Data Mining Approach for the Early Risk Assessment of Gestational Diabetes Mellitus

Saeed Rouhani, University of Tehran, Tehran, Islamic Republic of Iran
Maryam MirSharif, University of Tehran, Tehran, Islamic Republic of Iran

ABSTRACT

In this article, the authors proposed the method of medical diagnosis in gestational diabetes mellitus (GDM) in the initial stages of pregnancy to facilitate diagnoses and prevent the affection. Nowadays, in industrial modern world with changing lifestyle alimental manner the incidence of complex disease has been increasingly grown. GDM is a chronic disease and one of the major health problems that is often diagnosed in middle or late period of pregnancy, when it is too late for prediction. If it is not treated, it will make serious complications and various side effects for mother and child. This article is designed for answering to the question of: “What is the best approach in timely and accurate prediction of GDM?” Thus, the artificial neural network and decision tree are proposed to reduce the amount of error and the level of accuracy in anticipating and improving the precision of prediction. The results illustrate that intelligent diagnosis systems can improve the quality of healthcare, timely prediction, prevention, and knowledge discovery in bioinformatics.

KEYWORDS

Artificial Neural Network, Data Mining, Decision Tree, GDM, Risk Assessment

INTRODUCTION

With increasingly easier access to clinical databases and healthcare medical centers, diagnostic systems are now used in a myriad of medical domains. Large volumes of medical data accumulate in healthcare clinical centers. Having access to these data centers, it is possible to extract and discover tacit knowledge behind the data through analysis of real events and results. Nowadays, in the modern industrial world with changing lifestyle alimental manner and slake mobility the incidence of complex diseases like gestational diabetes is rapidly increasing. Recently, high rates of obesity and diabetes as well as low mobility and increasing marriage age, have resulted in increased occurrence of diabetes. Gestational diabetes mellitus is a kind of diabetes that appears for the first time during pregnancy (Ovesen, Jensen, Damm, Rasmussen, & Kesmodel, 2015). Its prevalence is reported to be 1-3% in the United States, 10.9% in Asian countries, 5.2% in Europe (Coustan, 2013) and ranges from 4-17% in all pregnancies (San, Ling, & Nguyen, 2012). GDM statistics are increasing on a daily basis, due to lifestyle changes in industrial urban areas as well as reduction of mobility and environmental factors.

DOI: 10.4018/IJKDB.2018010101

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
Having access to large amounts of data and database available in medical centers, it is possible to apply data mining methods. In the proposed research, after analysis, the primary data and output data of patients discovered that among the input and output data there were patients that they affect to GDM in future, but they have no initial warning symptoms or primary risk factors. Some data were unknown areas for symptoms and reasons of affection. Therefore, these exceptional cases may generate through encryption aspect of knowledge that are imperceptible for humans. However, human knowledge continues to grow and expand. If GDM is not treated or diagnosed in a timely manner, it may cause serious side effects for both the mother and the child. The diagnostic systems of the various complicate diseases like GDM are rapidly increased in recent years. Certain approaches such as expert systems, fuzzy logic, Artificial Neural Networks (ANNs), and decision-making support systems can help to improve the diagnosis and prediction of the disease. Different features of each approach can be used in diverse situations. For detection of these hide aspects of science and recognize the relationship between data simultaneously ANNs help to have a better and higher accuracy in prediction of complex disease like GDM. Complex diseases are often accompanied by unpredictable effects and are not free from false presumptions. Gestational diabetes mellitus is often diagnosed in middle or late period of pregnancy, when it is too late for prediction and the mother should be under medical treatment. In the proposed approach, the risk of GDM notices to the pregnant mother in the initial stages of pregnancy. However, it is possible by paying attention to this warning and taking a suitable food diet and necessary tendency the prevention of affection to GDM. This paper is designed to answer the question of “What is the best approach for the timely and accurate prediction of GDM?” Many data mining techniques have been developed that are applicable for many domains. For this purpose, the two most referenced, popular and basic methods of classification, the techniques of Artificial Neural Network (ANN) and Decision Tree (DT), are used to reduce the amount of error and make more accurate predictions:

