Journal of Theoretical and Applied Information Technology
15th February 2018. Vol.96. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN ELM PREDICTIVE MODEL FOR RISK ASSESSMENTS OF CVD IN IMPAIRED GLUCOSE TOLERANCE (IGT) PATIENTS VIA GENPCNN AND SLFNS ALGORITHM

[1]S.RAJESWARI, [2]DR.M.S.JOSEPHINE, [3]DR.V.JEYABALARAJA

[1] Research Scholar, Bharathiyar University,Coimbatore- INDIA

[2] Professor, Department of MCA, Dr.M.G.R Educational & Research Institute,Chennai - INDIA

[3] Professor, Department of Computer Science & Engineering, Velammal Engineering College,Chennai – INDIA

E-mail: [1]vrajee2008@gmail.com, [2]josejbr@yahoo.com,[3]jeyabalaraja@gmail.com

ABSTRACT

Medical diagnosis systems play a vital role in medical practice and are used for diagnosis and treatment by several medical practitioners. Diagnosing the risk factors of pre-diabetic (IGT) cases is quite difficult. There is a big challenge to improve the diagnosis system to recognize the risks factors of impaired glucose tolerance regarding to cardiac vascular disease. In this paper, ELM classifier is combined with the hybrid of genetic algorithm and pulse coupled neural network (GENPCNN). Especially, a Single-hidden layer feed forward neural networks are suitable for solving the complex classification problem. The datasets we collected from health care centre having 270 instances of pre-diabetic, Diabetic and non-diabetic data each was having 28 attributes. A combination of genetic algorithm based neural networks to select the features from the dataset. So, it will be reduced to 14 attributes. The best population of the GA will be passed as input for the PCNN. The features extracted from the GENPCNN are passed to ELM classifier SLFNs in which the hidden nodes are chosen randomly and logically determines the output weight. First, dataset is preprocessed in order to remove the noisy data, missing values or irrelevant values and also from 'curse of dimensionality' which have to make suitable for training. This algorithm tends to provide good generation performance and extremely fast learning speed. The classification accuracy obtained using this approach is 94%. The obtained results have shown very promising outcomes for the prediction of risk factors of CVD in impaired glucose tolerance and impaired fasting glucose.

Keywords - *Pulse Coupled Neural Network, Genetic Algorithm, Dysglycemia, ELM, Impaired Glucose Tolerance*

1. INTRODUCTION

Data mining is the process to discover unknown patterns or relations that provide a clear and valuable result of selecting, analyzing and modeling significant amounts of data in order and recently this technique has developed rapidly [10]. Studies have applied data mining to explore the factors which are unknown and predictive models have been built in the medical field [2,4]. Pre-Diabetics is the precursor stage to diabetes mellitus in which the required symptoms are not all included to label a person as diabetic, but blood sugar is abnormally high. This stage is often referred to as a grey area. Impaired Fasting Glycemia or Impaired Fasting Glucose (IFG) passes to a condition in which the fasting blood glucose is high above to be considered normal levels but is not high enough to be classified as diabetic mellitus, it is considered as a pre-diabetic state related with insulin resistance and increased possibility of cardio-vascular pathology [6]. Impaired glucose tolerance is a predictable state of dysglycemia, which is associated with insulin deficiency and increased risk of cardio vascular pathology.

IGT may precede Type 2 diabetes mellitus by many years [7].Pre-diabetes typically has no distinct signs or symptoms. Some of the causes are Family history of diabetes, Hypertension, Increase triglyceride level, Low levels of good Cholesterol (HDL), Overweight or Obesity, Elevated blood pressure and elevated fasting plasma glucose. Women who have had gestational diabetes has high birth weight babies [3]. Due to the abnormal elevation of blood glucose levels, leads to the diabetes which may cause damage to large and small blood vessels leading to cardiac vascular disease,

neuropathy, retinopathy and nephropathy the prolonged these are associated with insufficient insulin resistance and are risk factors for the development of type 2 diabetes mellitus. These in tries stratum (IGT or IFG) are at increased risk of CVD of the two, impaired glucose tolerance better predicts CVD and mortality.

Effect of the disease may affect larger blood vessels (atherosclerosis) or small blood vessels as seen damage to the retina of the eye, damage to the kidney and damage to the nerves [8].

1.1 Motivation and Justification of the proposed work :

The main objectives of this work are tc managing the pre-diabetic to CVD through the following multifold focus.

1. Detecting the early stages of diabetic using GTT.
2. Identification of risk factors regarding to pre-diabetic and complexities of cardiac vascular disease.
3. Predicting the cardiac vascular disease form the risk of pre-diabetic patients.

As more and more data are getting together, the processing and interpretation of data become more crucial order to turn in to knowledge towards the decision support system for pre-diabetic subjects in connection with CVD. The basic issues in data mining are to reducing the dimension and handling missing and noisy data. The motivation in this paper is to selecting the prime attributes and extracting those attributes; elimination of 'curse of dimensionality' problem and also to predict the risk factors of CVD from the IGT patients. The proposed model provides good results by predicting the risk factors accurately.

