



A Literature Survey of Different Type Diabetes Mellitus Detection Analysis

Rajeev Rajput¹, Dr Jitendra Agrawal², Prof . Sumit Sharma³

M.Tech Scholar CSE¹, Head of Department CSE³, Professor IT²

Vashnavi Institute of Technology (VIT)¹³, School of Information Technology RGPV²

Bhopal [M.P.]

rajeev.rajput183@gmail.com¹

Abstract— In this survey paper discuss the different type of diabetes mellitus detection methods. Now a day's diabetes is one of the most dangerous deceases in this whole world. There are number of different methods available for the analysis of diabetes patients on the basis of patient physical properties such as weight, body mass index, blood pressure and other parameters. There are different classification analysis method available diabetes patient such J-45, support vector machine, multi support vector, navbais and neural network. In this survey paper discuss the different methods of clustering and classification of diabetes patients.

Keywords— J-48, Svm, Multi Svm, Neural Network, Clustering and Classification etc...

I. INTRODUCTION

Diabetes Mellitus ("diabetes" for short) is a serious disease that occurs when your body has difficulty properly regulating the amount of dissolved sugar (glucose) in your blood stream. It is unrelated to a similarly named disorder "Diabetes Insipidus" which involves kidney-related fluid retention problems. In order to understand diabetes, it is necessary to first understand the role glucose plays with regard to the body, and what can happen when regulation of glucose fails and blood sugar levels become dangerously low or high. The tissues and cells that make up the human body are living things, and require food to stay alive. The food cells eat is a type of sugar called glucose. Fixed in place as they are, the body's cells are completely dependent on the blood stream in which they are bathed to bring glucose to them. Without access to adequate glucose, the body's cells have nothing to fuel themselves with and soon die. Human beings eat food, not glucose. Human foods get converted into glucose as a part of the normal digestion process. Once converted, glucose enters the blood stream, causing the level of dissolved glucose inside the blood to rise. The blood stream then carries the dissolved glucose to the various tissues and cells of the body. Though glucose may be available in the blood, nearby cells are not able to access that glucose without the aid of a chemical hormone called insulin. Insulin acts as a key to open the cells, allowing them to receive and utilize available glucose. Cells absorb glucose from the blood in the presence of

insulin, and blood sugar levels drop as sugar leaves the blood and enters the cells. Insulin can be thought of as a bridge for glucose between the blood stream and cells. It is important to understand when levels of insulin increase, levels of sugar in the blood decrease (because the sugar goes into the cells to be used for energy). The body is designed to regulate and buffer the amount of glucose dissolved in the blood to maintain a steady supply to meet cell needs. The pancreas, one of your body's many organs, produces, stores and releases insulin into the blood stream to bring glucose levels back down. The concentration of glucose available in the blood stream at any given moment is dependent on the amount and type of foods that people eat. Refined carbohydrates, candy and sweets are easy to break down into glucose. Correspondingly, blood glucose levels rise rapidly after such foods have been eaten. In contrast, blood sugars rises gradually and slowly after eating more complex, unrefined carbohydrates (oatmeal, apples, baked potatoes, etc.) which require more digestive steps take place before glucose can be yielded. Faced with rapidly rising blood glucose concentrations, the body must react quickly by releasing large amounts of insulin all at once or risk a dangerous condition called Hyperglycemia (high blood sugar) which will be described below. The influx of insulin enables cells to utilize glucose, and glucose concentrations drop. While glucose levels can rise and fall rapidly, insulin levels change much more slowly. When a large amount of simple sugar is eaten the bloodstream quickly becomes flooded with

glucose. Insulin is released by the pancreas in response to the increased sugar. The glucose rapidly enters the cells but the high levels of insulin remain in the bloodstream for a period of time. This can result in an overabundance of insulin in the blood, which can trigger feelings of hunger and even Hypoglycemia (low blood sugar), another serious condition. When blood glucose concentrations rise more gradually, there is less need for dramatic compensation. Insulin can be released in a more controlled and safer manner which requires the body experience less strain. This more gradual process will leave you feeling “full” or content for a longer period of time. For these reasons, it is best for overall health to limit the amount and frequency of sweets and refined sugars in your diet. Instead eat more complex sugars such as raw fruit, whole wheat bread and pasta, and beans. The difference between simple and complex sugars (carbohydrates) is exemplified by the difference between white (simple) and whole wheat (more complex) bread. Insulin is the critical key to the cell's ability to use glucose. Problems with insulin production or with how insulin is recognized by the cells can easily cause the body's carefully balanced glucose metabolism system to get out of control. When either of these problems occur, Diabetes develops, blood sugar levels surge and crash and the body risks becoming damaged.

