Early Gestational Diabetes Detection Using Neural Network

TANZINA RAHMAN HERA[1], MD. ASHIKUR RAHMAN KHAN[1], NISHU NATH[1,*]
Department of Information and Communication Engineering
Noakhali Science and Technology University
Noakhali-3814,
BANGLADESH

Abstract: - Gestational Diabetes Mellitus (GDM) is defined as any degree of glucose intolerance with onset or first recognition during pregnancy. Fifty percent of GDM patients develop type 2 Diabetes in next twenty years and as well as the newborn can also be affected by diabetes in their lifetime. So the long term complications for both the mother and the child cannot be ignored. In view of maternal morbidity and mortality as well as fetal complications, early diagnosis is an utmost necessity in the present scenario. In developing country like Bangladesh, early detection and prevention is not cost effective and usually troublesome. So, there is an urgent need for a well-designed method for the detection of gestational diabetes mellitus. The purpose of this study is to predict the GDM in the first trimester. This research presents and compares some Artificial Neural Network (ANN) models on the early detection of Gestational diabetes mellitus and chooses the best neural network model among them to detect GDM early.

1 Introduction

Gestational diabetes mellitus (GDM) is a condition which is quite separate from the other types of diabetes: type 1 and type 2. The term gestational refers to it occurring during pregnancy. For many women who are diagnosed, the diabetes will go away after their baby is born. However, there is a greater risk of developing type 2 diabetes for women who have already had gestational diabetes. Gestational diabetes occurs in up to 25% of all pregnancies. At around the 20th week of gestation, the usual processes and actions involved in insulin production become affected by pregnancy hormones. Gestational diabetes screening tests are done routinely for all women who are pregnant, whether they have a history or not. The most common time for it to occur is between weeks 24-28 of pregnancy though it can be detected earlier too. Gestational diabetes generally results with few symptoms; however, it does increase the risk of pre-eclampsia, depression, and requiring a Caesarean section. Babies born to mothers with poorly treated gestational diabetes are at increased risk of being too large, having low blood sugar after birth, and jaundice. If untreated, it can also result in a stillbirth. Long term, children are at higher risk of being overweight and developing type 2 diabetes. Gestational diabetes is caused by not enough insulin in the setting of insulin resistance. Risk factors include being overweight, previously having gestational diabetes, a family history of type 2 diabetes, and having polycystic ovarian syndrome (PCOS). According to the American Diabetes Association (ADA), gestational diabetes affects 18 percent of pregnant women. It affects 1% of those under the age of 20 and 13% of those over the age of 44 [1]. A number of ethnic groups including Asians, American Indians, Indigenous Australians, and Pacific Islanders are at higher risk.

Diagnosis of GDM is by blood tests. Blood sugar and weight of a woman are the biggest telltales of diabetes. Blood pressure is also one of the causes for the disease. Gestational diabetes symptoms can be subtle or even nonexistent and some can be mistaken for typical side effects of pregnancy such as, Blurred vision, tingling or numbness in the hands and/or feet, excessive thirst, frequent urination, sores that heal slowly, excessive fatigue. There are two subtypes of gestational diabetes namely Type A1 and Type A1 [2].

A number of screening and diagnostic tests has been used to look for high levels of glucose in plasma or serum in defined circumstances: Non-challenge blood glucose tests, screening glucose challenge test, and oral glucose tolerance test .The fact of diagnosis process of detecting Gestational diabetes Mellitus (GDM) [3] is comparatively time consuming and it is very laboured job for a pregnant women to go to the

hospital and do all the tests and wait for getting the test report to know whether she has diabetes or no. For this reason more research is needed to find the most effective way of screening for gestational diabetes. The need for an accurate predictor for the gestational diabetes is highly needed. Not only this, but also a predictor that is extremely automated and with less human interference. A diabetic predictor should meet the following specification; efficient modeling, applicability and accuracy and be trusted. Therefore, there is an urgent need to adopt cost-effective screening and diagnostic methods of gestational diabetes to reduce maternal and fatal complications. This research includes the method of detecting GDM by using various Artificial Neural Network models like Multilayer perceptron neural network (MLPNN), Radial Basis Function Neural Network (RBFNN), Self-Organizing Feature Map (SOM) Neural Network, Support Vector Machine (SVM) Network, Fuzzy Logic and exploring the most efficient model among them.

Our contribution in this paper is as follows: We develop a cost-effective screening and diagnostic method of detecting gestational diabetes mellitus (GDM) to reduce maternal and fetal complications. We study the performance of different Artificial Neural Network models by applying the same dataset of the patients. We conduct comparison among those ANN structures on the basis of finding the best network that shows the more accurate result and explore the efficient and better ANN structure for predicting the GDM.

2 Related Work

Few research works accomplished in order to prediction diabetes are discussed here Anthropometrical Body surface scanning data was used to construct a classification model for diabetes type II using decision tree, artificial neural network, logistic regression and rough sets [4]. A research that attempted to enhance the detection of diabetes based on set of attributes collected from the patients developed a mathematical model using Multigene Symbolic Regression Genetic Programming technique [5]. Genetic Programming (GP) showed significant advantages on evolving nonlinear model which can be used for prediction. The developed GP mathematical model was developed to provide a solution to the diabetic problem and to classify patient type. These evaluation criterions proved that Multigene GP is beneficial for diabetic patient classification. They showed 73% accuracy in training and 86% in testing phase. A work that focused on early detection of GDM for women who are pregnant

for the second time onwards (multigravida patients) was proposed and was decided to diagnosis by artificial neural network as the increasing demand of Artificial Neural Network applications for predicting the disease shows better performance in the field of medical decision-making [6]. A model was developed for the prediction of GDM from maternal characteristics and biochemical markers at 11 to 13 weeks' gestation [7]. It was stated that by developing a simple prognostic model using age and BMI at booking could be used for selective screening of GDM in Vietnam and in other low- and middle-income settings [8]. Fasting glucose and insulin were measured in the first trimester and the homeostasis model assessment-insulin resistance index (HOMA-IR) was calculated for each patient aiming to predict GDM in the first trimester [9]. The study included 271 patients who were between the 10th and 14th week of gestation. Fasting glucose and insulin were measured in the first trimester and the homeostasis model assessment–insulin resistance index (HOMA-IR) was calculated for each patient. These values were compared with the results of the second-trimester glucose tolerance test results. HOMA-IR values were higher in women with gestational diabetes. A semi parametric generalized linear mixed model (GLMM) method was proposed to detect gestational diabetes [10]. This method was applied in evaluating the impact of covariates on the accuracy of diagnostic test results by way of obtaining a common cut off value for screening 50g glucose challenge test (GCT) for the three trimesters of pregnancy.

A paper used fuzzy integral to structure the diagnostic model of gestational diabetes mellitus [11]. As the neural network is easy to get into local optimum, the algorithm of simulated annealing was used to optimize the neural network to obtain an approximate global optimal solution. A comparison of the diagnostic performances of 75g and 100g Oral Glucose Tolerance Tests in detecting GDM in Nigerian pregnant women and was conc1uded by that 100g OGTT criterion was more stringent than that of75g OGTT in identifying GDM [12]. A developed a new methodology on Gestational diabetes prediction was developed by using Case Based Fuzzy Cognitive Maps (CBFCM) decision support system [13]. One significant advantage of the proposed CBFCM-based decision-making system over other approaches, such as the Bayesian networks was that it resembles human decision-making, with its capacity for approximate reasoning and handling incomplete information. A review was conducted on European peer-reviewed literature [14], supplemented by other sources of information,

relating to the prevalence of gestational diabetes and current screening practices and barriers to screening. GA was used to initialize and optimize the connection weights of BPN to classify diabetes mellitus [15].

An application of automatic multilayer perceptron (AutoMLP) was developed which was combined with an outlier detection method Enhanced Class Outlier Detection using distance based algorithm to create a novel prediction framework for classifying DM [16]. A study was performed to verify the correlation between abdominal subcutaneous fat thickness (ASFT) measured by ultrasonography during the first trimester of pregnancy and gestational diabetes mellitus (GDM) of the second trimester in Korean women and to establish a standard of ASFT for predicting GDM [17]. A study was performed among 250 pregnant women aged 15-44 years in their first and second trimester attending antenatal clinics in Ebonyi State University Teaching Hospital, Mile Four Maternity Hospital and Federal Medical Center Abakaliki, Nigeria were seen within the period of June 2010 to December 2011 [18]. Their age, parity, body mass index, gestational age and family history of diabetes were taken, while their gestational diabetes mellitus was assessed using 100g oral glucose tolerance test (OGTT).

