

# Prediction of gestational diabetes diagnosis using SVM and J48 classifier model

S. Saradha<sup>1\*</sup>, P. Sujatha<sup>2</sup>

<sup>1</sup>Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies(VISTAS), Chennai, India.

<sup>2</sup>Department of Computer Science & Engineering, Vels Institute of Science, Technology & Advanced Studies(VISTAS), Chennai, India.

\*Corresponding author E-mail: saradha.research@gmail.com

## Abstract

Knowledge Discovery in Databases (KDD) process is also known as data mining. It is a most powerful tool for medical diagnosis. Due to hormonal changes, diabetes may occur during pregnancy is referred as Gestational diabetes mellitus (GDM). Pregnant Women with GDM are at highest risk of future diabetes, especially type-2 diabetes. This paper focuses on designing an automated system for diagnosing gestational diabetes using hybrid classifiers as well as predicting the highest risk factors of getting Type 2 diabetes after delivery. One of the common predictive data mining tasks is classification. It classifies the data and builds a model based on the test data values and attributes to produce the new classified data. For detecting GDM and also its risk factors, two classifier models namely modified SVM and modified J48 classifier models are proposed. The data set were collected from various hospitals and clinical labs and preprocessed with discretize filter using weka tool. Missing values are replaced by the suitable values. The final preprocessed data applied in the proposed classifier Model. The output of the proposed model is compared with all the other existing methodologies. Since the proposed model modified J48 classifier model produces more accuracy and low error rate against other existing classifier models.

**Keywords:** GDM, risk factors, classification, SVM, J48, hybrid model.

## 1. Introduction

### Data mining

From the huge amount of data, extracting information or mining knowledge is called Data mining. Descriptive Data mining and Predictive Data Mining are two major classification of data mining. The descriptive describes the interesting properties of data set in concise manner whereas the predictive constructs a set of data models and to predict the behaviour of new data sets. Class description, frequent patterns, Associations, correlations and clusters are some of the descriptive data mining functions. Some of the predictive models of data mining are classification rules, decision trees, mathematical formulae and neural networks. Data mining plays a major role in the areas of market analysis management, fraud detection, sports, medical, education, astrology and corporate analysis. This research work uses data mining to detect the gestational diabetes.

### Diabetes

Diabetes mellitus(DM) is a metabolic disease. Sometimes pancreas produce inadequate insulin or the body cells do not respond properly to the insulin. At that time, blood sugar levels are high. This is the root cause of diabetes. Our body mainly getting the energy from glucose. When our food is getting digested, it is converted into fats, proteins or carbohydrates. Then the carbohydrates are changed as glucose. This Glucose level is controlled by insulin which is secreted by pancreas beta cells. The insulin circulates the glucose metabolism in the tissues throughout

the body. The insufficient insulin cannot move the glucose into the body cells. As the blood glucose is utilized, the blood glucose levels would be lowered. The blood glucose will be increased either insufficient insulin secreted by beta cells or decreased insulin activity. If it is not treated properly, it affects our blood vessels and nerves. So, it may cause many complications like foot ulcers, vision problem, kidney diseases, and heart diseases.

### Types of Diabetes

Type -1 DM, Type-2 DM, Gestational diabetes are the three major types of diabetes mellitus.

- Type 1 DM also referred as 'Insulin Dependent Diabetes Mellitus' (IDDM) or 'juvenile diabetes'. In this case, pancreas is failure to produce insulin. Root cause of this type is unknown. Approximately 10% of diabetes people affected by Type 1 DM. Children are mostly affected by this type. It must be managed with only insulin injections.
- Type 2 DM is a category of 'Non-Insulin Dependent Diabetes Mellitus' (NIDDM) or 'adult-onset diabetes'. In this case, pancreas could not produce enough insulin or the body cells are failed to respond to insulin produced. Excessive weight and not enough exercise are the common factors. It may treat by proper exercise, diet and by medication with or without insulin. Approximately 90% diabetes people are of this type.

- Gestational diabetes affects the female only in the time of pregnancy. Gestational Diabetes is diabetes first recognition during pregnancy in which occurs in about 5% of all pregnancies. If it is untreated, it may cause many health problems for both mother and fetus like Miscarriage, preterm birth, fetal death and congenital malformation. Age, previous unknown stillbirth, family history and excess weight are some of the reasons for GDM. Various screening tests are done at the different weeks of pregnancy to diagnose the diabetes. At the first prenatal routine fasting glucose measurement screening test will be done to assess GDM. The fasting glucose level is  $<5.1\text{mmol/l}$ , then it is treated as normal. Suppose the glucose level is  $>7.0\text{mmol/l}$ , it is suspected pre gestational diabetes. For suspected GDM the glucose level should be in the range of  $5.1\text{-}7.0\text{mmol/l}$ . Oral Glucose Tolerance Test (OGTT) should be done at the 24 weeks of gestation. If the GDM symptoms are present the OGTT should be done again.

