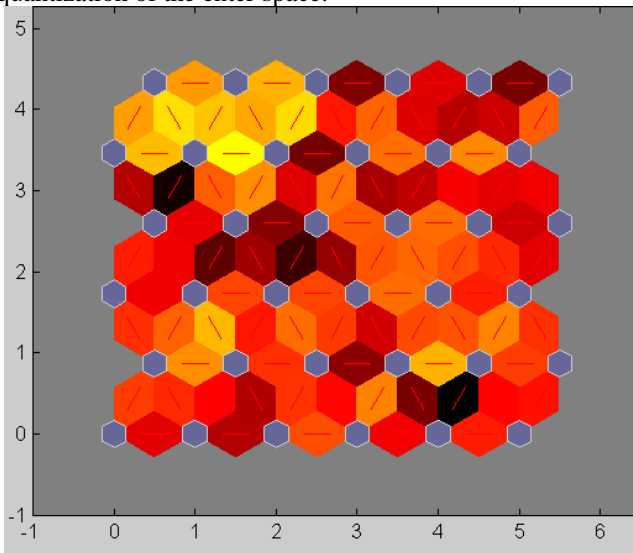


Fig 1.2: cluster diagram of SOM

Within the present look at, a technique that generated a couple of neighborhood linear fashions following a SOM-based vector quantification turned into deployed. SOM consisted of a dimensional grid of neurons (j), each of which became related to a weight vector (win) of identical size, a coefficient vector (v) of a linear autoregressive version with exogenous inputs and a weight output cost (W_{out}). All weights have been randomly initialized and subsequently updated at every new release of the education stage, based at the Euclidean distance among the modern example vector (x_{in}) and the load.

C. Teaching Learning Based

Recently, the nature-inspired evolutionary algorithms showed better performance and this benefit made them a popular tool in the optimization applications in a very short period [6]. Therefore, there is a wide range of heuristic algorithms which the most well-known can be named such as particle swarm optimization (PSO) algorithm, shuffled frog leaping algorithm, honey bee mating optimization algorithm, etc. In contrast with the better performance of the above evolutionary algorithms, the dependency on the adjusting parameters is their main drawback. In 2011, Rao *et al.* proposed a novel population based optimization algorithm inspired from the relationship between the teacher and learners. In this author illustrated the superiority of the TLBO compared with other famous algorithms. In the TLBO, two groups of the teacher and learners compose the population. The optimization procedure of the algorithm is simulated in two phases, first phase is Teacher Phase or gain knowledge from the teacher student's progress and second phase is the Learner Phase or gain.

Knowledge by mutual action among the students themselves. In the first Teacher Phase, the teacher intends to raise the knowledge of the class up to his/her level. In the second Learner Phase, the knowledge of the students will be improved among themselves by the use of interaction appropriately. As a matter of fact, this phase simulates what happens in the real world by means of presentations, friend communications, group debates, etc.

D. Optimization

For more explanation on TLBO mechanism, the algorithm generates a random initial population such that all the constraints are taken into account. Each row in the student population indicates a student and each column indicates grades which are obtained by the class in this population. After that, the Teacher Phase should be applied to the program. Here the best individual according to the objective function, is defined as the teacher *XTeacher*. Then the columnar mean value of the class is expressed as