Parastoo Rahimloo and Ahmad Jafarian performed a study to predict Diabetes by Using Artificial Neural Network, Logistic Regression Statistical Model and Combination of them [20]. In that research, the criteria were the performance to minimize the error function in neural network training using a neural network in a hybrid model which eventually came to the conclusion that the error function of the neural network was equal to 0.1 and combined neural network model was equal to 0.0002. A model was built using the hidden layers of a deep neural network to predict diabetes by Huaping Zhou, Raushan Myrzashova and Rui Zheng [21]. We tuned a number of parameters and used the binary cross-entropy loss function, which obtained a deep neural network prediction model with high accuracy. The experimental results show the effectiveness and adequacy of the proposed DLPD (Deep Learning for Predicting Diabetes) model. The best training accuracy of the diabetes type data set is 94.02174%, and the training accuracy of the Pima Indians diabetes data set is 99.4112%. In a paper Saumendra Kumar Mohapatra, Jagjit Kumar Swain and Mihir Narayan Mohanty tried to detect diabetes using multilayer perceptron neural network [22]. The case study was of Indian ladies with pregnancy suffered from diabetes. Data considered from PIMA database from UCI repository were used. Eight attributes were taken as features for each subject. The common MLP classifier was utilized for attributes and the experiment was learned with R studio platform. The performance found to be better as compared to earlier methods and verified in MATLAB platform as well. Nahla H Barakat, Andrew P Bradley, Mohamed Nabil H Barakat used several data mining and machine learning methods to diagnosis, prognosis diabetes [23]. They proposed utilizing support vector machines (SVMs) for the diagnosis of diabetes. Results on a real-life diabetes dataset showed that intelligible SVMs provide a promising tool for the prediction of diabetes, where a comprehensible ruleset had been generated, with prediction accuracy of 94%, sensitivity of 93%, and specificity of 94%. Md. Maniruzzaman, Md. Jahanur Rahman, Md. Al-Mehedi Hasan demonstrated that by replacing the missing values and outliers by group median and median values on dataset respectively and further using the combination of random forest feature selection and random forest classification technique yielded an accuracy, sensitivity, specificity, positive predictive value, negative predictive value and area under the curve as: 92.26%, 95.96%, 79.72%, 91.14%, 91.20% and 0.93 respectively [24]. The system was validated for its stability and reliability. RF-based model showed the best performance when outliers are replaced by median values.

3 Methodology

A model of detecting the GDM with high accuracy, less complex and has efficient performance is urgently needed. To address this need, the method identified in this study offers every pregnant woman the opportunity to know her risk early. Artificial Neural Network (ANN) model is becoming more popular in detecting various diseases so that the risk of the specific disease decreases. The early detection or classification of any disease helps anyone to decrease the risk of the disease. When c1assification is the goal, the artificial neural network (ANN) models often deliver close to the best fit. The present work is motivated in this direction.

This research of detecting Gestational diabetes Mellitus (GDM) is summarized by the following Fig. 1 given below. It starts with the collection of data like collection of the patient detail. The data of the patients is then is analyzed. Data should be normalized so that we can get better result. Then the input parameters should be normalized. After preparing the training dataset the different

artificial neural network will be applied. This study uses six neural network structures. The obtained result from them is then compared. The network that gives the more accurate result will be chosen. And GDM detection with the selected network configuration will be accomplished.
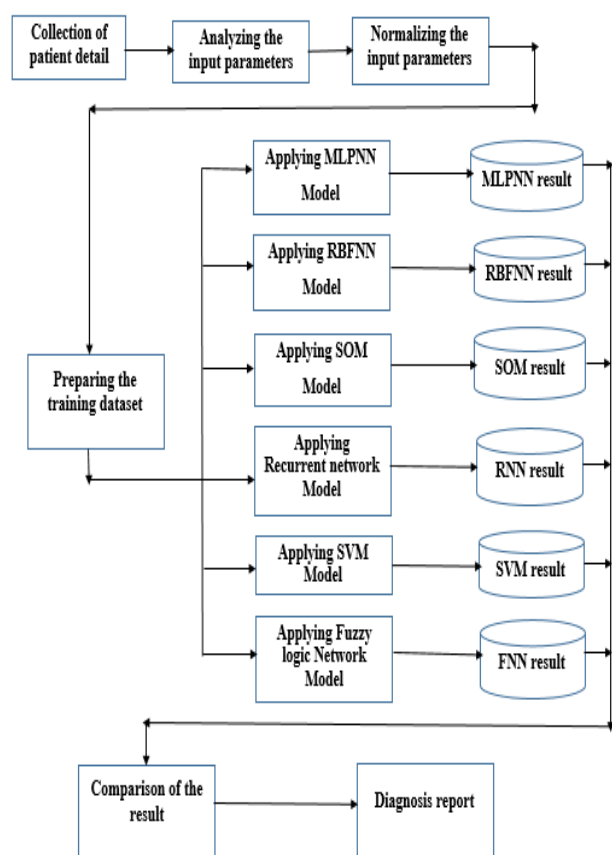

Fig. 1 Schematic diagram of proposed methodology

This fig shows that after the collection and preprossing of the dataset its starts the training, testing and the validation phase. After comparing the network structure the final method is selected according to their accuracy percentage.

3.1 Dataset description

Gestational Diabetes Mellitus (GDM) can occur in every pregnant women. However, women who have close relatives with the disease are somewhat more likely to develop it. Other risk factors include obesity, high cholesterol, high blood pressure and physical inactivity. The risk of developing GDM also increases, as women take baby at an older age. Women who have overweight are more likely to develop GDM. Women who develop diabetes while pregnant are more likely to develop full-blown diabetes later in life. Poorly managed diabetes can lead to a host of long-term complications among

these are heart attacks, strokes, blindness, kidney failure, blood vessel disease.

Dataset includes the following attributes (1-8 attributes as input and last attribute as target variable) number of times pregnant, Plasma glucose concentration a 2 hours in an oral glucose tolerance test, Diastolic blood pressure (mm Hg), Triceps skin fold thickness (mm), 2-Hour serum insulin (mu U/ml), Body mass index (weight in kg/ (height in m)^2), Diabetes pedigree function and Age (years). Class to be predicted is patient is tested-positive or tested-negative. A total of 768 cases are available in PIDD. 5 patients had a glucose of 0, 11 patients had a body mass index of 0, 28 others had a diastolic blood pressure of 0, 192 others had skin fold thickness readings of 0, and 140 others had serum insulin levels of 0. After deleting these cases there were 392 cases with no missing values (130 tested positive cases and 262 tested negative).

3.2 Distinct Neural Network and Techniques

3.2.1 Artificial Neural Network

An Artificial Neural Network (ANN) is inspired by observed process in natural networks of biological neurons in the brain. A very important feature of these networks is their adaptive nature, where "Learning by example" replaces "programming" in solving problems. This is a pattern recognition method which can recognize hidden patterns between independent and dependent variables.

Neural networks are non-linear statistical data modeling tools and can be used to model complex relationships between inputs and outputs or to find patterns in a dataset. Machine learning and statistical pattern recognition has been the subject of tremendous interest in the biomedical community as these approaches offer promise for improving the sensitivity and specificity of detection and diagnosis of disease. Moreover, these approaches reduce the potential for human error in the decision making process. ANN are widely used for classification and diagnosis in various thirst areas like effective decision making in medical field, signal processing and so on. In biochemical analysis, artificial neural networks have been used to analyze blood samples, track glucose levels in diabetes.

Neural network processes three different types of neuron such as input neurons, hidden neurons and output neurons [19]. Neurons are placed in the layer and the neurons of each layer operate in parallel. The first layer is the input layer. The activity of input units represents the non-processed information that entered the network; at that layer neuron does not

perform any computations. The hidden layer follows the input layer, and the activity of each hidden unit is determined from the activity of the input units and the weights at the connection of input and hidden units. A neural network model can have many or none hidden layers and their role is to improve the network performance. The last layer is the output layer. The output of the layer is the output of the whole network. Neurons of the output layer, in contrast to input layers perform calculations. The neuron numbers of the external layer is determined with the number of output parameters [20]. The weights are continuously modified until the neural network is capable of predicting the outputs within an acceptable user-defined error level.