- **Gestational Diabetes Mellitus:** GDM is one of the main disorders along pregnancy that is often diagnosed in middle or late pregnancy (Chu et al., 2007). There is a direct relation between performances of fetal and mother’s blood sugar, as which, by control of mother’s blood sugar it is possible to effectively avoid complication of diabetes. Gestational diabetes is a temporary form of diabetes in which the body does not produce adequate amounts of insulin to regulate sugar during pregnancy and with the onset or first recognition occurring during pregnancy (Kim, Cheng, & Beckles, 2008). Insulin is a hormone made in the pancreas. However, if gestational diabetes is not treated, it will affect the mother and child. If one suffers from gestational diabetes, her baby may be at increased risk of excessive birth weight. This can cause the baby to grow too large (macrosomia), early (preterm) birth and respiratory distress syndrome. Babies born early may experience respiratory distress syndrome, Low blood sugar (hypoglycemia), Type 2 diabetes later in life. Babies of mothers who have gestational diabetes, have a higher risk of developing obesity and type 2 diabetes later in life. Uncontrolled gestational diabetes can increase GFAP (Glial fibrillary acidic protein) in Muller cells and retinal thickness of retinal layer in rat offspring’s; therefore, uncontrolled gestational can damage the Muller cells (Larijani, 2009). Untreated gestational diabetes can result in a baby’s death either before or shortly after birth. Gestational diabetes may also increase the following risks for mother: high blood pressure, preeclampsia and future diabetes. Women with gestational diabetes, are more likely to get it again during a future pregnancy (David, Saeb, & Al Rubaean, 2013). It is also important that the baby be monitored for signs of diabetes after being born. In addition, women who have gestational diabetes have an increased risk of developing overt (Type II) diabetes later in life (Ovesen et al., 2015). There are three basic components in effectively managing gestational diabetes: monitoring blood glucose levels, adopting a healthy eating pattern, and physical activity. With well-timed alarm to the patient, self-management and self-care can provide to affection. Gestational diabetes can often initially be managed with healthy eating which is necessary for the rest of the pregnancy;
• **Data Mining:** Data mining techniques are used to find interesting patterns for medical diagnosis and treatment. There are various types of mining techniques which can be applied for predicting GDM like ANN and Decision Tree and logistic regression (Stafford, Kelley, Syka, Reynolds, & Todd, 1984; Vijayarani & Sudha, 2013);

• **Decision Tree:** Decision tree is a popular and important classifier which is easy and simple to implement. It does not have domain knowledge or parameter settings. It is capable of handling huge amounts of dimensional data. It is more suitable for exploratory knowledge discovery. The results attained from Decision Tree are easier to interpret and read (Lingaraj, Devadass, Gopi, & Palanisamy, 2015);

• **Neural Network:** An artificial neural network (ANN), often just called a “Neural network” (NN), is a mathematical model or computational model based on biological neural network. Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problem (San et al., 2012).

**BACKGROUND**

Numerous studies have been done around ANNs and used expert system, fuzzy approach and neural networks in the diagnosis of different cases, several of which are mentioned as follows.

The hybrid approach of adaptive neural fuzzy inference (ANFIS) is used by Physo san et al. to recognize the presence of Hypoglycemia or low blood glucose in Type I diabetes mellitus (T1MD) patients ANFIS is characterized by fuzzy inference system and adaptive neural network capabilities. The effectiveness of the proposed detection method is found to be satisfactory by giving better sensitivity, 79.09% and acceptable specificity, 51.82% (San et al., 2012). To provide the real-time estimations of the appropriate insulin infusion rate for Type I diabetes mellitus (T1DM), the system was developed based on a nonlinear model-predictive controller (NMPC). This paper consists of two compartmental models and a recurrent neural network and an algorithm based on fuzzy logic have been developed for the on-line adaptation of the NMPC control parameters (Zarkogianni, Vazeou, Mougiakakou, Prountzou, & Nikita, 2011). In an article entitled: “Diabetes mellitus forecast using artificial neural networks”, diagnosing diabetes by using artificial neural network and back propagation algorithm is proposed. There were seven inputs: age, BMI (body mass index), serum insulin, plasma glucose. They used artificial neural networks to forecast diabetes by considering the effective factors and choosing the BP algorithm for learning and testing the data and noticing to missing values in the data set (Jaafer & Ali, 2005). In an article published in 2010, an expert system is designed for diagnosis of diabetes type1 with the use of GA-NN (genetic algorithm) and ANFIS (adaptive neural fuzzy inference) intelligently (Sreedevi & Padmavathamma, 2012). There are papers and some researches in diagnosis of diabetes with the use of neural networks based on risk factors (Lakshmi & Padmavathamma, 2013; Sumathy, Thirugnanam, Kumar, Jishnujit, & Kumar, 2010). For diagnosis of diabetes mellitus based on risk factors with 16 inputs layer, an article published by Sumathy, P. et al. to ease the diagnosis and help the users to predict the affection with diabetes himself (Sumathy et al., 2010). An expert system is designed that diagnosis is based on fuzzy logic rules and ANN based on risk factors for Diagnosis of Gestational Diabetes Mellitus Based on Risk Factors (Lakshmi & Padmavathamma, 2013). In all of these methods, the input factors considered for the effective factors in diabetes or GDM are different.