### Diagnosing the diabetes

Tests for type 1 and type 2 diabetes and pre diabetes

## 2. Proposed methodology

In this research work, a systematic approach has been developed to detect the gestational diabetes using data mining classifier algorithms.

Proposed algorithms are,

- Modified Support Vector Machine
- Modified J48

### Steps Involved

1. Data set undergoes the Preprocessing technique to identify the missing or zero values.
2. Using the discretize filter, the missing or zero values replaced by the proper value.
3. The preprocessed final data set
4. Uploaded in weka as a csv format.
5. The data in the preprocessed data set can be separated by the risk factors of GDM.
6. The identified data set screening values are considered.
7. Proposed modified SVM and modified J48 applied on data set to classify the GDM and Non-GDM.
8. Decision tree is generated to find out the highest risk factor of Gestational Diabetes Mellitus.
9. Conclude the highest risk factor of Gestational Diabetes Mellitus
10. Compare the results with the three proposed methodologies and find the best classifier model with the evaluation measures.
11. Compare the results of proposed methodology with existing methodologies with the parameters like time taken to build a model, accuracy and error rate.

### Data set

Data collected from the various hospitals and labs from pregnant women. Age, Blood pressure, height, weight, number of times pregnant, hemoglobin, skin fold thickness, insulin level like various measures are taken into consideration. Totally 6629 instances are taken.

All these data are preprocessed using discretize filter before implemented to weka. The following table 2 gives the brief description about the data.