A multiple input neuron model is shown in Fig. 2, which consists of a single neuron with x_n inputs. A constant "1" is introduced to the neuron as an input and is multiplied by the bias, b.
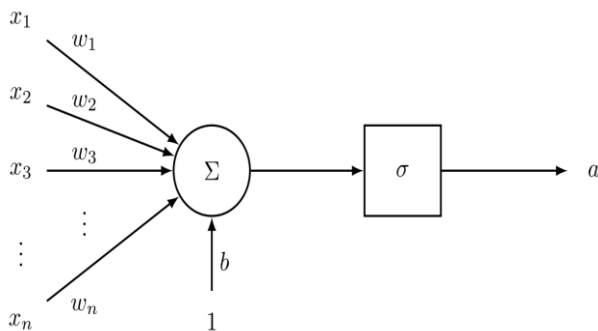
Fig. 2 Multiple input neuron model

Hence, the net input to the transfer function (f) is defined as the sum of the bias and the weighted inputs, such that:

$$net\_input = \sum_{n=1}^{n} W_n X_n + b \qquad (1)$$

where n represents the total number of input variables, xn shows input and wn represents synaptic weight. The output for the neuron is computed using the transfer function, such that

$$Output = f(net\_input) \qquad (2)$$

3.2.2 Multilayer Perceptron Neural Network

The multilayer perceptron (MPL) is one of the most common neural network architectures and has been used successfully in various applications. The MPL network is the most extensively used type of neural network. Multilayer perceptron (MLP) are layered feed-forward networks typically trained with static back-propagation. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that

they are easy to use, and that they can approximate any input/output map. The models are made up of multiple layers of nodes in a directed graph and each of the layers are connected with the adjacent one, hence the name multilayer perceptron. Multilayer perceptron are generally made up of three or more layers. These may consist of an input layer, output layer and one or more hidden layers. They are considered deep neural networks because each layer contains non-linearly activating nodes. Basically the nodes of one layer connect with a specific weight to the nodes of the next layer. The multilayer perceptron uses back-propagation method. Back-propagation is simply a generalization of the least mean squares algorithm in linear perceptron. This is a supervised learning technique where learning occurs by changing connection weights after data is processed. Data processing is done based on the amount of errors in the output compared to the expected results.
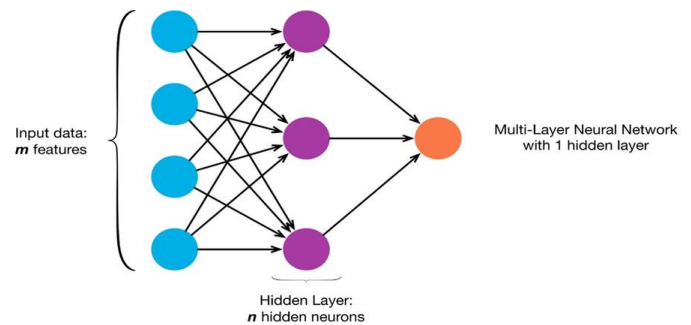

Fig. 3 Multilayer perceptron neural network with one hidden layer

MLP neural network architecture with one hidden layer is shown in Fig. 3. In this network structure, there are four nodes in the input layer, n nodes in the hidden layer, and one node in the output layer. Thus, the network structure can be defined as 4-n-1. For detecting GDM using multilayer perceptron neural network (MLPNN) this study uses the following network configuration.
Hidden layer = 1, Processing elements = 7, Epoch number = 20000, Activation function = SigmoidAxon, Learning Rule = Momentum, Time of runs = 3 times, Threshold = .000001.

3.2.3 Radial Basis Function Neural Network

Radial basis function (RBF) networks are nonlinear hybrid networks typically containing a single hidden layer of processing elements. This layer uses Gaussian transfer functions, rather than the standard sigmoidal functions employed by MLPs. The centers and widths of the Gaussian's are set by unsupervised learning rules, and supervised learning is applied to the output layer. These networks tend to learn much

faster than MLPs. Radial basis function neural network (RBFNN) typically have three layers: an input layer, a hidden layer with a non-linear RBF activation function and a linear output layer. The input can be modeled as a vector of real numbers. The output of the network is then a scalar function of the input vector. An input vector is used as input to all radial basis functions, each with different parameters. The output of the network is a linear combination of the outputs from radial basis functions or units.
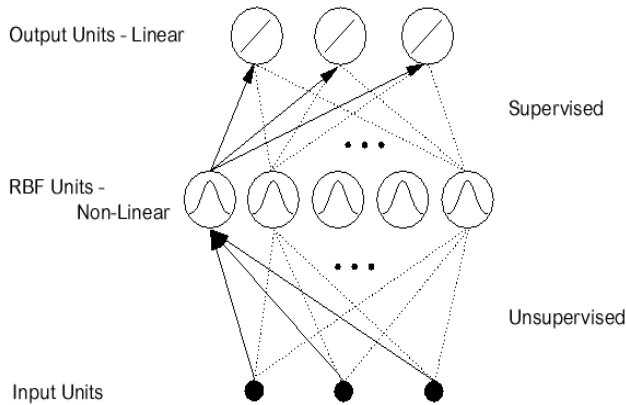

Fig. 4 Radial Basis Function neural network

For detecting GDM using Radial Basis Function neural network (RBFNN) this study uses the following network configuration. Hidden layer = 1, Processing elements = 5, Epoch number for unsupervised learning = 10000, Epoch number for supervised learning = 20000, Activation function = SigmoidAxon, Learning Rule = Momentum.

3.2.4 Self-Organizing Feature Map Neural Network

Self-organizing feature maps (SOM) transforms the input of arbitrary dimension into a one or two dimensional discrete map subject to a topological (neighborhood preserving) constraint. The feature maps are computed using Kohonen unsupervised learning. The output of the SOM can be used as input to a supervised classification neural network such as the MLP. This network's key advantage is the clustering produced by the SOM which reduces the input space into representative features using a self-organizing process. Hence the underlying structure of the input space is kept, while the dimensionality of the space is reduced. For detecting GDM using the network structure by using the Self-organizing feature map for this study uses two hidden layer one with seven processing element and 30000 epoch and the other with five processing element and 30000 epoch, Activation function Sigmoid Axon and Learning Rule Momentum factor.

3.2.5 Recurrent Network

A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. This allows it to exhibit dynamic temporal behavior for a time sequence. Unlike feed-forward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. Fully recurrent networks feed-back the hidden layer to itself. Partially recurrent networks start with a fully recurrent net and add a feed-forward connection that bypasses the recurrence, effectively treating the recurrent part as a state memory.
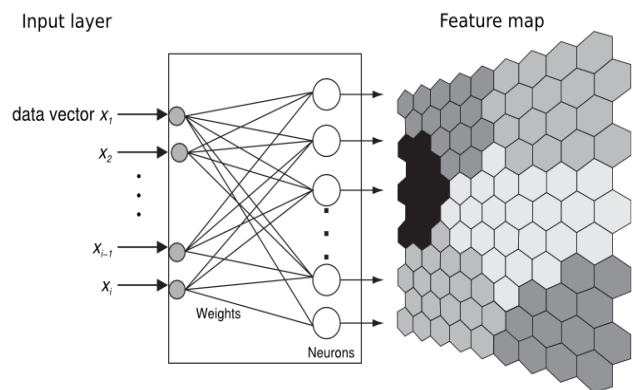

Fig. 5 Self-Organizing Feature Map neural network

For detecting GDM using the network structure by using the recurrent neural network for this study uses one hidden layer with five processing element and 50000 epochs. The activation function is SigmoidAxon and Learning Rule is Momentum factor.
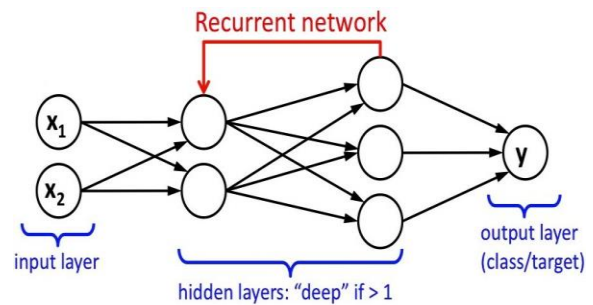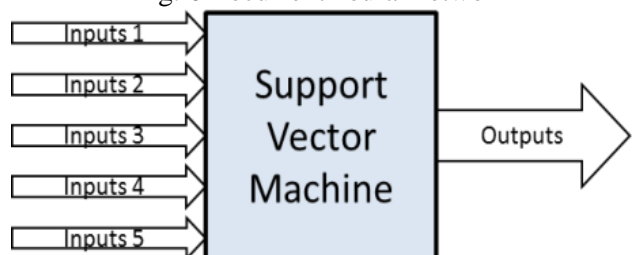

Fig. 6 Recurrent neural network

Fig. 7 Support vector machine

3.2.6 Support Vector Machine

Support vector machine (SVM) is a relatively new form of supervised machine learning. The Support Vector Machine is implemented using the kernel Adatron algorithm. The kernel Adatron maps inputs to a high-dimensional feature space, and then optimally separates data into their respective classes by isolating those inputs which fall close to the data boundaries. Therefore, the kernel Adatron is especially effective in separating sets of data which share complex boundaries. SVMs can only be used for classification, not for function approximation. For detecting GDM using the network structure by using the support vector machine (SVM) neural network for this study uses zero hidden layer with one processing element and 10000 epochs. The activation function and Learning Rule is same as the others that is SigmoidAxon and Momentum factor.

