



Rough Set Based Genetic Algorithm (RSBGA) For Assessing Hyperglycemia in Diabetic Patients

R. Manimaran¹ and Dr. M. Vanitha²

¹Research Scholar, PG and Research Department of Computer Science,
J.J.College of Arts and Science (Autonomous), Pudukkottai, Tamilnadu, India.
Email: rmanimaran_jjc@rediffmail.com

²Assistant professor, Department of Computer Applications,
Alagappa University, Karaikudi, India.
Email: mvanitharavi@gmail.com

Abstract—Diabetes mellitus is a chronic and important public health problem burdening patients economically and medically. It ranks seventh in human fatality and a leading cause for adult blindness, renal diseases, non-traumatic limb amputations and impotency. Diabetes may also trigger cardio vascular diseases. People aged less than forty and diagnosed with diabetes, carry the risk of a reducing their lifetime by nineteen years as reported in many studies. Though, there are no perfect curing methods for diabetes, diagnosis and management of diabetes has improved in the past three decades, mainly due to improvements in technology. Early application of aggressive interventions can delay progression of complications due to diabetes to other human body parts or help decrease mortality rates due to diabetes. Though multiple factors contribute towards diabetes, most common reasons for diabetes are diets, inflammations, genetics and environments. Diabetes results from higher glucose levels in humans, mainly due to lower production of insulin in the human

body. Hyperglycemia is a condition indicating diabetes in humans. Precise reasoning and prediction of Hyperglycemia can help physicians in diagnosis, treatment and management of Diabetes. Computer based diagnosis or applications or predictions have been complementing clinicians in assessment, cure and management of diabetes. This paper proposes a novel Rough Set Based Genetic Algorithm (RSBGA) for assessing hyperglycemia early in Diabetic Patients for better prediction and management of Diabetes Mellitus. The proposed technique can be of great help to clinicians in diagnosis and management of Diabetic Patients.

I. INTRODUCTION

India tops in the list of diabetics and is expected to increase by 2030 [1]. Diabetes is a disordered glucose metabolism caused by insulin shortage in the human body. Though multiple factors contribute towards diabetes, most common factors are diet, inflammations, hereditary traits and environment [2]. Diabetes Mellitus is mainly categorized as Type 1

(T1DM) and Type 2 (T2DM) [3]. Increasing incidences of both T1DM and T2DM has made it a public health issue. Complications due to diabetes range from acute hypoglycemia to chronic weakening of the human organ systems like retinopathy or cardiovascular disease. Prescriptions for its cure are based on glycemic control. Carbohydrate produces blood glucose in the body (Normal 50 g), also termed as Glycemic Index. It can be applied to meals or diets. Increase in glycemia (>140 mg/dL) can be related to the presence of diabetes in patients without previous history of diabetes [4]. The variations in glucose levels can be due to many reasons like not having nutrition control or discontinuing therapies leading to higher glucose levels [5].

Increase in blood glucose can lead to increased mortality [6]. Diabetes is patient specific and symptom related and varies on a case to case basis. Physicians take decisions on diabetes by evaluating current conditions and referring to their previous symptoms. Many computer aided decision systems exist for assisting physicians in diagnosis and prediction of complications in the body. Machine learning techniques have been used for forecasting T2DM risks [7]. Diabetes has been detected using neural networks back propagation [8].

Glucose levels in diabetic patients was predicted using Bayesian Modeling technique [9]. A combinations of data mining techniques have also been used to predict diabetes [10]. A boosting classifier model was proposed for prediction of T1DM and T2DM for achieving accuracy of 81% [11]. Thus, this paper aims to judge glycemic index in T2DM patients.

This paper proposes a novel Rough Set Based Genetic Algorithm (RSBGA) for assessing hyperglycemia early in Diabetic Patients for better prediction and management of Diabetes Mellitus. The proposed algorithm is implemented using data collected from real life patients, which can be used in researches as a starting point for improving implemented algorithmic accuracies.

A. Challenges and Issues in predicting Diabetes

Automatic diagnosis of diabetes is an important real-world medical problem as detection of diabetes in its early stages is the key for treatment of diabetes. Physicians need to refer to a patient's history for proper diagnosis. Diabetes being patient specific, its diagnosis is a complex issue. Though 13 different indicators were reported for assessment

of glycemic variability, no methods was clearly defined [12].