II. TYPES OF DIABETES

Diabetes is a group of diseases in which the body doesn't produce enough or any insulin, doesn't properly use the insulin that is produced, or exhibits a combination of both. When any of these things happens, the body is unable to get sugar from the blood into the cells. That leads to high blood sugar levels. Glucose, the form of sugar found in your blood, is one of your main energy sources. A lack of insulin or resistance to insulin causes sugar to build up in your blood. This can lead to many health problems.

The three main types of diabetes are:

I. TYPE 1 DIABETES

II. TYPE 2 DIABETES

III. GESTATIONAL DIABETES

i. Type 1 diabetes

Type 1 diabetes is believed to be an autoimmune condition. This means your immune system mistakenly attacks and destroys the beta cells in your pancreas that produce insulin. The damage is permanent. What prompts the attacks isn't clear. There may be both genetic and environmental reasons. Lifestyle factors aren't thought to play a role. Diabetes (previously called insulin-dependent or juvenile diabetes) is usually diagnosed in children, teens, and young adults, but it can develop at any age. If you have type 1 diabetes, your pancreas isn't making insulin or is making very little. Insulin is a hormone that enables blood sugar to enter the cells in your body where it can be used for energy. Without insulin, blood sugar can't get into cells and builds up in the bloodstream. High blood sugar is

damaging to the body and causes many of the symptoms and complications of diabetes.

Type 1 diabetes is less common than type 2—about 5% of people with diabetes have type 1. Currently, no one knows how to prevent type 1 diabetes, but it can be managed by following your doctor's recommendations for living a healthy lifestyle, controlling your blood sugar, getting regular health checkups, and getting.

For Parents

If your child has type 1 diabetes, you'll be involved in diabetes care on a day-to-day basis, from serving healthy foods to giving insulin injections to watching for and treating hypoglycemia (low blood sugar; see below). You'll also need to stay in close contact with your child's health care team; they will help you understand the treatment plan and how to help your child stay healthy.

II Type 2 diabetes

Type 2 diabetes starts as insulin resistance. This means your body can't use insulin efficiently. That stimulates your pancreas to produce more insulin until it can no longer keep up with demand. Insulin production decreases, which leads to high blood sugar. The exact cause of type 2 diabetes is unknown. Contributing factors may include:

- Genetics
- Lack of Exercise
- Being Overweight

There may also be other health factors and environmental reasons.

Diabetes, type 2: One of the two major types of diabetes, the type in which the beta cells of the pancreas produce insulin but the body is unable to use it effectively because the cells of the body are resistant to the action of insulin. Although this type of diabetes may not carry the same risk of death from ketoacidosis, it otherwise involves many of the same risks of complications as does type 1 diabetes (in which there is a lack of insulin).

The aim of treatment is to normalize the blood glucose in an attempt to prevent or minimize complications. People with type 2 diabetes may experience marked hyperglycemia, but many do not require insulin injections and can be treated with diet, exercise, and oral hypoglycemic agents (drugs taken by mouth to lower the blood sugar).

Type 2 diabetes requires good dietary control including the restriction of calories, lowered consumption of simple carbohydrates and fat with increased consumption of complex carbohydrates and fiber. Regular aerobic exercise is also an important method for treating both type 2 diabetes since it decreases insulin resistance and helps burn excessive glucose. Regular exercise also may help lower blood lipids and reduce some effects of stress, both important factors in treating diabetes and preventing complications.

Type 2 diabetes is also known as insulin-resistant diabetes, non-insulin dependent diabetes, and adult-onset diabetes.

- **Gestational diabetes**

Gestational diabetes is due to insulin-blocking hormones produced during pregnancy. This type of diabetes only occurs during pregnancy.