3.2.7 Fuzzy Logic

The CANFIS (Co-Active Neuro-Fuzzy Inference System) model combines two approaches, ANN and FL which integrates adaptable fuzzy inputs with a modular neural network to rapidly and accurately approximate complex functions. Fuzzy inference systems are also valuable as they combine the explanatory nature of rules (membership functions) with the power of "black box" neural networks.
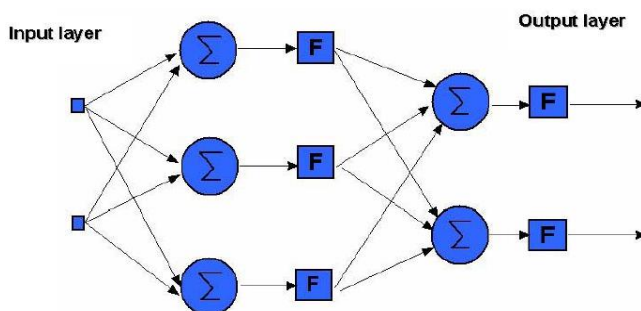

Fig .8 Architecture of Fuzzy neural network

For detecting GDM using the network structure by using the fuzzy logic network for this study uses 20000 epochs. The activation function and Learning Rule is same as the others, Sigmoid Axon and Momentum factor.

3.3 Normalization

Normalization is required so that all the inputs are at a comparable range. Because in input layer the multiplied value of weight and input variable should activate to very small less than 3 so it is necessary to get better result it should be normalized. Normalization (or scaling) is one of the main parts of ANN learning process. Normalization is important in ANNs because real data obtained from experiments and analysis, most times are distant from each other. The effect is great because the common activation functions such as sigmoid, hyperbolic, tangent and Gaussian produce result that ranges between [0,1] or [-1,1]. It is important to normalize the values to be in that range. If you we not normalize our inputs between (0,1) or (-1,1) you could not equally distribute importance of each input, thus naturally large values become dominant according to less values during ANN training. Normalization is the process of reorganizing data in a database so that it meets two basic requirements: (1) There is no redundancy of data (all data is stored in only one place), and (2) data dependencies are logical (all related data items are stored together). Therefore the experimental data are normalized according to Eq. (10) in order to train the neural network effectively. Normalization confirms that the neural network was trained effectively without any particular variable significantly distorting the result.

$$X_N = 2\frac{x - x_{min}}{x_{max} - x_{min}} - 1 \qquad (3)$$

where, X_N is the normalized value of the real variable
x is the measured value of the variable
x_{min} is the minimum value of the real variable
x_{max} is the maximum value of the real variable

3.4 Training, Testing and Validation

Once a network has been structured for a particular study, that network is ready to be trained. To start this process the initial weights are chosen randomly. There are two approaches to training - supervised and unsupervised. Supervised training involves a mechanism of providing the network with the desired output either by manually "grading" the network's performance or by providing the desired outputs with the inputs. Unsupervised training is where the network has to make sense of the inputs without outside help. Unsupervised training is used to perform some initial characterization on inputs. In supervised training, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs. Errors are then propagated back through the system, causing the system to adjust the weights which control the network. The set of data which enables the training is called the "training set." During the training of a network the same set of data is processed many times as the connection weights

are ever refined. The other type of training is called unsupervised training. In unsupervised training, the network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. This is often referred to as self-organization or adaption. The study of detecting Gestational Diabetes Mellitus (GDM) used 60% of total dataset for training approach.

Testing of the data is used to confirm the expected result, i.e. when test data is entered the expected result should come and some test data is used to verify the software behavior to invalid input data. Test data is generated by testers or by automation tools which support testing. Data validation is intended to provide certain well-defined guarantees for fitness, accuracy, and consistency for any of various kinds of user input into an application or automated system. This study uses 33% of total dataset for the testing purpose.

Data validation is the process of ensuring data have undergone data cleansing to ensure they have data quality, that is, that they are both correct and useful. It uses routines, often called "validation rules" "validation constraints" or "check routines", that check for correctness, meaningfulness, and security of data that are input to the system. The rules may be implemented through the automated facilities of a data dictionary or by the inclusion of explicit application program validation logic. This study uses 7% data from the dataset.

3.5 Performance Testing

Applying different artificial neural network structure for any specific study we can get different result. But we have to find the best network structure considering different parameters like Mean Squared Error (MSE), Minimum Absolute Error (MAE), Maximum Absolute Error and Linear Correlation Coefficient (r).

Mean Square Error: In neural network, the mean squared error (MSE) or mean squared deviation (MSD) of an estimator (of a procedure for estimating an unobserved quantity) measures the average of the squares of the errors—that is, the average squared difference between the estimated values and what is estimated. With the MSE performance function some other performance functions such as Mean absolute error (MAE), Root Mean Squared Error (RMSE) will be used. The MSE and RMSE will be obtained by Eq. (4) and Eq. (5), respectively.

$$MSE = 1/N \sum_{o=1}^{N}(t_T - y_o)^2 \qquad (4)$$

$$RMSE = \sqrt{1/N \sum_{o=1}^{N}(t_T - y_o)^2} \qquad (5)$$

where, N is the total number of training pattern.

Mean Absolute Error: MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It's the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. The MAE can be obtained by Eq. (6).

$$MAE = 1/n \sum_{j=1}^{n} |y_j - \hat{y}_j| \qquad (6)$$

Min Absolute Error and Max Absolute Error: Min Abs Error and Max Abs Error can be calculated by the following equations respectively

$$Min\,Abs\,Error = min\,|y_i - t_j| \qquad (7)$$
$$Max\,Abs\,Error = max\,|y_i - t_j| \qquad (8)$$

where, yi is desired output, tj is computed output.

Normalized Mean Square Error: The NMSE (Normalized Mean Square Error) is an estimator of the overall deviations between predicted and measured values. It is defined as:

$$NMSE = RMSE / (y_{max} - y_{min}) \qquad (9)$$

NMSE generally shows the most striking differences among models. If a model has a very low NMSE, then it is well performing both in space and time. On the other hand, high NMSE values do not necessarily mean that a model is completely wrong.

Linear correlation coefficient, R: R value, the coefficient of correlation or determination are used to measure the correlation between actual and predicted value. It measures the direction and strength of the linear relationship between actual and predicted value. The r value is always between -1 and +1. The highest value of r is 1. The r value is a measure of how well the variation in the output is explained by the target. There is perfect correlation between target and output when this value is equal to 1. A positive r indicates a positive relationship and a negative r indicates a negative relationship.

4 Results and Discussion

4.1 Multilayer perceptron neural network

4.1.1 Training
The training results of the multilayer perceptron neural network with a specific network configuration are shown below.

Table 1 is showing the training result of all the runs of multilayer perceptron neural network with a training minimum value and training standard deviation that gives the value of average of minimum MSE and average of final MSE.

Table 1 Training result of MLPNN

All Runs	Training Minimum	Training Standard Deviation
Average of Minimum MSEs	0.02816889272 26686	0.00681187592606 33
Average of Final MSEs	0.02816889272 26919	0.00681187592604 67

Table 2 is showing the training result of the best network among all network configurations with their run no, epoch no, minimum MSE and final MSE.