The abundance and frequency in research around ANNs and use of neural network in recent years indicates the ability and appropriateness of ANNs in medical science and researches that make acceptable and more accurate in replication to medicine probe. This paper applied ANN method using real data to extract the knowledge and patterns behind it and anticipate in new term and predicted the risk of GDM in initial stages of pregnancy for timeous prevention and incidental medical cures. Accordingly, the medical cost and the acute complications of affection to GDM will decrease.
increasingly. Expert systems are knowledge-based systems that act and inferences like experts in the situation of no access to experts and specialists. Sometimes, the designed system needs to user interface and fuzzy semantic as well as application of the fuzzy concepts like high or low risk, then the use of fuzzy logic approach is appropriate. Designing an expert system and fuzzy model approach was applied to diagnose gestational diabetes by M. Mirsharif, et.al. It defines and extracts the roles of fuzzy expert system (Figure 1) and fuzzy inference to diagnose GDM in condition of no access to doctors by the author. Expert systems are knowledge-based systems that act as experts when experts and specialists cannot be accessed (Mirsharif & Alborzi, 2014).

Gaussian function was used in order to convert the non-fuzzy variants (numerical) to fuzzy. In designing expert fuzzy system, the relevant knowledge related to determine the inputs and output of the system and also the inference rules were obtained by specialist doctors and also using the knowledge of experts. The MSE error level was obtained as 0.227. In the fuzzy approach, the roles are defined based on the opinions of experts and specialist doctors while the accuracy and correctness of system results depend on the rules that are defined in the knowledge base. It may be associated with incompetency and false or some effective factors and hidden features existence in affection to gestational diabetes which did not inform to them. After extracting the roles, creating the knowledge base system, and comparing the results with actual values, it was discovered that there were patients who were not identified as GDM-prone by the system but were likely to suffer from GDM in the future. Some data were out of defined rule ranges of affection to GDM and there were unknown areas for symptoms and reasons of affection. Therefore, these exceptional cases increase the error of prediction and may generate through encryption aspect of knowledge that are imperceptible for humankind. Medical science continues to evolve every day. In order to detect the hidden aspects of science and recognize the relationship between data simultaneously data mining approach will help to have a better and higher accuracy in prediction of complex disease like GDM. For these purpose, the author applied data-driven techniques to reduce the rate of error in prediction.

MATERIALS AND METHODS

Many data mining techniques have been developed that are applicable for many domains. However, in this paper, the authors used the two most referenced, popular and basic methods of classification, i.e. decision trees and neural networks, to reduce the amount of error and produce more accurate predictions.

Artificial Neural Networks

Artificial Neural Networks with layered architecture have been widely used in different fields like diagnosing, forecasting, economy science and etc. It is a very popular model for its useful applicable attributions. ANNs have a three-layered architecture consist of input layer, hidden layer and output layer. Each layer consists of neurons or nodes. The input units feed into the hidden units, which finally feed into the output layer. Each of the hidden nodes is a linear function of its input. In the proposed

Figure 1. Fuzzy expert system structure (Mirsharif & Alborzi, 2014)
method at first, real number of inputs was given to the network and each input has a real amount as output. Then the output unit is a linear function of its input. ANNs have the parameters for all of their linear functions. These parameters can be tuned to achieve the minimum error on training in prediction. We can train the network by a set of examples to find parameters and eventually setting the minimize amount of error. ANNs have a natural propensity for storing experiential knowledge and making it available for future use (Srivastava & Tripathi, 2011). ANN can save past data and knowledge and make it available for use after learning it (Hirose, Yamashita, & Hijiya, 1991).