2.1 Pharmacological treatment

Old approaches to the treatment of this chronic progressive disease include diet modification and oral hypoglycemic medications, which have proven inadequate, while insulin therapy only solves the problem temporarily. Even with the newest pharmacotherapy, patients continue to develop macro- and micro vascular complications. Diabetes is associated with increased cardiac- and stroke-related deaths, kidney failure, blindness, and 60% of non-trauma lower-limb amputations (National diabetes fact sheet, Atlanta 2004). Alternative treatments targeting different models of this disease require careful and responsible examination. As shown below, apart from insulin treatment, it is possible to gain diabetes control after gastrointestinal bypass surgery

III Literature Survey

Kopitar, Leon, et al. "Early detection of type 2 diabetes mellitus using machine learning-based prediction models", In this research work studied differences in performance, calibration and interpretability of machine learning-based prediction models and multivariable regression models. Our results show that using new data in the EHR system to rebuild prediction models not only improves prediction performance, but also stability of the variable importance ranking, although not equally in different machine learning prediction models. Our results found no clinically relevant improvement when employing machine learning-based models over the more conventional regression models in terms of predictive performance. Even with calibration of the models, visualisation of the observed versus actual FPG showed some advantages in using simpler models. When observing the stability of variable ranking based on relative importance of variables, one can notice that a method like LightGBM results in much more stable results in comparison to other methods, which were more prone to high variance in variable importance. Both regression-based methods also proved as comparable alternatives. Since regression-based prediction models have been regularly used in clinical practice they could represent a better alternative in some clinical environments. The results in this study show significant improvement in terms of AUC, AUPRC and RMSE for all tested methods as the amount of collected data increases. For all tested predictive models in most of the experiments, we were able to show that additional data availability positively correlates with improved predictive performance and more stable variable importance-based ranking of variables. The opportunity of updating models arises as additional routine data become available over time. Future research needs to explore the

implementation of different approaches of building ensemble methods. In this case, stacking and blending of different prediction models could be taken into consideration. However, such systems bring along even more challenges in terms of interpreting the results that should support decisions of the healthcare experts[1].

Krishnan, Devi R., et al. "Evaluation of predisposing factors of Diabetes Mellitus post Gestational Diabetes Mellitus using Machine Learning Techniques.", In this study, we tested various ML classification models on 15 feature dataset of 77 patients. In clinical practice, predicting the risk of DM after GDM is a challenging task. In our research, we make use of prognostic machine learning (ML) techniques that inputs 15 predisposing factors and outputs the risk of DM. The results from our study demonstrated that Random Forest and Gaussian Naive Bayes classifies the data comparatively better than the other models[2].

Vigneswari, et.al. "Machine Learning Tree Classifiers in Predicting Diabetes Mellitus", Machine learning tree classifiers were analyzed for predicting diabetes mellitus in this study. The accuracy of these classifiers were calculated based on the True Positive, True Negative, False Positive, and False Negative of the classifiers. In this study, Logistic Model Tree (LMT) classifier achieved an accuracy of 79.31% with an average time of 0.49 sec to build the model which is higher than the Random Forest tree classifier with 78.54% accuracy and 0.04 sec. Machine learning tree classifiers other than the ones considered here can be analyzed in future for constructing hybrid tree classifier algorithms[3].

Alaassaf, Reem A. et.al., "Preemptive Diagnosis of Diabetes Mellitus Using Machine Learning" In this paper, Saudi Arabia medical data has been explored for the first time in diagnosing diabetes mellitus. Correlation coefficient and recursive feature elimination were used for feature selection, then four classification algorithms, namely: ANN, SVM, NB, and K-NN were compared on the basis of classification accuracy, precision, recall and f-measure. ANN outperformed the rest with classification accuracy of 77.5%. Potential future work could be in improving the classification accuracy by investigating different feature selection methods or classifiers. Moreover, decreasing the FN rate in misclassification, has high significance in the context of medicine[4].

Ioanniskava Kiotiset,[2017] "Machine Learning and Data Mining Methods in Diabetes Research", The growth of biotechnology and health sciences have led to a significant production of data, such as high throughput genetic data and clinical information, generated from large Electronic Health Records (EHRs). To this end, application of machine learning and data mining methods in biosciences is presently, more than ever before, vital and

indispensable in efforts to transform intelligently all available information into valuable knowledge. Diabetes mellitus (DM) is defined as a group of metabolic disorders exerting significant pressure on human health worldwide. Extensive research in all aspects of diabetes (diagnosis, etiopathophysiology, therapy, etc.) has led to the generation of huge amounts of data. The aim of the present study is to conduct a systematic review of the applications of machine learning, data mining techniques and tools in the field of diabetes research with respect to a) Prediction and Diagnosis, b) Diabetic Complications, c) Genetic Background and Environment, and e) Health Care and Management with the first category appearing to be the most popular[5].