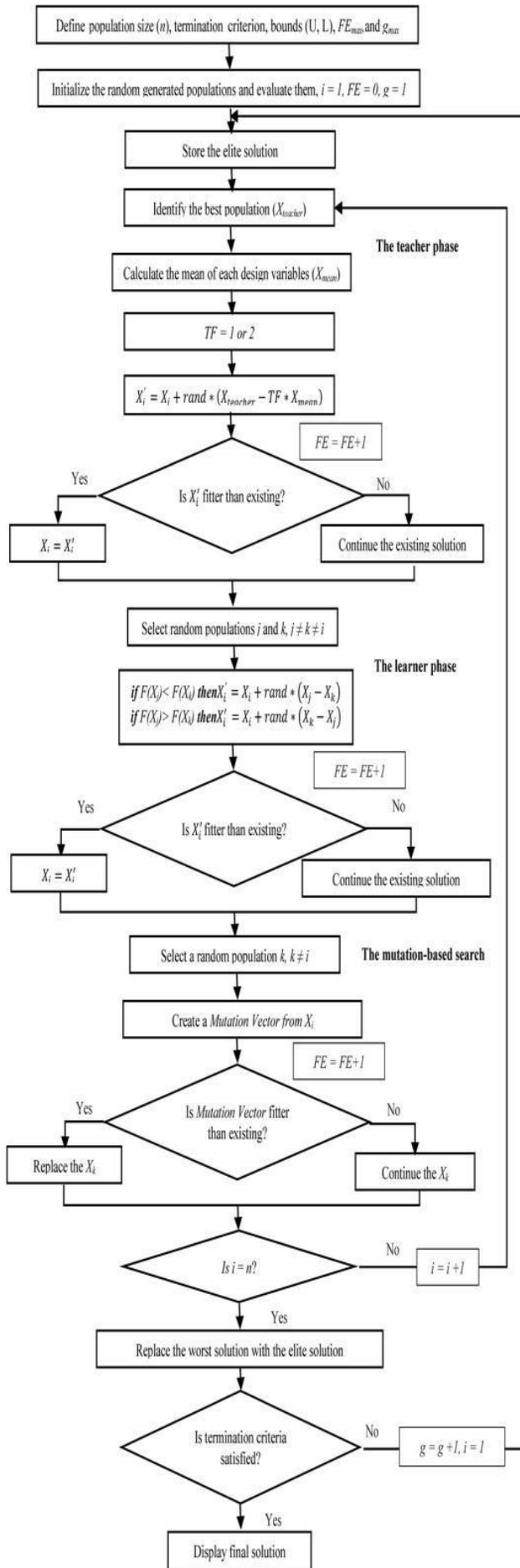
$$Mean\ Q = [m_1, m_2, m_3, \dots, m_N]$$

Now, the teacher tries to improve the mean grade of the class toward his/her position. Then a new position generated for each of the students as follows:

$$X_{new, Q} = X_{old, Q} + \gamma (X_{Teacher} - TF \cdot MQ)$$

If the student gains higher knowledge than before, in other words, *X_{new, Q}* has better position than *X_{old, Q}* then accept *X_{new, Q}*. Now the learner phase should start its process among the students considering that *i* ≠ *j*, two students *i* and *j* are selected and the following equation is implemented to simulate the Learner Phase

Predicting Cardiovascular Disease as a Long-Term Diabetes Complication using SOM



E. Modified Teaching Learning Based Optimization

For $i = 1: N_{class}$

If $F(X_i) < F(X_j)$
 $X_{new,i} = X_{old,i} + \mu_1(X_i - X_j)$
 If $F(X_i) < F(X_j)$
 $X_{new,i} = X_{old,i} + \mu_2(X_i - X_j)$
 End if
 End for

Where $F(X)$ is the objective function value of the student X . Also, μ_1 and μ_2 are random values in the range of (0, 1).

As recently discussed, TLBO is a simple population based optimization algorithm which has proved its superiority over the other renowned algorithms. The main conspicuous characteristics of TLBO can be count as, 1) easy implementation 2) undependability of the algorithm on its adjusting parameters 3) fast convergence to the optimal solution 4) high performance stability 5) little memory requirement for implementing, etc. Nevertheless, in some optimization problems, the Learner Phase simulation causes an increase in probability of being trapped in local optima. This may be the consequence of an inappropriate knowledge interaction among the students that results in a deceptive movement of all the whole population toward the local optima solution. Therefore, in order to improve the output solution of the Learner Phase, the following by the aid of selecting three students ($d1, d2, d3$)

$$X_{muted} = X_{d1} + \gamma \times (X_{d2} - X_{d3})$$

Where μ is a random value in the range of (0, 1). At this moment, three new modified learners are produced by employing $X_{Teacher}$, X_{muted} and X_s

$$x_{muted}, j, \text{ if } \theta_1 \leq \theta_2$$

$$x_{mut1}, j = x_{Teacher}, j, \text{ Else}$$

$$X_{mut,3} = \varphi \times X_{Teacher} + \rho \times (X_{Teacher} - X_{muted})$$

Where

$$\theta_1, \theta_2, \theta_3 \text{ and } \rho \text{ are random values in the range of } (0,1).$$

III. RESULT AND DISCUSSION

The main concept of this paper is to show the better accuracy among three machine learning algorithms SOM, HWNN, MTBO. By comparing all three algorithms for the imbalanced dataset. Balanced dataset has no difficult in showing better accuracy whereas imbalanced data with large number of data which never take lots of time and it never gives better accuracy.