In a study, only standard deviation of glycemic index was used to assess mortality [13]. CDSS have been widely been applied for diagnosis [14] from electronic medical record (EMR) data [15], but variations in interactions are multidimensional creating a problem for efficient assessment. Models proposed to solve issues in management of diabetes are mostly restricted to specific type of diabetes. Lack of understanding about the cause of diabetes is another problem which limits prediction or management of diabetes.

Moreover, dimensionality and vagueness of patient's data also creates a challenge in processing. Results based on assumptions of statistical techniques alone can also create problems in accuracy and efficiency.

Translating findings in diabetic research to clinical trials is also difficult. An additional challenge is the heterogeneity in diabetic patients.

Thus, many practical challenges exist in diabetic research both clinically and academically.

II. METHODOLOGY

Diabetes continues to present a significant clinical and economic burden on healthcare systems. New medications for treating T2DM, have been unable to control glycemic levels.

Maintaining glycemic control, as measured by hemoglobin (Hb)A1c levels, in patients with T2DM is crucial for reducing serious clinical complications and subsequent economic costs associated with this chronic disease.

The baseline blood glucose is assessed prior to admissions in hospitals [18]. Acute blood glucose levels and mortality are related to each other [17], making it imperative to monitor glucose levels continuously.

Hence, this research work concentrates on the glycemic levels.

The methodology followed in this research work is data set creation, preprocessing and preparing a data set for application RSBGA classification of diabetic patients.

The methodology followed in this research work is detailed in Figure 1.

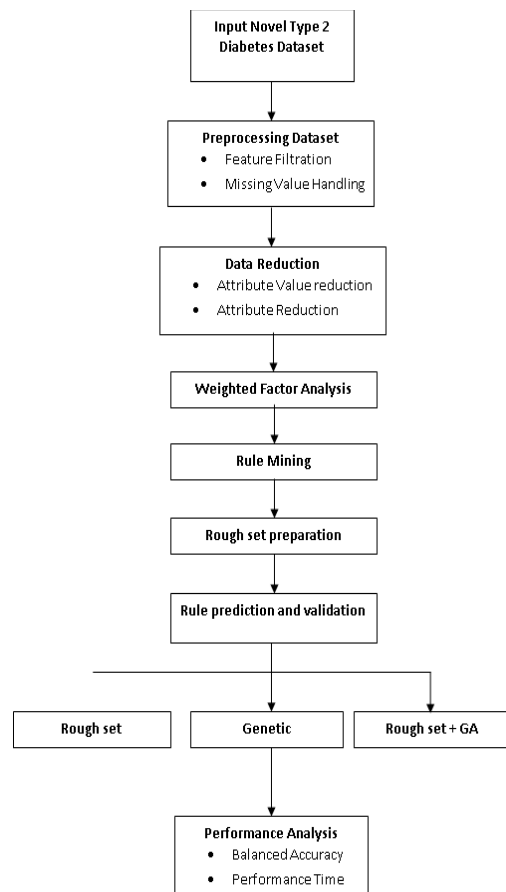


Fig. 1 – Research Methodology

III. RESULTS AND DISCUSSIONS

Multiple studies have shown that adherence to antidiabetic therapies is suboptimal. Moreover, as diabetes is patient specific and symptom related, data collected from diabetic patients vary or may be incomplete in certain aspects. Further, Type 2 diabetes (T2DM) is an example of a complex trait that results from the contribution of many genes, many environmental factors including diet and the interactions among those genes and environmental factors. Various but differing combinations of these factors can produce the same clinical features. T2DM also presents with related physiological responses of

insulin resistance, and other complications. Each of these physiological responses may result from gene nutrient (and environment) interactions. Hence, this research work was implemented using a data set called MV, that was created from real life patient's details. The database consists of Thousand Twenty-Four patient's data on generic, symptoms and disease related factors. The created MV dataset of this research work is depicted in Figure 2, while Table 1 lists the Attributes of the MV dataset.