1.2 Outline of Proposed Work

The proposed model consists of three challenging methodologies. The first methodology contains the Attribute Selection using GENPCNN.The second methodology consists of Attribute Extraction and final one is the ELM prediction with SLFNs. It is shown in Fig.1. The paper is organized as follows; Related work is dicussed in Section 2. Clinical Dataset is given section 3. Attribute Selection using GENPCNN is discussed in Section 4. Classifiers are discussed in Section 5. Section 6 presents the Experiments and Performance Evaluation. Results & Discussions is given in section 7 and Conclusion is given in Section 8.

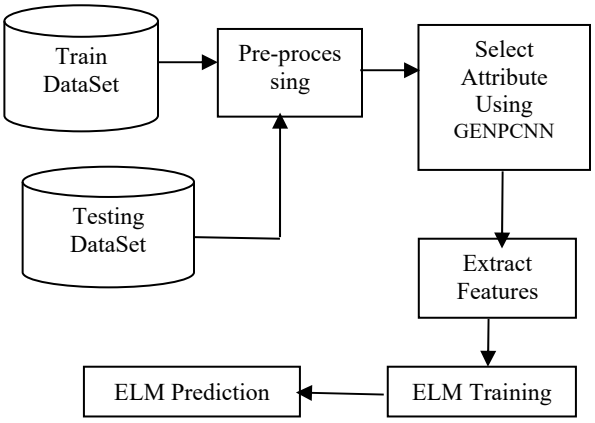


Figure.1. System Architecture of Proposed Model

2. RELATED WORK

The techniques [22] such as clustering is combined with K-means, Gaussians and self-organizing maps The techniques such as clustering is combined with K-means, Gaussians and self-organizing maps (SOM) and Neural Gas(NG) have been used for the diagnoses of T2DM. While some of the classification techniques such as Support Vector Machine(SVM), Neural Networks such as Multilayer, back-propagation, radial basis function and neuro fuzzy inference systems are used to classify diabetic or non-diabetics [8,5].

Optimization techniques such as Genetic Algorithm developed by John H.Holland by generating the effective chromosomes from the population of Chromosomes in optimized one[17]. According to the phenomenon of synchronous pulse burst in cat's visual cortex, Eckhorn introduce the linking field network[9]. In [18], promotes high dimension data processing capacity of PCNN model, which has more neuron structure patterns and pulse transmission properties for Image processing.

In [19], generating a data reduction method, original training instances are mapped into some pulse coupled neurons and a firing algorithm is presented for determining which instances locate in border regions and filtering instances. PCNN is unique from all other techniques due to its synchronous pulsed output, adjustable threshold and controllable parameters are hence used in image processing [15]. In [20], used Cascaded Neural Network (CNN) and SVM with radial basis function to analyze the hidden units for predicting the CVD.

In [21], a hybrid system of SVM and Random Forest(RF) of Ensemble learning approach for generating the rules for screening the diabetes. An

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Ensemble classifier with SVM and SVMJ for Decision Support system and also reducing the missing data and imbalance data [1].

3. CLINICAL DATASETS

The dataset used in this study was derived from the public health care centre's, it was about 270 patients. It divides the corresponding datasets into DM, Cardiovascular Disease (CVD) and Hypertension (HT) as a chord of diseases, which are also identified in the dataset. The collected datasets contains both clinical variables as well as anthropometrics variables. The clinical variables include GTT, Fasting, Post prandial Blood Sugar(Pbs), HbA1c, cholesterol profile: LDL, HDL, triglyceride (TG) concentrations in blood; inflammatory and oxidative stress markers from urine; peripheral vascular function parameters; blood pressure (BP) measurements, both systolic (SBP) and diastolic (DBP); anthropometrics, including Height, weight, Body Mass Index(BMI), Smoking, Alcohol consumption Diet, Work stress in the medical history. Data on 28 attributes from 270 patients have been collected from different health care centers. The temporal data underwent compression to the instance of patient classes as an alternative to attendances by calculating longitudinal means on the amount of available data.

3.1 Glucose Tolerance Test

According to the American Diabetes Association 2016 (ADA) [11,12] Screening for pre-diabetes can be done using FPG or 2-hr PG after 75-g OGTT criteria. The glucose tolerance test is a medical test in which the glucose were given orally and blood samples are taken to determine the pre-diabetic risk, insulin resistance, reactive hyperglycemia or rarer disorders of carbohydrate metabolism.

Categories of Increased Risk for Pre-diabetes FPG 2-hr PG A1C 100-125 mg/dL (5.6-6.9 mmol/L),IFG- 140-199 mg/dL (7.8-11.0 mmol/L) Impaired glucose tolerance (IGT) 5.7-6.4% (39-46 mmol/mol). The glucose tolerance test is a medical test in which glucose is given directly and samples of blood were taken to find out how fast it is cleared from the blood. Basically, the test is used to test for dysglycemia, insulin resistance, impaired function of beta cell and sometimes reactive hypoglycemia or rarer disorders of metabolism in carbohydrate. In the most frequently performed version of the test, an oral glucose tolerance test (OGTT), a standard dose of glucose is taken by patient and blood levels are checked after two hours. A considerable number of variation of the GTT have been designed over the years for various purposes, with changes in standard doses of glucose, various routes of administration, intervals and durations of sampling, and various substances measured in addition to sample blood glucose.CVD risk factors should be identified and treated. Glucose Tolerance test is only for finding the pre-diabetics cases or the risk of type 2 diabetics.