In this paper, the input parameters are based on the most common symptoms of gestational diabetes. At first, we directed the parameters with which the patient had been diagnosed. For this purpose, data sets were collected through a survey for extraction of required information from the patient’s files and records in a special clinic of hospital. One thousand and two hundred patient’s files in a hospital laboratory system database were examined among which, two hundred and forty-eight had complete information from the beginning to the end of pregnancy.

Information includes all the clinical characteristics of patient and medical records such as blood pressure, age, weight, the number of pregnancy, thyroid, and abortion. The required information consists of effective factors that affect GDM and the results of Oral Glucose Challenge Test (OGCT) at 24-28 weeks of the pregnancy. There are several effective factors in the emergence of an illness. These factors are different in degree of importance. Based on physiological parameters and based on the result of previous researches and investigations on medical reference books and opinion of obstetrician doctors, four factors of fast blood sugar, body mass index (BMI), blood pressure and age of pregnancy have the highest degree of importance among all effective factors that are operative in emerging GDM (Cunningham et al.). Also, we need the result of Oral Glucose Challenge Test (OGCT) at 24-28 weeks of pregnancy, which assists in the diagnose of GDM. The results of OGT test can be compared to the system output. All inputs and outputs given to the network. By comparing the ANN output with the result of (OGT), the correctness of outputs and the accuracy of prognostication can be examined. Various standards exist for OGT test like: carpenter-coustan, NDDG (the national diabetes data group), ADA (American Diabetic Association). The result of OGT test is interpreted by ADA method as below.

The Evaluation Engine diagnoses the GDM Patient by the following Screening process proposed by ADA method (Lakshmi & Padnavathamamma, 2013):

1. Perform a 75-g OGTT, with OGTT, with plasma Glucose measurement fasting and at 1 and 2h at 24-28 weeks of gestation in women not previously diagnosed with overt diabetes;
2. The OGTT should be performed in the morning after an overnight fast of at least 8h. The diagnosis of GDM is made when any of the following plasma glucose values are exceeded:

   Fasting ≥ 92 mg/dl (5.1 mmol/l)
   1h ≥ 180 mg/dl (10.0 mmol/l)
   2h ≥ 153 mg/dl (8.5 mmol/l)

The scheme of three-layered architecture in the proposed network with its inputs and output is given in Figure 2.

The supervised multilayer feed-forward network with back propagation algorithm is used to predict GDM in the initial stage of pregnancy by MATLAB software. ANNs can recognize the pattern behind data and learn its knowledge for future use. The output of the system is defined as: the value between (0 to less than 0.5) is treated as no danger for affection to GDM and the output between (0.5 and 1) is treated as high risk of affection to GDM. Existence of the high risk of GDM is the notification that if patients ignore this notification, they may be affected by GDM in remained months of pregnancy.
However, by attention to this warning and by taking a suitable food diet and obligatory surveillance, it is possible to prevent from being affected by GDM in initial stages of pregnancy:

\[
\text{Output} = \begin{cases} 
0.0-0.5 & \text{no risk in affection to GDM} \\
0.5-1 & \text{high risk in affection to GDM}
\end{cases}
\]

The number of nodes in the input and output layer depends on the application under consideration. The number of optimal hidden layers and nodes is generally one of the important and difficult problems. A trial-and-error approach seems to be the best network structure. To find the number of hidden nodes by minimum error, an algorithm by Y. Hirose, et al. proposed that found the number of hidden nodes by changing the number of hidden nodes dynamically until a minimum number is found for each convergence occurrence (in each iteration MSE or mean squared error is less than a predetermined value) (Hirose et al., 1991).

The network has one input layer and an output layer and eight nodes in the hidden layer. As shown in Table 1, eight nodes in hidden layer have the least amount of MSE and the maximum amount of \( R^2 \).