Fig. 2–MV Dataset

Attribute	Values
Age	1.20-30,2.31-40,3.41-50,4.Above 50
Sex	Male, 2. Female
Marital status	1.Married,2.Single,3.Widowed, 4.Divorced
Nature of works occupation	1.Physical,2.Mental,3.Both 1.Govt.2.Private,3.Business,4.Student, 5.Other
Working Environment	1. City, 2. Village
Working hour	1.6hrs, 2.8hrs, 3.>8hrs
Job Satisfaction	1. Yes, 2. No
Smoking	1. Yes, 2. No
Alcohol	1. Yes, 2. No
Tobacco	1. Yes, 2. No
Food	1.Veg., 2.NV, 3. Both
Water	1. Normal, 2. R.O, 3. Well, 4. All
Affected Age Diabetes	1.20-30,2.31-40,3.41-50,4.Above 50
Heart disease	1. Yes, 2. No
Kidney failure	1.Yes, 2. No
Giddiness	1.Yes, 2. No
Stroke	1.Yes, 2. No
Exercise	1.Yes, 2. No
Treatment	1.Yes, 2. No
High BP	1. Yes, 2. No
Blood Urea	1.20-40mg, 2.Less than, 3. Greater than
Blood Sugar After Food	1.80-140mg, 2.Less than, 3. Greater than

Blood Sugar	1.70-110mg, 2.Less than, 3. Greater than
Before Food	
Weight	1.30-50,2.51-70,3.71-90,4.Above 90

The captured values and attributes are exhaustive and complicated, hence this research work implemented a weighted factor analysis on the attributes for reducing the size of the data set. Common traits possessed by T2DM patients and selected for the reduction process. The objective of this process was to remove unrelated value attributes or neutralize outliers by filling a mean mode value. Each attribute was set using a weight, based on the importance of the attribute. The weighted data set, a reduced version of the MV dataset can be stored easily. This data set reduced processing time and was aimed to improve accuracy and efficiency. Figure 3 depicts the weighted dataset.

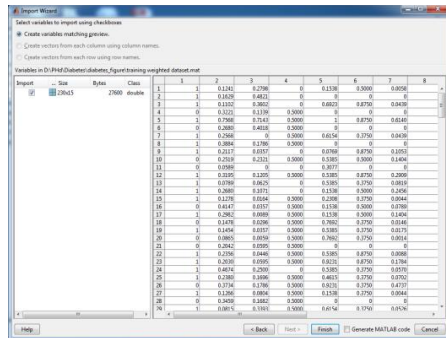


Fig. 3 – Weighted Dataset

RSBGA

The proposed novel algorithm Rough Set Based Genetic Algorithm (RSBGA) for detection of T2DM was implemented and evaluated on MATLAB 6.6 on intel core2 duo system with 4gb ram running windows. The proposed algorithm takes the weighted dataset as an input and generates a decision rule set for core attributes based on weightage of the attributes. Algorithm 1 lists RSBGA

Scheme Evaluation

Data: Patients Data Set

Result: The mean performance values

Step 1: $M=12$:No of Data Set

Step 2: $i=1$;

Step 3: while $i \leq M$ do

Step 4: Read Patients Data Set $D[i]$;

Step 5: Split Data set Instances using % split;

Step 6: Train[i]=60% of D ; % Training Data;

Step 7: Learning(Train[i],scheme);

Step 8: Test Data= $D[i]$ -Train[i];% Test Data;

Step 9: Result=TestClassifier(Test[i],Learner);

Step 10: end

Prediction Algorithm

Input: Diabetic type II Data set with attributes

Output: Diabetics disease prediction accuracy rate measure

Step 1: Initialize diabetics type II data set

Step 2: Analysis Label information

Step 3: Generate random wise to input weight of attributes

Step 4: Calculate weights based on the rule prediction

Step 5: Classify prediction tool based on max weight

Step 6: Regression classification data with weighted value

Step 7: populate weighted matrix for selection process

Step 8: Find maximum fitness of weighted attributes.