Table 2 Training result of the best MLP neural network

Best Network	Training
Run	2
Epoch	19991
Minimum MSE	0.0235771022072017
Final MSE	0.0235771022072026

Fig. 9 is showing graphical view of the convergence of average MSE with respect to epoch no. The solid line is showing the average MSE with respect to epoch of training and the dashed line is for the standard deviation.
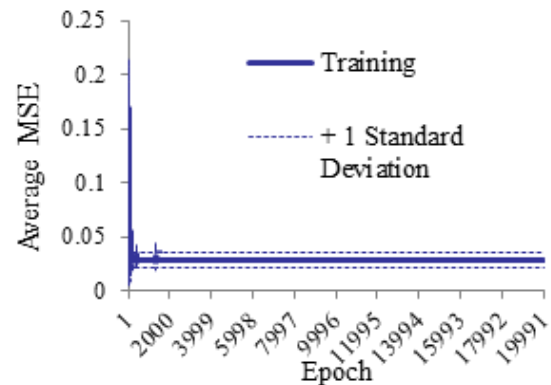

Fig. 9 Convergence of MSE with respect to epoch in training

Fig. 10 is showing graphical view of the Convergence of MSE with respect to number of iteration or epoch. Here MSE is calculated by Run1, 2 and 3.
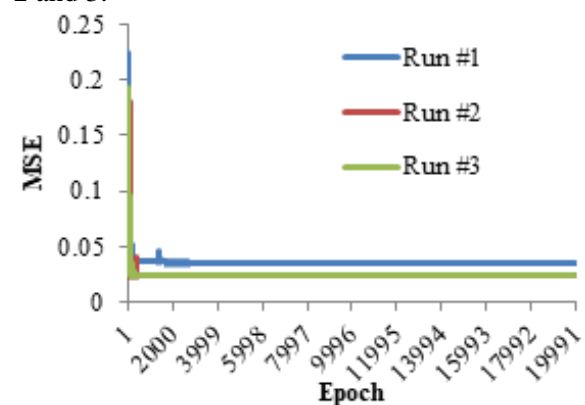

Fig. 10 Convergence of MSE with respect to number epoch

4.1.2 Testing
The testing result of the built network configuration with computed and desired output number is showing on the Table 4.3 Here the outcome of 0's and 1's is given by the matrix form.

Table 3 Testing results of desired output (0) and output (1) of MLPNN

Output / Desired	Outcome (0)	Outcome (1)
Outcome(0)	57	15
Outcome(1)	125	57

Table 4 shows the result of testing performance parameters like mean square error (MSE), normalized mean square error (NMSE), mean average error, min absolute error, max absolute error, linear correlation coefficient (r) and percent of correct output values. Linear correlation coefficient (r) is used to measure the performance of neural network model in testing with their outcome's.

Table 4 Testing result showing the performance and outcome of MLPNN

Performance	Outcome (0)	Outcome (1)
MSE	0.254838194	0.24088323
NMSE	0.404754398	0.304754398
MAE	0.504754398	0.490712546
Min Abs Error	0.487110916	0.46964414
Max Abs Error	0.527091662	0.512081184
r	.53200	.50012
Percent Correct	31.31868132	79.16666667

4.1.3 Validation
Validation result of a network structure proves the validity of a research. By using the MLP neural network this study has achieved only 64% accuracy. In the fig there is the comparison of desired and computed neural network output.
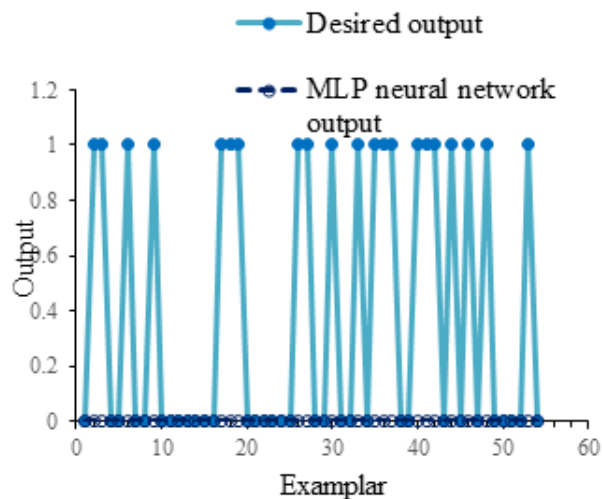
Fig. 11 Desired versus actual MLP neural network output

4.2 Radial basis function neural network
The radial basis function neural network (RBFNN) configuration for this research shows the more accurate result than the other network configuration. The training, testing and validation results are shown by the figures and tables given below.

4.2.1 Training
The training result of radial basis function neural network structure for the research is shown for the epoch number of 10000.
In the following table the computed minimum mean square (MSE) error and final MSE for the best network configuration structure is given.

Table 5 Training result of the best network of RBFNN

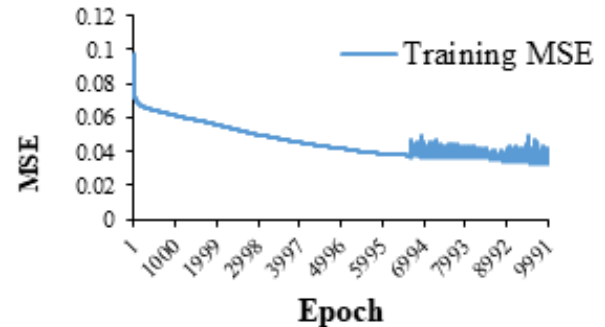| Best Network	Training
Epoch	9990
Minimum MSE	0.033558199
Final MSE	0.033598587

Fig. 12 Training MSE vs. Epoch

In the following Fig. 12 training mean square errors vs. epoch is shown for the specific epoch. The solid line is showing the training MSE with respect to the epoch number.

4.2.2 Testing
The testing result of the Radial Basis Function (RBFNN) is shown by the figure given below. The desired output and actual radial basis function neural network output are shown concurrently.
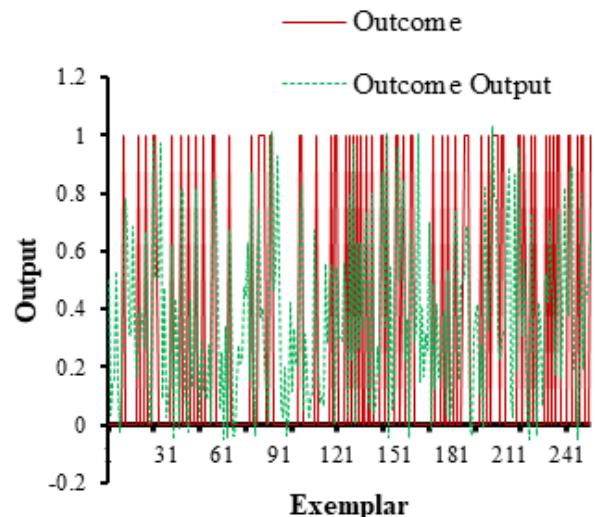

Fig. 13 Testing graph of desired output and actual output

Table of testing performance parameters like mean square error (MSE), normalized mean square error (NMSE), mean average error, min absolute error, max absolute error, linear correlation

coefficient (r) and percent of correct output values and their calculated outcome is given below.

Table 6 Testing result with performance and output of RBFNN

Performance	Outcome
MSE	0.159367498
NMSE	0.782761984
MAE	0.31359339
Min Abs Error	0.004629013
Max Abs Error	1.050074899
r	0.71565647

4.2.3 Validation
Validation result of this study by using Radial Basis Function neural network (RBFNN) has achieved 87.3% accuracy. This figure shows the graphical view in basis of the comparison between the desired output and Radial Basis Function neural network (RBFNN) output.
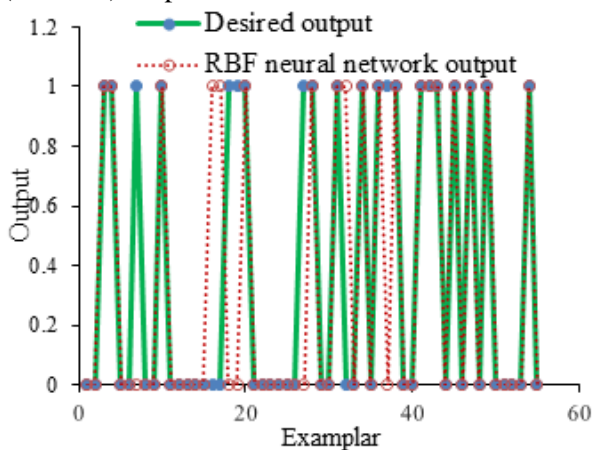
Fig. 14 Desired versus actual RBF neural network output

4.3 Self-organizing feature map
The network structure by using the Self-organizing feature map for this study uses two hidden layer one with seven processing element and 30000 epoch and the other with five processing element and 30000 epochs. The train, test and validation result is given by the following tables and figures.