4. ATTRIBUTE SELECTION USING GENPCNN

It is a hybrid model of genetic algorithm and pulse coupled neural network. A novel method for selecting the best attribute by using the GENPCNN method. Automated medical diagnostic approaches are mainly based on machine learning (ML) algorithm. A enormous number of attributes that can surpass the number of data by themselves often characterizes the data used in ML. The problem "curse of dimensionality" which generates a great challenge for various ML applications for decision support system. The above said problems will leads to risk of redundant attributes may cause lower classification accuracy. Therefore, the process of eliminating irrelevant attribute is a vital phase for designing a decision support system with high accuracy [22,23].

The proposed model is a new hybrid feature selection method by combining Genetic and Pulse coupled neural network (PCNN) approaches. The weight value which is used in PCNN is obtained by using the genetic algorithm.

4.1 Genetic Algorithm

Genetic Algorithm widens the ideas of natural evolution called Darwinian selection. It denotes the 'survival of the fittest '. The concept of the genetic algorithm is the population of candidate solution (called phenotypes) to an optimization problem for better solution. The above said candidate solution contains the set of individuals called chromosomes or genotype. This can be mutated or altered to find optimized way. The binary values of strings 0s and 1s are used represent the solutions. The evolution will begins with the randomly generated individuals and in each generation the fitness value of the every individual is calculated. Based on their fitness, multiple individuals are stochastically selected from the recent population and mutation will takes place to form a new population. Then the new population is

Journal of Theoretical and Applied Information Technology
15th February 2018. Vol.96. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

used for the next iteration of the algorithm. Whenever a maximum number of generations or a satisfactory fitness level can be reached for the given population the algorithm terminates. If the algorithm terminates due to the maximum number of iterations, an optimized solution may or may not be reached [17]. By using the genetic representation, the fitness function is defined and measures the quality of the represented solution. The fitness function always problem oriented. After generating the fitness and the genetic representation, the GA proceeds by repetitive application of the mutation, crossover, inversion and selection operator. In genetic programming a tree-like representations were explored and graph –form representations are explores in evolutionary programming.

4.2 Pulse Coupled Neural Network

Pulse coupled Neural Network is a self-organizing network model that does not require training and the network is simulating the activity of mammal's visual cortex neurons. The PCNN could be a single layer; two dimensional, network which laterally connected to integrate and fire neurons with 1:1 correspondence between the data and network neurons. The structure of PCNN neuron is shown in Fig. 2. The neuron consists of an input part, linking part and pulse generator. The feeding and linking part gives the input signal to the neuron. From the neuron receptive area, the primary input is the feeding input. The neuron receptive area consists of the neighbouring data of corresponding in the input part. The secondary input is linking input with lateral connections of neighboring neurons. The difference between these explores in evolutionary programming. Inputs from the feeding connections have a slower characteristic response time constant than the linking connections. The standard PCNN model is described by the following equations with different iterations [15, 16].

From the below formulations,

Sij is the input stimulus such as the best values from GA

$F_{ij}[n]$ – Feedback input of the neuron in (i,j)

$L_{ij}[n]$ – Linking item

$U_{ij}[n]$ – Internal activity of neurons.

$T_{ij}[n]$ – Dynamic Threshold.

$Y_{ij}[n]$ – Stands for the pulse o/p of neuron and it gets either the binary value o or 1. It is also said output updation.

According to the following equations, here α_F ,α_L, α_T are represented as a initial constant value of decay term for linking , feeding and thresholds.