Step 9: Crossover weighted population and regenerate population

Step 10 : Classify attributes with weight to predict diabetics

Step 11: Using confusion matrix to measure accuracy

The dependencies like partial or total dependencies in data are identified in RSBGA and redundant data is eliminated, thus clearing the data for rule generation. This refined data is then used for rough set creation and then a genetic algorithm is executed on the rough set to clearly predict T2DM in patients from collected attributes. Table 2 lists the accuracy table of the proposed RSBGA, while Figure 4 shows the time complexity and accuracy plot of RSBGA

TABLE II
ACCURACY TABLE OF RSBGA

iteration	Trainin g time	Testin g time	Avg training accurac y	Avg Testing Accurac y	Elapse d Time
1	0.0072	0	0.7780	0.7795	16.632
2	0.0031	0.0019	0.7868	0.7757	32.884
3	0.0140	0.0006	0.8004	0.7636	49.44
4	0.0112	0.0006	0.8049	0.7641	65.34
5	0.0144	0.0016	0.8127	0.7668	81.582
6	0.0184	0.0028	0.8169	0.7607	98.118
7	0.0337	0.0037	0.8206	0.7615	115.43
8	0.0309	0.0025	0.8237	0.7560	131.98
9	0.0437	0.0028	0.8297	0.7481	148.95
10	0.0415	0.0025	0.8330	0.7459	165.93

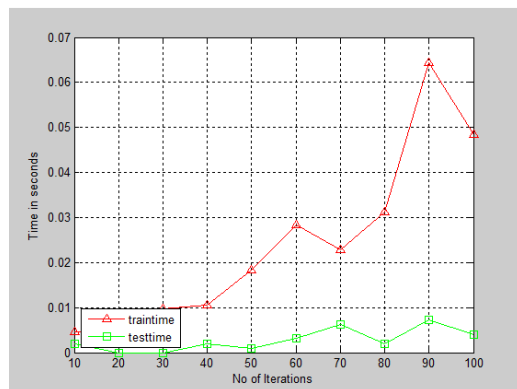


Fig. 4 – RSBGA Time Complexity and Accuracy Plot.

A. Validity of RSBGA

Diabetic complications include vision loss, dialysis, kidney transplant, amputations and stroke. Most of these severe conditions are preventable with proper care. The damage due to complications is linked to the duration and control of glycemic levels and measured by HbA1c over a period of three months. Co-pathologies that occur commonly with diabetes are high blood pressure and abnormal lipids. Hence, current clinical approaches that aim to prevent diabetic complications are focused on achieving targets for glycemic, blood pressure, and lipid control. Better glycemic control has a positive impact on complications due to T2DM and T1DM. Thus, Hypoglycemia is a marker of disease severity or can be directly associated with adverse events in diabetic patients. It can be seen from Figure 4 and Table 2 that RSBGA performs well on T2DM predictions on the MV data set fulfilling the target of glycemic index assessments with more than 75% accuracy. Clinicians would benefit from creative approaches like RSBGA that can reliably predict longer term endpoints. Sensitive tests that can detect and correlate functional changes will be helpful. These approaches need to be validated in clinical trials for its further developments.

IV. CONCLUSIONS AND FUTURE DIRECTIONS

Patient clinical charts and online guidelines are the most important available knowledge sources for physicians. They make future predictions regarding specific observations of the patient. Automatic prediction systems can help physicians to understand patient and make proper decisions. This work designed a hybrid Rough set reasoning model from generated rules in the mining process. RSBGA reduces generation and

understandable decision rules extraction using ARM algorithm from the original MV dataset. Experimental results for the prediction model reveal that performance of the model is above 75% in classifying diabetes types. Correlation-based trends analysis results suggest insights of the patient conditions to physicians in an appropriate way and assisting them in controlling risky stages. Although the proposed hybrid model performs well, further experiments and improvements are required. Additionally, the proposed model can be applied to prediction problems in other fields. Thus, it can be concluded that RSBGA has performed decently with less processing time for prediction of T2DM in MV data set.