4.3.1 Training
The following table shows the best network configuration of training result with the epoch number, minimum MSE, final MSE.

Table 7 Training result of SOM neural network

Best Network	Training
Epoch	2601
Minimum MSE	0.051060059
Final MSE	0.065314239

4.3.2 Testing
The testing result of the Self-organizing feature map neural network is shown by the figure given below. The desired output and actual neural network output is shown concurrently by this graphical view.

This table shows the numerical values of testing performance parameters like MSE, NMSE, MAE, Min Absolute Error, Max Absolute Error and linear correlation coefficient (r) value.

Table 8 Testing result with performance parameters and outcomes of SOMNN

Performance	Outcome
MSE	0.173062508
NMSE	0.850027475
MAE	0.349155235
Min Abs Error	0.001765262
Max Abs Error	0.923917657
r	0.504145591

The testing result of the Self-organizing feature map neural network is shown by the figure given below. The desired output and neural network output is shown concurrently.
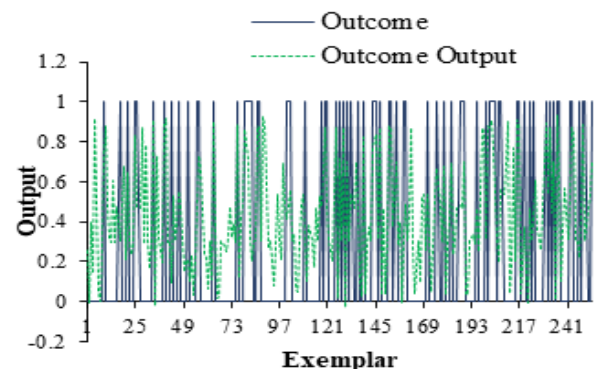

Fig. 14 Testing output versus exemplar

4.3.3 Validation
Validation result of this study by using Self-organizing feature map neural network has achieved 83.64% accuracy. This figure shows the graphical view in basis of the comparison between the desired output and neural network output.
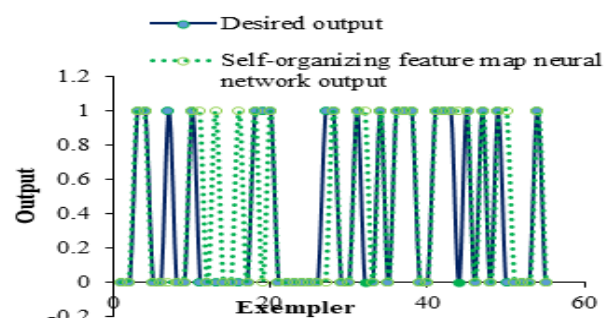

Fig. 15 Desired output versus actual network output

4.4 Recurrent network

The network structure by using the recurrent network for this study uses one hidden layer with five processing element and 50000 epochs. The train, test and validation result is given by the following tables and figures.

4.4.1 Training

The following table shows the best network configuration of training result with the epoch number, minimum MSE, final MSE.

Table 9 Training result of recurrent neural network

Best Network	Training
Epoch	3315
Minimum MSE	0.065820911
Final MSE	0.067774138

In the following figure training mean square error vs. epoch is shown for the specific epoch. The solid line is showing the training MSE with respect to the epoch number.

4.4.2 Testing

This table shows the numerical values of testing performance parameters like MSE, NMSE, MAE, Min Absolute Error, Max Absolute Error and linear correlation coefficient (r) value and their outcome's.

Table 10 Testing result of recurrent network

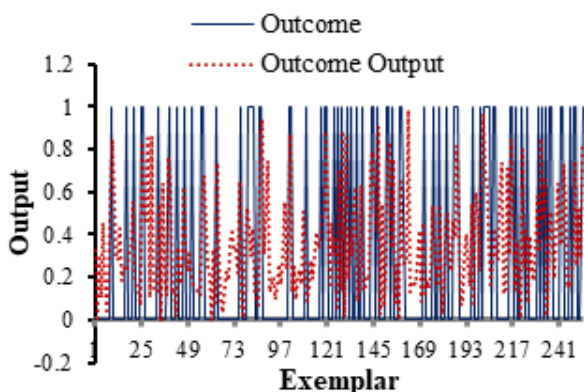| Performance	Outcome
MSE	0.140747968
NMSE	0.691308831
MAE	0.310484452
Min Abs Error	0.000834192
Max Abs Error	0.984280028
r	0.577301842

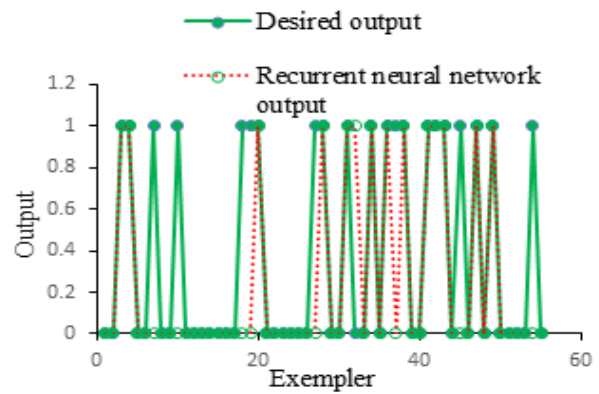Fig. 16 Desired output and actual network output

Fig. 17 Desired versus actual recurrent network output

The testing result of the recurrent neural network is shown by the figure given below. The desired output and actual neural network output is shown concurrently by this graphical view.

4.4.3 Validation

Validation result of this study by using recurrent neural network has achieved 83.64% accuracy on detecting gestational diabetes. This figure shows the graphical view in basis of the comparison between the desired output and recurrent neural network output.

4.5 Support vector machine

The network structure by using the Support Vector Machine neural network for this study does not use any hidden layer but with one processing element and 10000 epochs. The train, test and validation result is given by the following tables and figures.

4.5.1 Training

The following table shows the best network configuration of training result with the epoch number, minimum MSE, final MSE.

Table 11 training result of SVM neural network

Best Network	Training
Epoch	10000
Minimum MSE	0.036260495
Final MSE	0.036260495

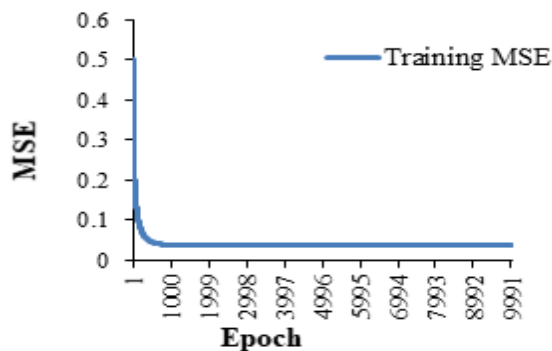Training mean square error with respect to the epoch number of the training phase is given by the following figure.

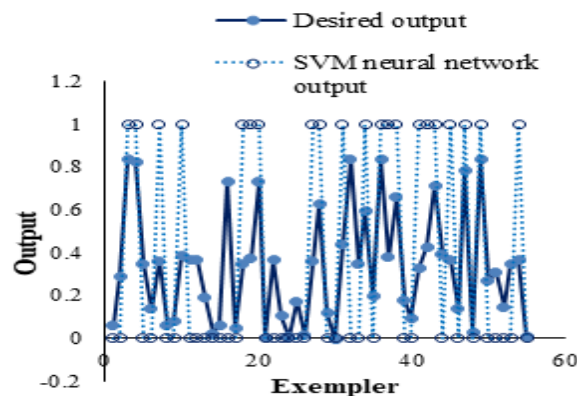Fig. 18 Training MSE vs. Epoch of Support Vector Machine neural network

4.5.2 Testing

The numerical values of testing performance parameters like MSE, NMSE, MAE, Minimum Absolute Error, Maximum Absolute Error and linear correlation coefficient (r) value and their outcomes are given by the following table.