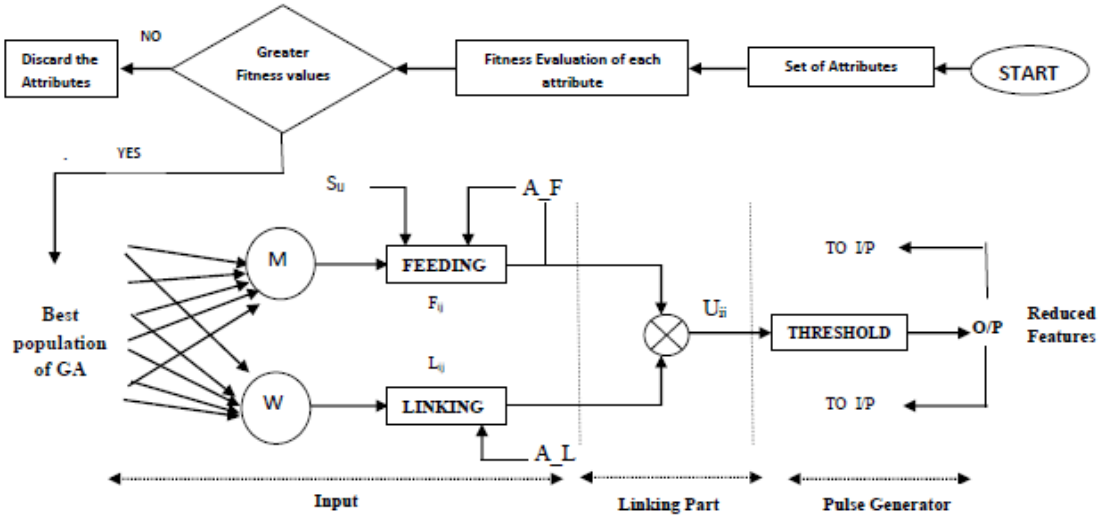
$$\mathbf{F_{ij}[n] = e^{-\alpha_F} Fij[n-1] + V_F \sum_{K,I} W_{i,j,K,I}\, Y_{ij}\,[n-1] + S_{ij}} \quad \mathbf{(1)}$$

$$\mathbf{L_{ij}[n] = e^{-\alpha_L} Lij[n-1] + V_L \ \sum_{K,I} M_{i,j,K,I}\, Y_{ij}\,[n-1]} \quad \mathbf{(2)}$$

$$\mathbf{U_{ij}[n] = F_{ij}[n](1 + \beta L_{ij}[n])} \quad \mathbf{(3)}$$

$$\mathbf{Y_{ij}\,[n] = \begin{cases} 1 & Uij[n] > Tij[n] \\ 0 & otherwise \end{cases}} \quad \mathbf{(4)}$$

$$\mathbf{T_{ij}[n] = e^{-\alpha_t} Tij[n-1] + V_T \ Yij[n-1]} \quad \mathbf{(5)}$$



Fig(2). Structure of GENPCNN Model

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Algorithm of GENPCNN:

Input: Attributes (X) and its values, Cross over Probability (CP), Mutation Probability (MP)
Output: Best attribute (S) & Fitness Value (S).

1. Generate random population from each Attribute and Consider as the Initial Population P.
2. Evaluate the fitness f(X) of each attribute (X) in the population (P) by using the below formula.

$$f(x) = \frac{1}{N}\sum_{i=1}^{N} x_i \quad (6)$$

Where X_i is ith attribute values in the attribute X and N is the total values in the attribute

3. Create a new population (NP) by repeating following steps until the new population (NP) is complete
4. Select attributes from a population P according to their fitness. If the fitness value is high, then those attributes as selected as best fitness value (W), otherwise rejected.
5. With a crossover probability (CP) cross over the selected attributes to form a new attribute. If no crossover was performed, new attribute is an exact copy of selected attributes.
6. With a mutation probability (MP) mutate selected attribute at each locus.
7. Place selected attribute in a new population (NP).
8. Use new generated population (NP) for a further run of algorithm.
9. If the end condition is satisfied, stop, and return the best solution(S) in current attribute.
10. Go to step 2.
11. Finally the output of the genetic algorithm is the fitness value (W) and the best attribute (S).
12. Give the best attribute (S) as the weight value of the PCNN and execute PCNN.
13. From the PCNN and return the best attribute (S) as the output.
    Finally the output of the genetic algorithm is the fitness value and the best population. It is given into the PCNN as the input.

PCNN INPUTS

W= [best population of Genetic Algorithm] } Output of genetic algorithm
M = [best population of Genetic Algorithm] }
Alpha_F = 0.1 Decay term for feeding
Alpha_T = 1.0 Decay term for threshold
V_L = 0.2 Magnitude scaling term for linking
Beta = 0.1 Linking strength
F = zeros(size(S)) Initial values for F
Y = F Initial values for Y
T = Ones(size(S)) Initial values for T
Alpha_L = 1.0 Decay term for linking
V_F = 0.5 Magnitude scaling of feeding
V_T = 20.0 Magnitude scaling of linking
Num = 40, the number of iterations
L = F Initial values for L
U = F Initial values for U
S = im2double(S) Normalizing to lie within [0,1]

for n = 1:num
F = exp(_Alpha_F) *F + V_F*conv2(Y,W) + S; Update the feeding input
L = exp(_Alpha_L) * L + V_L*conv2(Y,M); Update the linking input
U = F_ *(1 + Beta*L); Compute the internal activation
Y = double(UiT); output updation
T = exp(_Alpha_T) *T + V_T*Y ; Updating of threshold input
end
Here W is the genetic algorithms fitness value. And the S is the genetic algorithms best population

feature are calculated by using the below formula

4.3 Feature Extraction

In this module the feature are extracted from the best attributes. the first order feature such as mean and standard Deviation are used. The mean

$$M = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i)} \quad (7)$$

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Where xi is the value of each best attribute and N is the total no of patients.