After choosing the best architecture (Figure 3) and building the network with eight neurons in the hidden layer, the progress of training and tasting the model begins. Back-propagation learning algorithm is used to train the multi-layered perceptron (MLP). Back-propagation (BP) model is a popular and successful application model and attracted most research interest among all the existing models (Rumelhart, Hinton, & Williams, 1986). This method was discovered by D.E. Rumelhart, et al. BP is the gradient descent search method that minimizes the system error or sum-of-squares error (MSE). Each input pattern is presented to the network and the train of the BP network is started. BP determines the value of output nodes, passes the error on the hidden nodes, and then updates the weights based on the derivation of the error. By training the BP, the system error determined between

<table>
<thead>
<tr>
<th>n</th>
<th>MSE[0,1]</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>0.0738</td>
<td>0.9198</td>
</tr>
<tr>
<td>9</td>
<td>0.1543</td>
<td>0.7744</td>
</tr>
<tr>
<td>10</td>
<td>0.1365</td>
<td>0.5921</td>
</tr>
</tbody>
</table>
the known output given by the user (actual output) and the output from the model (network output decreased) (Phien & Siang, 1993).

**Decision Tree**

One of the data mining tools in relevance to solving real world disease problems is decision tree (DT). The decision tree classified the diabetes dataset. A decision tree is a wonderful classification model, which uses the dataset. The features are extracted from the dataset to give decision. The Class Distribution is defined as Class value “1” means affection to gestational diabetes and 0 means negative gestational diabetes mellitus. The steps of decision tree algorithm are as below: after importing the inputs from dataset for each attribute it finds the information gain from splitting on them and lets the best of them be the attribute with the highest standardized gain after that creates a decision node that splits on the best attribute. Recursive on the sub-list obtained by splitting the best and adding those nodes as children of nodes (YUVARANI & SELVARANI, 2016). In this paper, the decision tree C4.5 algorithm of data mining technique was applied. C4.5 is defined to handle noisy data, better missing data, pre- and post-pruning of decision trees, attributes with continuous values and rule derivation. Data mining is the core part of Knowledge Discovery in Database (KDD) process, which helps to find the patterns and regularities in the large database (Devi & Shyla, 2016). In this paper, The DT model (C4.5) performed in CLEMENTINE software. The classification accuracies of decision tree model are obtained about 0.9300 that shown an acceptable amount of efficiency.

**DISCUSSION**

**ANN**

Artificial neural networks are very suitable in inferring semantics from complicate and ambiguous data. It can be concluded that the pattern recognizes manners that are cryptic and complex to realize by individual and other computerize technique. ANNs can recognize the patterns among information that nobody may know in nowhere about them. Learning ANN is considered as a successful performance in medicine and applied in pattern recognition and function estimation. In some medical researches that achievement the manual results among largest data is not easy, neural networks are very suitable to be an assistance in diagnose and prediction.

The method of calculating MSE Bias and Mad are given as below. Mean squared error measures the expected squared distance between an estimator and the true underlying parameter:

\[ \text{MAD} = \frac{1}{n} \sum \left| \text{actual} - \text{forecast} \right| \]

\[ \text{BIAS} = \frac{1}{n} \sum \left( \text{actual} - \text{forecast} \right) \]

\[ \text{MSE} (\theta^*) = E[(\theta^* - \theta)^2] \]
After the training and testing the network and comparing the actual result by ANNs forecasting, the amount of $R = 0.906$ and $MSE = 0.074$ are estimated in testing level. Other indexes like BIAS and MAD can be used for assessing the accuracy and the validity of this approach (Table 2). Results show that the neural network model has the desirable operation in prediction. The bias (or bias function) of an estimator is the difference between the estimator’s expected value and the true value of the parameter being estimated. It is referred to any kind of systematic error in design or analysis in a study. Bias is a methodology for evaluating the model’s prediction. An estimator with zero bias is called unbiased. Regression shows how scattering data in the data set. In other words, Whatever R value be closer to one the optimal performance is expected. The network performance more than 0.88 is acceptable. The amount of MSE in the same research that done by M. Mirsharif (2014) in order to design a diagnostic fuzzy expert system to predict the risk of GDM was estimated about 0.227. Thus, the accuracy and speed of prediction by data mining approach was less than the error of fuzzy inference system. For detection of hide aspects of science and recognize the relationship between data simultaneously data mining methods such as ANNs and classification by DT can help to improve the better and higher accuracy in prediction of complex disease like GDM.