Method	Accuracy in %
SOM	98.6
HWNN	96.5
MTLOB	99.5

IV. CONCLUSION AND FUTUREWORK

Inside the structure of the present examination, the utilization of promotion vanned machine learning methods dependent on HWNNs, SOMs and MTLBO towards creating CVD hazard scores for the T2DM populace was investigated. The finish of the paper is the patients with long termT2DM experience the ill effects of CVD here with less parameter the sickness is anticipated with the better exactness. It got outcomes demonstrate that the algorithm calculations HWNN, SOM & MTLBO a performs well regardless of whether the dataset utilized incorporates few CVD. The acquaintance of large datasets relating with patients with various ethnicity and race will broaden the material of the model to different companions of patients.



Mrs. K. Blessing Christiana, has received the Master’s degree in Information Technology (2012) from Hindustan Institute of Technology and Science, Chennai, India. She is currently working as Assistant Professor in Department of Information Technology at New Prince Shri Bhavani College of Engineering and Technology, Chennai from 2013. Her research interest includes Artificial Intelligence, Cryptography and Networks.

REFERENCE

1. K. Zarkogianni, V. G. Karamanos, A. C. Thanopoulou, and K. S. Nikita, “A hybrid genetic algorithm for the selection of the critical features for risk prediction of cardiovascular complications in Type 2 Diabetes patients,” presented at the *IEEE 13th Int. Conf. Bioinformatics Bioengineering*, Nov. 2013, pp. 1–4.
2. P. Brindle, A. Beswick, T. Fahey, and S. Ebrahim, “Accuracy and impact of risk assessment in the primary prevention of cardiovascular disease: A systematic review,” *Heart*, vol. 92, pp. 1752–1759, 2006.
3. Sai Buddi , Thomas Taylor, Chad Borges, Randall Nelson “SVM Multi- classification of T2D/CVD Patients Using Biomarker Features” presented at the IEEE 10th int . cof machine learning and application, 2011 ,pp .338 - 341
4. Berina Ali , Lejla Gurbeta , Almir Badnjevi , “Machine Learning Techniques for Classification of Diabetes and Cardiovascular Diseases”, presented at the IEEE 6th editerranean conf . Embedded computing , 2017, pp
5. V. Almendra and D. En’achescu, “Using self-organizing maps for binary classification with highly imbalanced datasets,” *Int. J. Numerical Anal. Modelling.*, vol. 5, no. 3, pp. 238–254, 2014.
6. M. Matheny, “Systematic review of cardiovascular disease risk assessment tools,” Evidence Synthesis no. 85, AHRQ Publication 11-05155-EF-1, Rockville, MD, USA: Agency Healthcare Res. Qual., May 2011.
7. K. Valavanis, S. Mougiakakou, K. Grimaldi, and K. Nikita, “A multifac- torial analysis of obesity as CVD risk factor: Use of neural network based methods in a nutrigenetics context,” *BMC Bioinformat.*, vol. 11, no. 453, pp. 1–10, Sep. 2010.
8. M. Skevofilakas, K. Zarkogianni, B. G. Karamanos, K. S. Nikita, “A hybrid Decision Support System for the Risk Assessment of retinopathy development as a long term complication of Type 1 Diabetes Mellitus,” *32nd International Conference of the IEEE Engineering in Medicine and Biology Society*, Sep. 2010.

AUTHORS PROFILE



Mrs K. Rajathi, M.Tech., CSE, currently working as an Associate Professor and a Research Scholar in Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology. She worked as an assistant professor in various Engineering colleges. She has received her Master’s degree in Bharathidasan

University in the year 2008. Her area of interest is environment perception for autonomous vehicle, Lidar Data processing and self-organized map.



Mrs. Asmetha Jeyarani .R., presently working as an Assistant Professor in the department of CSE at Veltech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. She also worked as an Asst. prof. in the department of CSE at Shivani Engineering College, Trichy before joining this institution. She has received the Master’s degree in Information Technology (2012) from Hindustan Institute of Technology and Science, Chennai, India. Her Field of research includes Cryptography and Network security, Internet of Things. She has published many international journal papers.