The standard deviation formula is calculated by using the below formula.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \quad (8)$$

Where xi is the value of each best attribute values, μ is the mean value and N is the total no of patients. These features of all three transform are combined to form the single feature vector and then these features will trained the ELM.

5. CLASSIFIERS

Naives Bayes: The classification is based on Bayes theorem with independent assumptions. It represents the supervised learning method as well as the statistical method for classification. Bayesian classifiers can predict the class membership probabilities such as the tuple belongs to the particular class. It is based on posterior and prior probability.

Decision Tree: It is a predictive model that maps the observation about an item to conclusions about the target value. A tree-like structure where each non-leaf node represents a test case on an attribute, each branch represents the outcome of the test and each terminal node holds the class label[13,14].

K-NN: Nearest neighbor classifier is based on analogy learning, by comparing the test tuples with the training tuples that are similar to it. It is an instance-based classifier from the unknown features using the distance or similarity measures [13].

Ensemble: A fusion of multiple classifiers may learn a more expressive concept class than a single classifier.

5.1 ELM Prediction

In this module the extracted features are predicted using the Extreme Learning Machine (ELM). A new learning algorithm called Extreme Learning Machine (ELM) for single-hidden layer feedforward neural networks (SLFNs) which randomly select the hidden nodes and analytically conclude the output weights of SLFNs. This algorithm tends to provide good generalization performance at extremely fast learning speed.

ELM Algorithm

1. Generate the hidden nodes randomly as well as randomly assign the weights Wij for i= 1..m; j=i..n
2. Activate the input values Xi, i=1....n and calculate the net input to hidden layer using Hij= $\sum_{i=1}^{n} Xi\, Wij, j = 1 \ldots m$
3. From the hidden layer to output layer, Repeat the step 2 to calculate the net output

Here W_{ij} is the weight, which is given as random to train the machine and X_{ij} is the each best attribute which is combined with the random weights to train the ELM machine for the net output.

6. EXPERIMENTS AND PERFORMANCE EVALUATION

Experiments are conducted in order to evaluate the performance of the proposed system. We carried out the experiments for diagnosing the risk factors of pre-diabetic cases. Our main aim is to build an intelligent classifier that is capable of recognizing a new patient and assigning it to one of two classes: presence or no presence of heart disease in pre-diabetic patients. For this purpose, we experiment with well-known machine learning algorithms to train classifier and use it for classification. The dataset is collected from several public health care centre's. The dataset consists of 270 individual clinical records of which it is collectively having diabetic, non-diabetic and pre-diabetic individuals. These datasets having missing feature values and we have replaced these missing values with the average of all values for those specific features across all samples. The clinical dataset is represented with 28 features in which only 14 features are selected to implement our experiments which are shown in Table.1.

To compare the performance the ELM is compared with other four classifier technique such as KNN, Navie Bayes, Decision Tree and Ensemble classifier The classifier will show the diagnosis of the normal patient as pre-diabetic as well as the risk of dysglycemia with the threshold values in Table1.

6.1 Performance Evaluation

To evaluate the performance of the proposed method various performance metrics are available. Our proposed model uses the Precision Rate, Recall Rate, Sensitivity, Specificity and F-Measure to analyses the performance [10, 13].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Performance Measures:

1) ***True Positive Rate***: Number of positive instances that were labelled correctly by the classifier

2) ***True Negative Rate***: Number of Negative samples that were labelled correctly by the classifier

3) ***False Positive Rate***: Number of Negative samples that were incorrectly labelled by the classifier.

4) ***False Negative Rate***: Number of positive instances that were labelled incorrectly by the classifier.

5) ***Precision Rate***: The precision is the fraction of retrieved instances that are relevant to the find.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

where TP = True Positive (Equivalent with Hits), FP = False Positive (Equivalent with False Alarm)

6) ***Recall Rate***: The recall is the fraction of relevant instances that are retrieved according to the query.

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (10)$$

Where TP = True Positive (Equivalent with Hits), FP = False Negative Equivalent with Miss).

7) ***F-Measure***: F-measure is the ratio of product of precision and recall to the sum of precision and recall. The f-measure can be calculated as,

$$F_m = (1+\alpha) * \frac{\text{Precision}*\text{Recall}}{\alpha*(\text{Precision}*\text{Recall})} \quad (11)$$

8) ***Specificity***: Specificity measures the proportion of negatives which are correctly identified such as the percentage,

$$\text{Specificity} = \frac{\text{TN}}{(\text{FP+TN})} \quad (12)$$

9) ***Sensitivity***: Sensitivity also called the true positive rate or the recall rate in some fields, measures the proportion of actual positives.

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP+FN})} \quad (13)$$

Where, TP – True Positive (equivalent with hit), FN – False Negative (equivalent with miss).
where TN – True Negative (equivalent with correct rejection), FP – False Positive (equivalent with false alarm).

6.2 Performance Evaluation of Each Classifier

From the Table 2, the precision rate analysis of the proposed classifier and other existing classifier approaches are shown. The precision rate value of the proposed model is higher than the other existing model, which will be observed from the table 2.The graph of Fig 3. of precision rate of four existing models is compared with ELM model which shows the higher precision rate. The precision rate of the ELM classifier is 94%. According to the training dataset, the precision rate of the ELM classifier will be high for the testing dataset.