### DT

Decision tree is an important tool of data mining, which can produce an optimal prediction to woman with highest risk of diagnostic OGTT at week 24-28 of pregnancy. DT provides decision rules and handles noisy data. It helps to find regularities, rules direction and patterns behind the largest database. The tree structure formed expressed in terms that the end user can be take accurate decision based on the input parameters and easily can be understand by the end user. The calculated values of “$R$” are shown in (Table 3) to compare the results.

Based on the results of experiments, it appears that the data mining methods used in this paper had the lowest amount of estimated error in anticipation and the highest level of accuracy that can be used as tools to improve prediction.

Gestational diabetes mellitus is often diagnosed in middle or late period of pregnancy, when it is too late for prediction and mother should be under medical treatment. In the proposed approach, the risk of GDM notices to the pregnant mother in the initial stages of pregnancy. If the patient ignores this notification, she may be affected by GDM in remaining months of pregnancy. However, by paying attention to this warning and taking a suitable food diet and necessary tendency the prevention of affection to GDM is possible in the initial stages of pregnancy. Babies of mothers who have gestational diabetes have a higher risk of developing Type II diabetes later in life and it may increase the risk of developing overt (Type II) diabetes later in life for both the mother and the child. They should be under view of manual control tests.

<table>
<thead>
<tr>
<th>$R$</th>
<th>BIAS</th>
<th>MAD</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.906</td>
<td>0.0004</td>
<td>0.0161</td>
<td>0.074</td>
</tr>
</tbody>
</table>

### Table 3. Comparison of $R$ values

<table>
<thead>
<tr>
<th>Method</th>
<th>$R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT (Clementine)</td>
<td>0.930</td>
</tr>
<tr>
<td>ANN (Matlab)</td>
<td>0.906</td>
</tr>
</tbody>
</table>
CONCLUSION

There are many methods used to generate accurate predictions such as artificial neural network, expert systems, regression and decision tree. Based on the results of experiments, it appears that data mining methods can be used as tools to improve prediction. Among them ANN and DT are the most referenced and popular techniques in classification. There are many successful applications of neural networks in various fields of prediction such as AN networks along with their drawbacks and review their use in the prediction of survival in the field of gastroenterology (Sapra, Mehrotra, & Nundy, 2015). Therefore, in prediction of many kinds of complex disease or other medicine researches, the accuracy and acceptability of ANN is defined. Neural networks are a popular and successful approach to inference the semantics from complicated data and extracted and recognized patterns that approximately are complicated to perception by other computerized techniques and human kind. Also, among data mining methods the decision tree has a good performance. One of the similar applications of DT in prediction is “An Analysis of Decision Tree Models for Diabetes” (2016). This paper claimed that in the analysis of data mining techniques and tools modified J48 Classifier gives 99.87% of highest accuracy using WEKA & MATLAB tool. Since the diabetes is a chronic disease it has to be prevented before it affects people. In the future, diabetes can be prevented using gene analysis and previous history of the diabetes (YUVARANI & SELVARANI, 2016). In a review paper, data mining techniques are applied in health care sector in order to predict diabetes for this purpose, analyzed the Publications and journals and data mining techniques to find out efficient ways to treat them as well. The result show that data mining is a technique used to extract useful information from existing large volumes of data which enable us to gain more knowledge (Lingaraj et al., 2015).

Gestational diabetes mellitus is a chronic disease that needs to be prevented before a person affected. Data mining techniques enable us to discover tacit knowledge behind data and extract medical knowledge, analysis the important clinical parameters and find out efficient ways of treating disease in health care sectors. ANN and DT perform classification with high accuracy and reduce complexity (Table 3). Nevertheless, timely treatment with minimum error is one of the results and outcomes of using information technology in modern world.

Result

There are several other effective risk factors in access to GDM like family history, personal habits, physical examinations, past history, life style that can considered in more researches and investigations about the influence of each of them in local studies. By using other evolitional algorithms and data mining methods for diverse data and various illnesses, the extraction and detection of apt patterns, cryptic and tacit knowledge behind data sets is possible. By accessing the database available in medical centers, the prediction by usage of datasets is possible with data mining methods. In term of inaccessibility to initial data the model driven methods like fuzzy logic and fuzzy expert systems seem to be helpful to be used by the patients as decision-making support systems. It believed that data mining can significantly improve the accuracy and timely prediction that will be helping to improve the quality of health care and prevent from increasing the numbers of mothers and children prone to diabetes in the later life.
REFERENCES