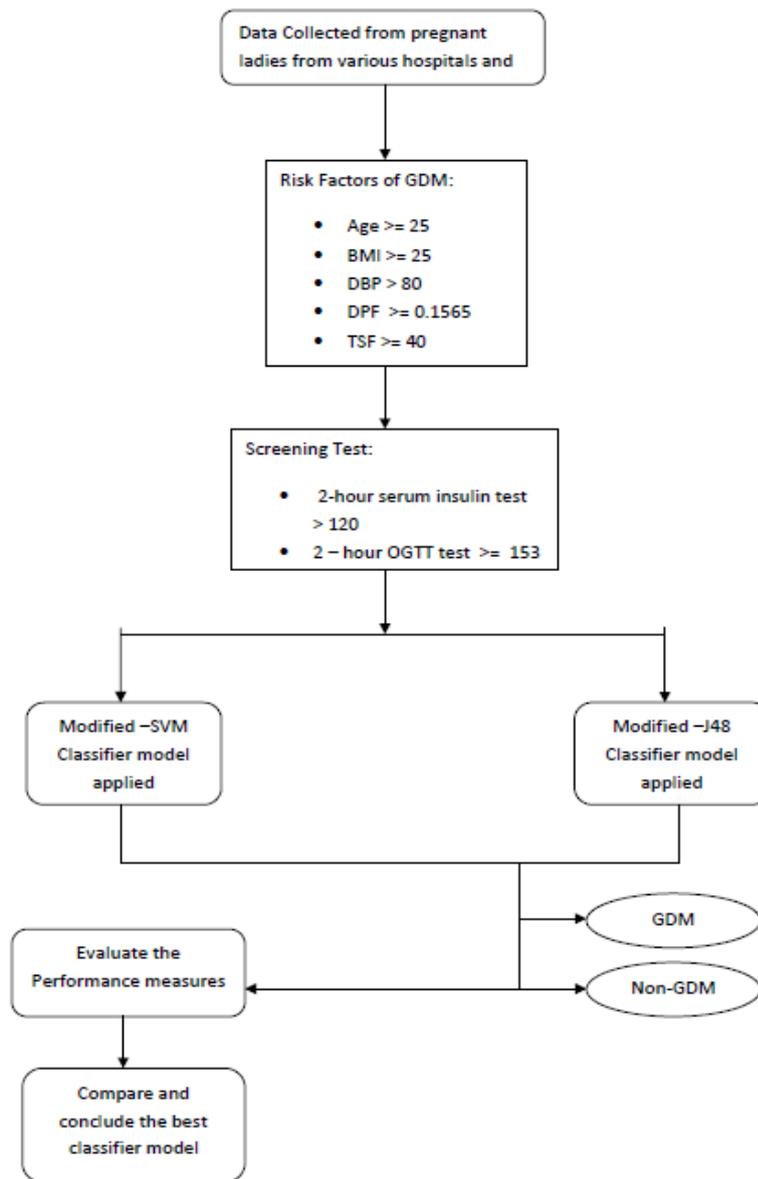
**Table 1:** Diagnosing Diabetes

Tests	Normal Level	Pre Diabetes Level	Diabetes Level
A1C	5.7 percent	5.7 -6.4 percent	6.5 percent and above
Random Blood Sugar	80 – 139 mg/dl	140 – 199 mg/dl	200 mg/dl and above
Fasting Blood Sugar	80 -99 mg/dl	100 -125 mg/dl	126 mg/dl and above
OGTT	Lower than 140 mg/dl	141- 199 mg/dl	200 mg/dl and above

**Table 2:** Data set description

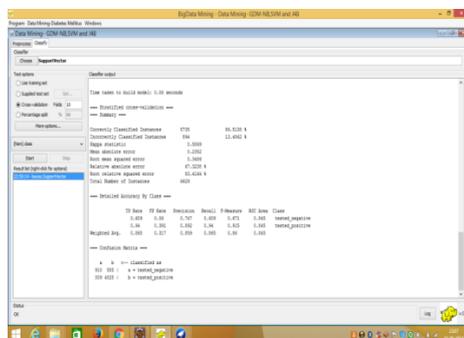
S.No	ATTRIBUTE	DESCRIPTION	TYPE
1	Age	Age (years)	Numeric
2	Ht	Height (in cms)	Numeric
3	Wt	Weight (in kgs)	Numeric
4	Bp	Blood Pressure	Numeric
5	Hmg	Heamoglobin	Numeric
6	Preg	Number of times pregnant	Numeric
7	Plas	Plasma glucose concentration (glucose tolerance test)	Numeric
8	Pres	Diastolic blood pressure (mm Hg)	Numeric
9	Skin	Triceps skin fold thickness (mm)	Numeric
10	Insu	2-Hour serum insulin (mu U/ml)	Numeric
11	Mass	Body mass index (weight in kg/(height in m) <sup>2</sup> )	Numeric
12	Pedi	Diabetes pedigree function(Family history)	Numeric
13	HbA1c	HbA1c test value	Numeric
14	Class	Class variable (1-tested positive for diabetes, 0- tested negative for diabetes)	Nominal

### Schematic diagram of the proposed



### 3. Results and discussions

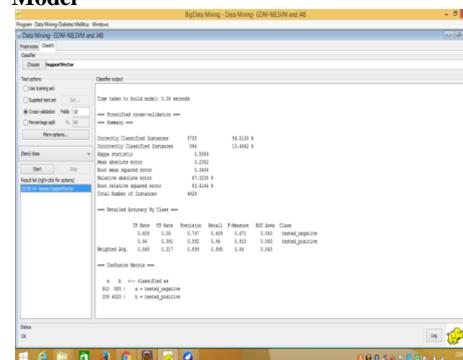
#### Output of support vector machine (SVM)



Time taken to build a model : 0.06 seconds  
 Correctly classified instances : 5735  
 Incorrectly classified instances: 894  
 Accuracy: 86.51%  
 Error rate:0.13  
 Precision: 0.74  
 Recall: 0.60

#### Output of Modified J48 classifier

##### Model



Time taken to build a model : 0.12 seconds  
 Correctly classified instances : 6420  
 Incorrectly classified instances: 209  
 Accuracy: 96.84%  
 Error rate:0.03  
 Precision: 0.92  
 Recall: 0.93

**Table 3:** Comparison with Existing Methods

Classifier	Time Taken to Build a Model	Accuracy Value (in %)	Error Rate	Precision	Recall
Bayes.NaiveBayes	0.01 seconds	78.38%	0.21	0.68	0.69
Rules.ZeroR	0.01 seconds	65.10%	0.34	0	0
Trees.J48	0.08 seconds	78.26%	0.21	0.70	0.63
Meta.AdaBoostM1	0.06 seconds	80.51%	0.18	0.72	0.61
Modified SVM	0.06 seconds	86.51%	0.13	0.74	0.60
Modified J48	0.12 seconds	96.84%	0.03	0.92	0.93

**Table 4:** Comparison with Proposed Methods

SL.NO	Classifiers	Accuracy %
1	Modified SVM	86.51
2	Modified J48	96.84

According to the results, clearly see that decision tree the high accuracy 96.84%. J48 gives the next highest accuracy among all the classifiers in case of diabetes disease.