Table 2. Performance of Classifiers

Classifier	PRE (%)	REC (%)	SEN (%)	SPE (%)	F-MES (%)
ELM	94	97	95	97	97
DT	85	87	86	87	87
KNN	82	84	83	83	85
ENS	87	89	88	89	89
NB	79	81	80	81	80

PRE- precision, REC- Recall rate, SEN- Sensitivity SPE- Specificity, F-Mes- Fmeasure.

The performance comparison of the Recall Rate value of the proposed method and other four existing approaches such as Naives Bayes – 81%, K-NN – 84%, Decision Tree – 87% and Ensemble – 89% is shown in the Table 2. From the observation it is shown that the recall rate value of the ELM method is 97% which is higher than the other existing approaches. The relevant instances given by the ELM classifier according to the training datasets will be high when compared with above said classifier, because the retrieving capacity of the following classifiers will remain slow. The graph of recall rate analysis is shown in Figure 4.

The F-Measure analysis of the proposed method and other four existing approaches such as Naives Bayes, K-NN, Decision tree and Ensemble is shown in the Table.2. From the table it is shown that the F-Measure analysis value if the ELM method is 95% higher than the other existing approaches. As the Naives Bayes – 80%, KNN-83%, Decision Tree-86% and Ensemble-88% have the lower F-Measure analysis when compared to the proposed method. The Sensitivity analysis and It is analyzed by the precision rate and recall rate of training and testing datasets. The graph of

Journal of Theoretical and Applied Information Technology
15th February 2018. Vol.96. No 3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

F-Measure is shown in Fig.5.

The sensitivity and specificity analysis of the proposed model and other four existing classifier is shown above in the table.2. From the table it is shown that the proposed model ELM classifier is having high sensitivity and specificity rate. The graph of sensitivity & specificity analysis is shown in the Fig,6 & 7. From the Table.2 the sensitivity % of the following classifiers are the Naives Bayes – 81%, K-NN – 83%, Decision Tree – 87%, Ensemble – 85% and ELM – 97%. Sensitivity shows the true positive rates in the diagnosing system which shows the actual results from the patients having the CVD risks.

Specificity shows the proportion of negatives from the testing data which will be learned from the training datasets. When comparing with the existing classifier our proposed model , ELM -97% shows the high specificity rate for the detecting as well as diagnosing the risk factors of CVD from the Impaired Glucose Tolerance(IGT) or pre-diabetic patients or dysglycemic patients.