Table 12 Testing result of SVM neural network

Performance	Outcome
MSE	0.210023444
NMSE	1.031567726
MAE	0.366041382
Min Abs Error	0.004239389
Max Abs Error	1.418548728
r	0.427297873

The testing result of the Support Vector machine neural network (SVMNN) is shown by the figure given below. The desired output and actual neural network output is shown concurrently by this graphical view.
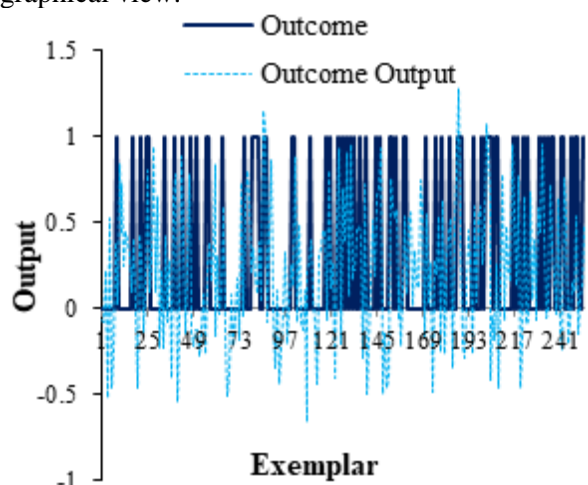

Fig. 19 Testing output versus exemplar

Fig. 20 Desired vs. Computed recurrent SVM neural network output

4.5.3 Validation

Validation result of this study by using SVM neural network has achieved only 76.37% accuracy on detecting gestational diabetes. This figure shows the graphical view on basis of the comparison between the desired output and recurrent neural network output.

4.6 Fuzzy logic
4.6.1 Training

The network structure by using the Fuzzy logic neural network for this study use 20000 epochs. The train, test and validation result is given by the following tables and figures.

The following table shows the best network configuration of training result with the epoch number, minimum MSE, final MSE.

Table 13 Training result of fuzzy network

Best Network	Training
Epoch	14034
Minimum MSE	0.036260495
Final MSE	0.036260495

Training mean square error with respect to the epoch number of the training phase is given by the following figure.
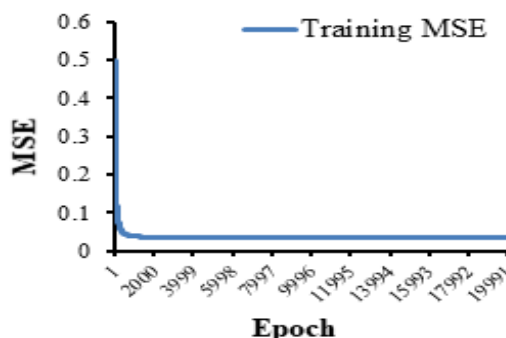

Fig. 21 Training MSE vs. Epoch of Fuzzy logic neural network

4.6.2 Testing

The numerical values of testing performance parameters like MSE, NMSE, MAE, Min Absolute Error, Max Absolute Error and linear correlation coefficient (r) value and their outcome's.

The testing result of the recurrent neural network is shown by the figure given below. The desired output and actual neural network output is shown concurrently by this graphical view.

Table 14 Testing result of fuzzy network

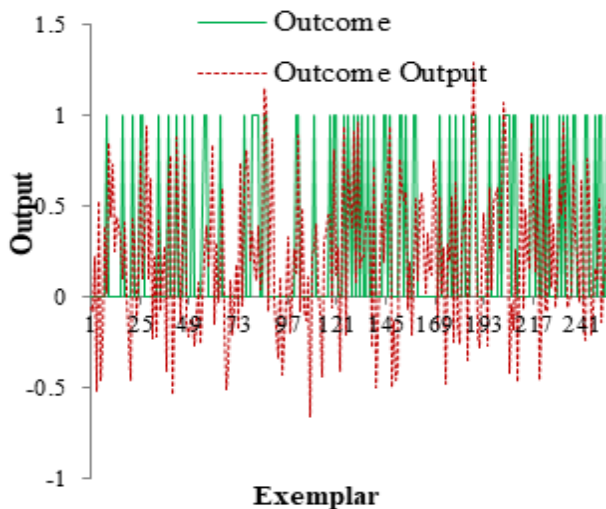| Performance	Outcome
MSE	0.210023444
NMSE	1.031567726
MAE	0.366041382
Min Abs Error	0.004239389
Max Abs Error	1.418548728
r	0.427297873

Fig. 22 Convergence of testing output with respect to exemplar

4.6.3 Validation

Validation result of this study by using Fuzzy logic neural network has achieved 74.6% accuracy on detecting gestational diabetes. This figure shows the graphical view on basis of the comparison between the desired output and fuzzy logic neural network output.
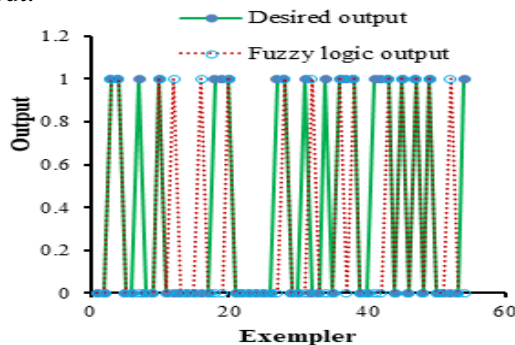

Fig. 23 Desired vs. Computed recurrent fuzzy neural network output

4.7 Comparison Between Distinct Techniques

The different neural network structures vary with their varying epoch number, processing elements, hidden layer, transfer function, activation function. The result for a specific study varies by altering them and they build different network structures that shows different results. Some neural network structures show more accurate result than the others. In every research, the best network is chosen considering the parameters like mean square error (MSE), standard deviation, linear correlation coefficient (r) and many others. In this study there is developed such a comparison among the built neural network structure. There is a table showing all the attributes that differ one network from another in this study. Table is showing the parameters and their values.

Table. 15 Comparison of the neural network structures by training and testing results

Neural network structures	Training	Testing	
	MSE	MSE	r
Multilayer perceptron	0.0235771022072026	0.254838194	0.532001001
Support Vector machine	0.036260495	0.210023444	0.427297873
Recurrent network	0.067774138	0.140747968	0.577301842
Fuzzy logic	0.036260495	0.210023444	0.427297873
Self-organizing feature map	0.065314239	0.173062508	0.504145591
Radial Basis Function	0.033598587	0.159367498	0.71565647

Different neural network structures show different Mean Square Error (MSE) and linear correlation coefficient (r). The value of MSE for each network structure is kept as low as possible. On the other hand a network structure is considered to be better than other when its r value is near to 1. From the table of comparison we can see that the multilayer perceptron neural network shows the lowest MSE of 0.0235771022 in training but in testing its increases to 0.254838194. On the other hand the recurrent neural network structure shows the lowest MSE in testing that is 0.140747968.

From validation we get the average percentage of accuracy. Accuracy of a neural network finally decides which network structure will be more reliable

for this research. Though the r value of recurrent network is good compared to the other except RBFNN and the MSE of multilayer perceptron is low from other structures, their accuracy is lower than radial basis function neural network (RBFNN) structure. Multilayer perceptron shows only 64% accuracy of detecting gestational diabetes. Where support vector machine neural network shows 76.37%, recurrent network 83.64%, radial basis function neural network (RBFNN) 87.3%, and fuzzy logic network 74.6% and self-organizing feature map shows 83.64% accuracy. The tabular form of the comparison between them is given by the following table.

Table. 16 Comparison of the neural network structures by validation results

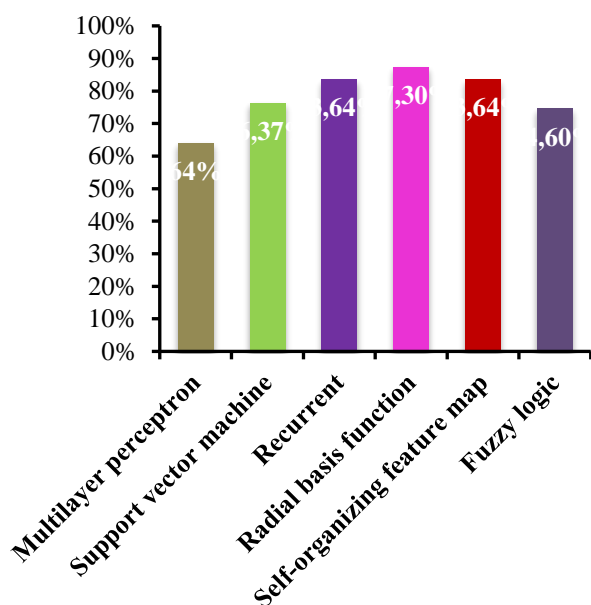| Neural network structure	Accuracy
Multilayer perceptron	64%
Support Vector machine	76.37%
Recurrent network	83.64%
Fuzzy logic	74.6%
Self-organizing feature map	83.64%
Radial Basis Function	87.3%

Fig. 24 Statistical view of the comparison of different neural network structures

The statistical view of the comparison is shown by the diagram given above. RBFNN shows 87.3% which is near to approximately 88%. From the previous research of detecting gestational diabetes the rate of accuracy obtained by radial basis function neural network is only 82%. Thus, the radial basis function neural network structure for detecting gestational diabetes is much better than other research's. Thus, this research of detecting Gestational Diabetes Mellitus (GDM) selects the network structure of Radial Basis Function Neural Network (RBFNN) as it shows the highest accuracy.