REFERENCES

- [1] <https://www.ndtv.com/food/world-diabetes-day-2017-number-of-diabetics-to-double-in-india-by-2023-1775180>
- [2] Stahl F. Licentiate Thesis. Department of Automatic Control, Lund University; Lund, Sweden: 2012. Diabetes Mellitus Glucose Prediction by Linear and Bayesian Ensemble Modeling.
- [3] Thirugnanam M., Kumar P., Srivatsan S.V., Nerlesh C.R. Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach. *Procedia Eng.* 2012;38:1709–1718
- [4] Smith FG, Sheehy AM, Vincent JL, Coursin DB. Critical illness-induced dysglycaemia: diabetes and beyond. *Crit Care.* 2010;14(6):327.
- [5] Ali NA, O'Brien JM, Jr, Dungan K, Phillips G, Marsh CB, Lemeshow S, et al. Glucose variability and mortality in patients with sepsis. *Crit Care Med.* 2008;36(8):2316–2321
- [6] Meyfroidt G, Keenan DM, Wang X, Wouters PJ, Veldhuis JD, Van den Berghe G. Dynamic characteristics of blood glucose time series during the course of critical illness: effects of intensive insulin therapy and relative association with mortality. *Crit Care Med.* 2010;38(4):1021–1029.
- [7] Mani S, Chen Y., Elasy T., Clayton W., Denny J. Type 2 diabetes risk forecasting from emr data using machine learning. *AMIA Annu. Symp. Proc.* 2012;2012:606–615.
- [8] Zolfaghari R. Diagnosis of diabetes in female population of pima indian heritage with ensemble of bp neural network and svm. *Int. J. Comput. Eng. Manag.* 2012;15:2230–7893.
- [9] Stahl F. Licentiate Thesis. Department of Automatic Control, Lund University; Lund, Sweden: 2012. Diabetes Mellitus Glucose Prediction by Linear and Bayesian Ensemble Modeling.
- [10] Thirugnanam M., Kumar P., Srivatsan S.V., Nerlesh C.R. Improving the prediction rate of diabetes diagnosis using fuzzy, neural network, case based (FNC) approach. *Procedia Eng.* 2012;38:1709–1718.
- [11] Ali R., Siddiqi M.H., Idris M., Kang B.H., Lee S. Ubiquitous Computing and Ambient Intelligence. Personalisation and User Adapted Services. Springer; Berlin/Heidelberg, Germany: 2014. Prediction of diabetes mellitus based on boosting ensemble modeling; pp. 25–28.
- [12] Eslami S, Taherzadeh Z, Schultz MJ, Abu-Hanna A. Glucose variability measures and their effect on mortality: a systematic review. *Intensive Care Med.* 2011;37(4):583–593.
- [13] Meynaar IA, Eslami S, Abu-Hanna A, van der Voort P, de Lange DW, de Keizer N. Blood glucose amplitude variability

as predictor for mortality in surgical and medical intensive care unit patients: a multicenter cohort study. *J Crit Care*. 2012;27(2):119–124

[14] 1T.K.Sethuramalingam, 2B.Nagaraj, “A COMPARATIVE APPROACH ON PID CONTROLLER TUNING USING SOFT COMPUTING TECHNIQUES”, *International Journal of Innovations in Scientific and Engineering Research (IJISER)*, Vo.1, No.12, pp.460-465, 2014.

[15] Sheard NF, Clark NG, Brand-Miller JC, et al. Dietary carbohydrate (amount and type) in the prevention and management of diabetes: a statement by the American Diabetes Association. *Diabetes Care* 2004;27:2266–2271

[16] Zhiheng H, King GL. Microvascular complications of diabetes. *Endocrinol Metab Clin North Am* 2004;33:215–38

[17] Qaseem A, Humphrey LL, Chou R, Snow V, Shekelle P. Clinical Guidelines Committee of the American College of Physicians. Use of intensive insulin therapy for the management of glycemic control in hospitalized patients: a clinical practice guideline from the American College of Physicians. *Ann Intern Med*. 2011;154(4):260–267.

[18] Egi M, Bellomo R, Stachowski E, French CJ, Hart GK, Taori G, et al. The interaction of chronic and acute glycemia with mortality in critically ill patients with diabetes. *Crit Care Med*. 2011;39(1):105–111]

[19] Kavanagh BP, McCowen KC. Clinical practice. Glycemic control in the ICU. *N Engl J Med*. 2010;363(26):2540–2546. Erratum in. *N Engl J Med*. 2011;364(12):1182. [PubMed]