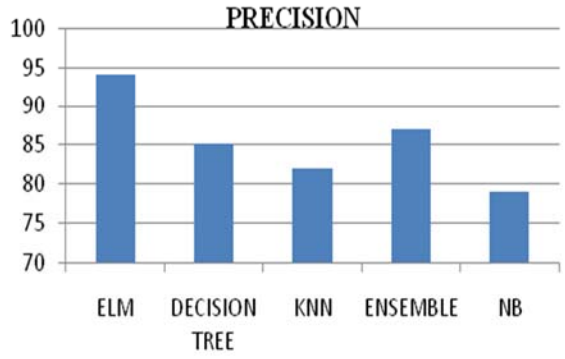


Figure. 3 Precision Rate

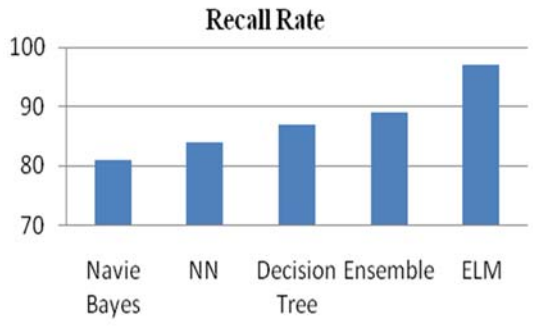


Figure.4 Recall Rate

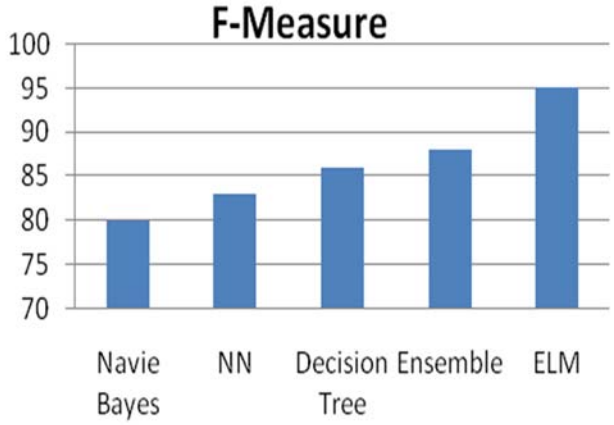


Figure.5 F-Measure

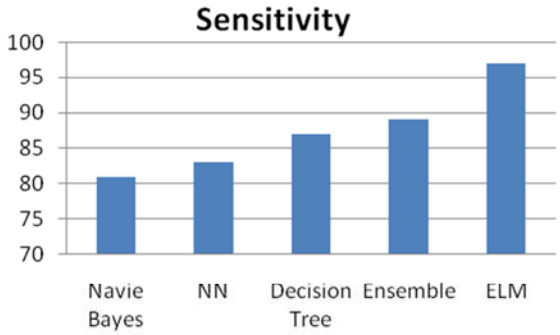


Figure. 6 Sensitivity

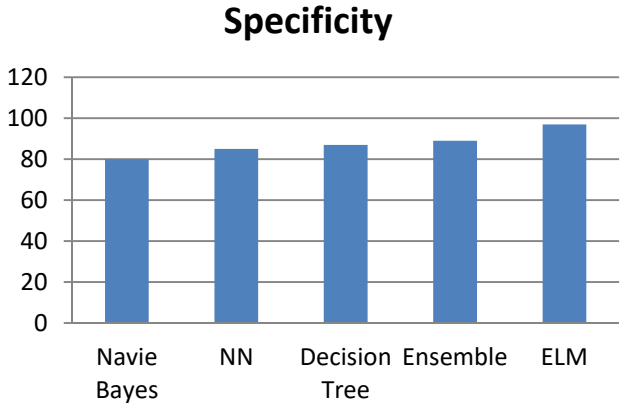


Figure. 7 Specificity

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

7. RESULTS & DISCUSSIONS

From the proposed classification ELM is compared with Naives Bayes, Neural Network, Decision Tree and Ensemble methods. The dataset collected from the health care centre contains 28 attributes which is reduced to 14 attributes by using GENPCNN algorithm, the extracted features will be given as training set to the ELM classifier.

As per the [17], when compared with the existing one the reduced attributes will not be optimized one and iterations will be increased and also individually compared with the existing PCNN [18,19] model the attributes will not be reduced more. Our objective function is to improve the classification accuracy for non-linear datasets, because the clinical datasets are non-linear.

When we use our proposed GENPCNN model, the output will be optimized with less number of iterations and the classification accuracy is gradually increased [22,23]. The highest percentage of correct classification obtained where 97%, 94% etc. Several risks factors are found from the available attributes which includes dysglycemic events. When the IGT≥140mg/dl with the standard deviation and mean as 98.5±0.7071, PBS≥200mg/dl with 127±7.071, HBA1C –(5.7-6.4%) with 6.3±0.8485. The BMI is calculated from the patients height and weight, the value should be >25 or greater with22.5±0.9899, Uric acid with 4.3±6.0811 these factors are extracted from the dataset with the corresponding values. Patients who are enter the dysglycemia will undergo the IGT & IFG test with pre-diabetic segment and having these criteria as TG, LDL-C, VLDL-C and HDL-C with 264±226.2742, 77±70.107, 62.8±59.3970, 43.2±8.4853, diastolic and systolic as 128±2.8234, 82.5±0.7071.

The classifier is trained with above risk factors. By these factors, ELM classifier will recognize the CVD risk factors. The proposed model will link some of the hidden nodes like smoke, Alcohol intake, Work Stress also taken in the ELM Machine to recognize the CVD disease from the pre-diabetic patients. From the above results, sensitivity and specificity of the ELM machine refers the tests ability to correctly detect the patients who are these risk conditions. The sensitivity shows the proportion of people who are having the test positive (TP) for the disease among those who have the CVD disease.

Table. 1 Mean And Standard Deviation For Factors of IGT

	Features	units	Mean ±Sdv
1	Imapired Glucose Tolerance (IGT)	>=140 mg/dl	144 ± 1.4142
2	Impaired Fasting Glucose (IFG)	>=126 mg/dl	98.5 ± 0.7071
3	Post Prandial Blood Sugar (PBS)	>=200 mg/dl	127 ± 7.0711
4	HBA1C	(5.7-6.4)%	6.3 ± 0.8485
5	Height (Ht)		157 ± 1.4142
6	Weight(Wt)		55.50 ± 3.5355
7	Body Mass Index (BMI)	>25 or greater	22.5 ± 0.9899
8	Uric Acid (UA)		4.3 ± 6.0811
9	Triglyceride (TG)	>=150	264 ± 226.2742
10	Low Density Lipoprotein (LDL-C)	>=100	77 ± 70.7107
11	Very Low Density Lipoprotein (VLDL)	>=200	62.8 ± 59.3970
12	High Density Lipoprotein (HDL-C)	<40 or <50	43.2 ± 8.4853
13	Systolic Blood Pressure	<140	128 ± 2.8284
14	Diastolic Blood Pressure	<90	82.5 ± 0.7071

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

8. CONCLUSION & FUTTURE ENHANCEMENT

Lifestyle changes, insufficient diet control and a lack of physical fitness leads to the irregular glucose metabolism. The epidemic of pre-diabetic to CVD has caused huge loses and caught the attention of society. It is imperative to predict the risk factors of dysglycemic events individually in its early phase and the appropriate intervention to prevent the occurrence of acute events. One of the major challenges is prediction of CVD factors from the Pre-diabetic patients without knowing how the patients' blood glucose level concentration will respond.