The proposed classifier model used to identify the highest risk factors of the gestational diabetes. All the patients' data are analysed by using different classifiers such as SVM and J48 Decision Tree. From the results of proposed modified J48 classifier model, conclude that Age factor and strong family history of diabetes, especially first degree of relatives are the two main factors for Gestational Diabetes Mellitus. Fitting predictive models like those utilized as a part of this study might be utilized to create individual/clinical decision support systems, to keep up wellness or enhance the administration of chronic diseases, for example, heart disease and diabetes. Recognizable proof of real hazard factors and creating decision support system, and powerful control measures and wellness programs will decrease in the chronic disease mortality.

#### 4. Conclusion

This paper developed a framework for automatic detection of gestational diabetes using hybrid classifier algorithms of data mining. The data collected from various hospitals and labs are preprocessed and applied on two proposed models namely modified SVM classifier and modified J48 classifier model. The result of the proposed model is compared with all the other existing methodologies and also with each other. Accuracy, error rate, precision and recall are taken as evaluation measures. The decision tree J48 classifier model shows that the proposed model produce more accuracy and low error rate against other existing classifier models.

#### References

- [1] Takashi S, "Management of Gestational Diabetes Mellitus", *Journal of the Japan Medical Association*, Vol.139, No.10, (2011), pp. 2089 -2094.
- [2] Feizi-Derakhshi MR & Zafarani E, "Review and comparison between clustering algorithms with duplicate entities detection purpose", *International Journal of Computer Science & Emerging Technologies*, Vol.3, No.3, (2012).
- [3] Karthikeyani V, "Comparative of Data Mining Classification Algorithm (CDMCA) in Diabetes Disease Prediction", *International Journal of Computer Applications*, Vol.60, No.12, (2012).
- [4] Hasdemir PS, Terzi H & Koyuncu FM, "Recent advances in the diagnosis and management of gestational diabetes", *Turkish journal of obstetrics and gynecology*, Vol.11, No.3, (2014), pp.181-185.
- [5] Lingaraj H, Devadass R, Gopi V & Palanisamy K, "Prediction of Diabetes Mellitus using Data Mining Techniques: A Review", *Journal of Bioinformatics & Cheminformatics*, Vol.1, No.1, (2015), pp.1-3.
- [6] Iyer A, Jeyalatha S & Sumbaly R, "Diagnosis of diabetes using classification mining techniques", *International Journal of data mining & knowledge management process*, Vol.5, No.1, (2015).
- [7] Habibi S, Ahmadi M & Alizadeh S, "Type 2 diabetes mellitus screening and risk factors using decision tree: results of data mining", *Global journal of health science*, Vol.7, No.5, (2015), pp.304-310.
- [8] Nagarajan S & Chandrasekaran RM, "Design and implementation of expert clinical system for diagnosing diabetes using data mining techniques", *Indian Journal of Science and Technology*, Vol.8, No.8, (2015), pp. 771-776.
- [9] Kalaiselvi C & Nasira GM, "Prediction of Heart diseases and cancer in diabetic patients using data mining techniques", *Indian Journal of Science and Technology*, Vol.8, No.14, (2015).
- [10] Saradha S & Sujatha P, "Analysis and significance study of clustering techniques", *IJETER*, Vol.4, No.9, (2016), pp.31-33.
- [11] Durairaj M & Kalaiselvi G, "Prediction of diabetes using back propagation algorithm", *International Journal of Emerging Technology and Innovative Engineering*, Vol.1, No.8, (2015).
- [12] Deepika N & Poonkuzhali S, "Design of hybrid classifier for prediction of diabetes through feature relevance analysis", *IJISSET*, Vol.2, No.10, (2015).
- [13] Amin MN & Habib MA, "Comparison of different classification techniques using WEKA for hematological data", *American Journal of Engineering Research*, Vol.4, No.3, (2015), pp.55-61.
- [14] Saradha S & Sujatha P, "A performance evaluation of classification algorithms for diagnosing gestational diabetes", *JARDCS*, (2017), pp.268 -274.
- [15] Bortolon LNM, Triz LDPL, de Souza Faustino B, de Sá LBC, Rocha DRTW & Arbex AK, "Gestational diabetes mellitus: new diagnostic criteria", *Open Journal of Endocrine and Metabolic Diseases*, Vol.6, No.01, (2016), pp.13-19.
- [16] Sujatha P & Saradha S, "A study of data mining concepts and techniques", *IJAER*, Vol.9, No.27, (2014), pp.9648 -9651.
- [17] Marlina L, "Data mining classification comparison(Naive bayes and C4.5 Algorithms)", *International Journal of Engineering Trends and Technology*, Vol.38, No.7, (2017).