5 Conclusion

Gestational Diabetes metabolic disorder is highly prevailing among pregnant women nowadays. Various risk factors are associated with this disorder thus leading to complications to both mother and neonatal. There can be long term risk to both mother and infant if not treated. Therefore, to maintain optimal glycaemic control there is a need to adopt appropriate screening and diagnose method. Even though shortly after delivery glucose tolerance usually returns to normal, there is strong evidence that women with GDM have a high risk for developing diabetes in the course of their lives. Usually only women who have risk factors such as obesity or a family history of GDM are screened earlier on in pregnancy. Therefore, women who develop GDM and do not have these common risk factors often remain undiagnosed until the second trimester and a delay in diagnosis often means therapies for GDM are less effective. This research clearly shows that when considering the inputs to the models, there is at least one input value for which the patient should get the help of a doctor or a hospital staff. The future research can be a system that will help pregnant women in the early stage in diagnosing GDM using newly designed attributes, without even taking a blood test and hence is cost effective. This study offers every pregnant woman the opportunity to know her risk early on without even going to the hospital. As GDM is widely prevalent, many pregnant women have the fear of acquiring it. The risk during pregnancy is less if GDM is diagnosed earlier. Therefore early identification of women at risk of GDM is recommended to prevent complications. The increasing demand of Artificial Neural Network applications for predicting the disease shows better performance in the field of medical decision-making. Considering the great potential of this technique, this research aims to build different artificial neural network model to detect GDM and to compare the models for early prediction of women at risk for the development of gestational diabetes mellitus (GDM) and to choose the best network model among them.

References:

[1] Solis-Herrera C, Triplitt C, Reasner C, et al. *Classification of Diabetes Mellitus*, South Dartmouth (MA): MDText.com, Inc.; 2000-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK279119/

[2] Rani PR, Begum J, Screening and diagnosis of gestational diabetes mellitus, where do we stand, *Journal of clinical and diagnostic research: JCDR*, 2016 Apr;10(4):QE01.

[3] C.-T. Su, C.-H. Yang, K.-H. Hsu, and W.-K. Chiou, Data mining for the diagnosis of type ii Diabetes from three-dimensional body surface anthropometrical scanning data, *Computers & Mathematics with Applications,* vol. 51, no. 6-7, pp. 1075–1092, (2006).

[4] Ahlam A. Sharief and Alaa Sheta, (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 3, No.10, (2014).

[5] Priya ShirleyMuller, S. Meenakshi Sundaram, M. Nirmala, An unconventional Approach forEarly Detection of Gestational Diabetes Mellitus using Influential Parameters, *International Conference on Computational Intelligence and Advanced Manufacturing Research*, vol. 1, pp144-148, (May 2014).

Gestational Diabetes Mellitus, *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 4, pp 296-300, (Nov 2009).

[11] J. K. Olarinoye, A. E. Ohwovoriole, G. O. Ajayi, Diagnosis of Gestational Diabetes Mellitus in Nigerian pregnant women - comparison between 75g and 100g Oral Glucose Tolerance Tests, *West African journal of Medicine*, vol. 23 (3), (July-Set, 2004).

[12] Nassim Doualia, Julien Dollonb, Marie-Christine Jaulenta, Personalized prediction of gestational Diabetes using a clinical decision support system, *INSERM UMR 1142 – Pierre and Marie Curie University and Medical School*, Paris, France, Microsoft, One Microsoft Way, Redmond, WA, 98074, USA, (2013).

[13] B. S. Buckley, J. Harreiter, P. Damm, R. Corcoy, A. Chico, D. Simmons, A. Vellinga, F. Dunne, Gestational diabetes mellitus in Europe: prevalence, current screening practice and barriers to screening, *Diabetic medicine*, (6 December 2011).

[14] Asha Gowda Karegowda, A.S. Manjunath, M.A. Jayaram, Application of genetic algorithm optimized neural network connection weights for medical diagnosis of pima indians diabetes,

[6] Surabhi Nanda, Mina Savvidou, Argyro Syngelaki, Ranjit Akolekar and Kypros H. Nicolaides, *Prediction of Gestational Diabetes Mellitus by maternal factors and biomarkers at 11 to 13 weeks, Prenatal Diagnosis*, Published in Wiley Online Library, vol. 31 (2), pp 135-141, (Jan 2011).

[7] Thach S.Tran, Jane E. Hirst, My An T. Do, Jonathan M. Morris, Heather E. Jeffery, Early Prediction of Gestational Diabetes Mellitus in Vietnam: Clinical impact of currently recommended diagnostic criteria, *Diabetes Care*, Volume 36, pp 618–624, (2013).

[8] Ozcimen EE, Uckuyu A, Ciftci FC, Yanik FF, Bakar C, Diagnosis of gestational diabetes mellitus by use of the homeostasis model assessment-insulin resistance index in the first trimester, *The official Journal of the International Society of Gynaecological Endocrinology*, vol. 24 (4), pp 224-229, (April 2008).

[9] Okeh U. M and Oyeka I. C. A, Receiver Operating Characteristic Curve Analysis Of Diagnostic Tests Results For Gestational Diabetic Mellitus, *Journal of Mathematics*, vol. 8(1),pp 11-17, (Sept - Oct 2013).

[10] Caipo Zhang, Jinjie Song, Zhilong Wu, Fuzzy Integral be Applied to the Diagnosis of *International Journal on Soft Computing (IJSC)*, Vol.2, No.2, (May 2011).

[15] Maham Jahangir, Hammad Afzal, Mehreen Ahmed, Khawar Khurshid, Raheel Nawaz, An Expert System for Diabetes Prediction using Auto Tuned Multi-Layer Perceptron, *Intelligent Systems Conference* 2017, (7-8 September 2017) | London, UK.

[16] Sung Hee Yang, Changsoo Kim, Hyun Sook An, Hyun An and Jin Soo Lee, Prediction of Gestational Diabetes Mellitus in Pregnant Korean Women Based on Abdominal Subcutaneous Fat Thickness as Measured by Ultrasonography, *Diabetes & Metabolism journal*, (Mar. 6, 2017).

[17] Stich, M. T. J., Spoerre, D. J. K and Velasco, D. T. 1999-2000, The application of artificial neural networks to monitoring and control of an induction hardening process, *Journal of industrial technology*, 16(1): 1-11.

[18] R.A. Mahdavinejad, Optimisation of electro discharge machining parameters, *J. Achievements Mater. Manuf. Eng.* 27(2) (2008) 163–166.

[19] Sreedevi E, Vijaya Lakshmi K, Chaitanya Krishna E, Padmavathamma M., "Modelling effective diagnosis of risk complications in

gestational diabetes mellitus: an e-diabetic expert system for pregnant women" SPIE Proceedings, *Fourth International Conference on Digital Image Processing (ICDIP 2012),* vol. 8334, (May 2012).

[20] Parastoo RAHIMLOO, Ahmad JAFARIAN, "Prediction of Diabetes by Using Artificial Neural Network, Logistic Regression Statistical Model and Combination of them", in *Bulletin de la Société Royale des Sciences de Liège*, vol. 68, pp. 1148-1164, 2016.

[21] Huaping Zhou, Raushan Myrzashova, Rui Zheng, "Diabetes prediction model based on an enhanced deep neural network" in *EURASIP Journal on Wireless Communications and Networking*, pp. 1-13, 2020.

[22] Saumendra Kumar Mohapatra, Jagjit Kumar Swain and Mihir Narayan Mohanty, "Detection of Diabetes Using Multilayer Perceptron" in *Advances in Intelligent Systems and Computing*, pp. 109-116, September 2019.

[23] N. Barakat, A. P. Bradley, and M. N. H. Barakat, "Intelligible support vector machines for diagnosis of diabetes mellitus", *IEEE transactions on information technology in biomedicine*, vol. 14, pp. 1114–1120, 2010.

[24] M. Maniruzzaman, N. Kumar, M. M. Abedin, M. S. Islam, H. S. Suri, A. S. El-Baz, and J. S. Suri,"Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm", *Computer methods and programs in biomedicine*, vol. 152, pp. 23–34, 2017.