Being able to predict the blood glucose level would simplify the management. Therefore the early identification of risk factors of CVD in pre-diabetic patients with a clear accuracy will be a challenging one. In this paper, we developed an Extreme Learning Machine for diagnosing the risk factors of CVD in Impaired Glucose Tolerance (IGT). In particular, we utilized GENPCNN for reducing the dimension for 'Curse Of Dimensionality' problem and ELM SLFNs classifier for diagnosing the dysglycemic risk factors.

By training, the ELM classifier with the following conditions will provide a second opinion for diagnosis and a tool to screen the individuals with undiagnosed diabetes by lay users. This will provide an enhanced opportunity for timely and appropriate intervention to apply, which may reduce the incidence of pre-diabetics and its complications. Results show that our proposed model has high quality in terms of diagnosis with precision, which points out the diagnosis ability of the model.

REFERENCES

[1] Maciej Zieba, "Service-Oriented Medical System for supporting Decisions with missing and Imabalanced data", *IEEE Journal of Biomedical and Health Informatics*, vol.18(5),Sept 2014, pp-1533-40.

[2] Bellazzi R, Zupan B. "Predictive data mining in clinical medicine:current issues and guidelines" *Int J Med Inform* ;77:81-97.

[3] Chen HY, Chuang CH, Yang YJ, Wu TP. "Exploring the riskfactors of preterm birth using data mining" .*Expert Syst Appl* ;38:5384-7, 2011.

[4] Chang CD, Wang CC, Jiang BC. Using data mining techniques for multi-diseases prediction modelling of hypertension and hyperlipidemia by common risk factors. *Expert Syst Appl*;38, pp.5507-13, 2011.

[5] Aslan K, Bozdemir H, Sahin C, Ogulata SN. "Can neural networkable to estimate the prognosis of epilepsy patients according to risk factors?" *J Med Syst* ;34, pp.541-50, 2010.

[6] M. Borgman and E. McErlean, "What is the metabolic syndrome?:Prediabetes and cardiovascular risk," *J.Cardiovasc. Nurs*., vol.21,pp. 285–290, Feb. 2006.

[7] T. A. Samuels et al., "Delayed diagnosis of incident type 2 diabetes mellitus in the ARIC study," Amer. *J. Manag. Care,* vol. 12, no. 12,pp. 717–724, Dec. 2006.

[8] D. O. Shankaracharya et al., " Computational intelligence in early diabetes diagnosis: A review," *Rev. Diabetic Stud.,* vol. 7, no. 4,pp.252–262, Feb. 2010.

[9] R.Echorn, H.J.Reitboeck et al., "Feature linking via synchronization among distributed assemblies: simulation of results from cat cortex", *Neural Computation*, vol.2 .293-307, 1990

[10] P.-N. Tan, Introduction toDataMining. Reading, *MA:Addison-Wesley*,2006.

[11] ADA American Diabetes Association 2016

[12] https://en.wikipedia.org/wiki/ Glucose tolerance_test

[13] Jiawei Han and Micheline Kamber, Data mining concepts and techniques , 2nd Edition.

[14] http://en.wikipedia.org/wiki/Decision_ tree_learning

[15] M . Monica Subashini et al " Pulse coupled Neural Network and its applications" Expert systems with Applications(41), Elsevier, 2014.

[16] Zhaobin Wang et al "Review of pulse coupled neural networks" Image and Vision Computing (28)(5-3), Elsevier, 2010.

[17] David E.Goldberg and John H.Holland, "Genetic Algorithms and Machine Learning", Machine Learning, vol.3,pp.95-99, October 1988.

[18] Luping Ji, L.Shang, G.Liu, " On High dimension- Pulse Coupled Neral Network", machineilab.org.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[19] Yang Sheng song, Zhang Yi et al., "Spatial point- data reduction using pulse coupled neural network", Neural process let(32), pp. 11-29, May 2010.

[20] R.Chitra, Dr.V.Sreenivasagam, "Heart Disease prediction system using supervised learning classifier", Bonfring International Journal of Software Engineering and Soft Computing,vol.32(1) pp.1-5, March 2013.

[21] Longfei Han et al., "Rule Extraction form support vector machines using Ensemble Learning Approach:An Application for Diagnosis of Diabetes" IEEE Journal of Biomedical and Health Informatics, Vol.19(2), pp.728-734, March 2015.

[22] Paolo Melillo et al., "Classification Tree Risk Assessment in patients Suffering from Congestive Heart failure via Long-term Heart Rate Availability", IEEE Journal of Biomedical and Health Informatics,vol.17(3) pp.727-733, May 2013.

[23] Yifeng li, Alioune Ngom, "Nonnegative Least Square Methods for the classification of High-Dimensional Biological Data", IEEE/ACM Transactions on Computational Biology and bioinformatics, vol.10 (2), March2013.