PhattharatSongthung et.al, [2016] “Improving Type 2 Diabetes Mellitus Risk Prediction Using Classification”, Diabetes is a chronic disease that contributes to a significant portion of the healthcare expenditure for a nation as individuals with diabetes need continuous medical care. In order to prevent or delay the onset of type 2 diabetes, it is necessary to identify high risk populations and introduce behaviour modifications as early as possible. Screening the population to identify high risk individuals is an important task. One of the most accurate tests of diabetes is through the analysis of fasting blood sugar, but it is invasive and costly. Furthermore, it is only useful when the individual is already displaying symptoms i.e., making a diagnosis, which is considered too late to be an effective screening mechanism [6].

MadhuriPanwaret.al. [2016] ,“K-nearest neighbour basedmethod-ology for accurate diagnosis of diabetes mellitus”, Diabetes is one of the leading causes of death, disability and economic loss throughout the world. Type 2 diabetes is more common (90-95% worldwide) type of diabetes. However, it can be prevented or delayed by taking the right care and interventions which indeed an early diagnosis. There has been much advancement in the field of various machine learning algorithms specifically for medical diagnosis. But due to partially complete medical data sets, accuracy often decreases, results in more number of misclassification that can lead to harmful complications. An accurate prediction and diagnosis of a disease becomes a challenging research problem for many researchers. Therefore, aimed to improve the diagnosis accuracy we have proposed a new methodology, based on novel pre-processing techniques, and K-nearest neighbour classifier [7].

Purushottam et.al[2015] “Diabetes Mellitus Prediction System Evaluation Using C4.5 Rules and Partial Tree”.Diabetes disease prediction is a progressive area of research in the healthcare Sector. Although many data mining techniques have been employed to assess the main causes of diabetes, but only few sets of clinical risk factors are considered. Due to this, some important factors

like pre-diabetes health conditions are not considered in their analysis. So the results produced by such techniques may not represent appropriate diabetes pattern and risk factors appropriately. In this study, we have designed a system that can efficiently discover the rules to predict the risk level of patients based on the given parameter about their health. Then we evaluate and compare this system using C45 rules and partial tree. [8].

KiarashZahirniaet. al, [2015] “Diagnosis of type 2 diabetes using cost-sensitive learning”, Diabetes is the fourth cause of death in the world and has some complications such as amputation, visual impairment, kidney disorder and early death. 80% of diabetes symptoms is avoidable by early diagnosis. There are standard methods to diagnose diabetes by measuring the plasma glucose concentration. However, screening all people is impossible due to financial shortages especially in developing countries. Therefore, it is proposed that the people more than 20 years old who are prone to diabetes be tested. Identifying diabetic people is possible by using different methods including machine learning algorithms. Standard machines assume balance in data and use all the available and related features to achieve lower error rates. However, in medical applications misclassification cost should be minimized as misclassification costs for healthy and patient instances are different. In addition, we are facing imbalanced data in the most medical issues including diabetes diagnosis[9].

Bum Ju Lee et.al, [2014] “Prediction of Fasting Plasma Glucose Status Using Anthropometric Measures for Diagnosing Type 2 Diabetes”, It is well known that body fat distribution and obesity are important risk factors for type 2 diabetes. Prediction of type 2 diabetes using a combination of anthropometric measures remains a controversial issue. This study aims to predict the fasting plasma glucose (FPG) status that is used in the diagnosis of type 2 diabetes by a combination of various measures among Korean adults. A total of 4870 subjects (2955 females and 1915 males) participated in this study. Based on 37 anthropometric measures, we compared predictions of FPG status using individual versus combined measures using two machine-learning algorithms. The values of the area under the receiver operating characteristic curve in the predictions by logistic regression and naive Bayes classifier based on the combination of measures were 0.741 and 0.739 in females, respectively, and were 0.687 and 0.686 in males, respectively. Our results indicate that prediction of FPG status using a combination of anthropometric measures was superior to individual measures alone in both females andmales.We show that using balanced data of normal and high FPG groups can improve the prediction and reduce the intrinsic bias of the model toward the majority class [10].

SooBeom Choi et.al, [2014]. “Screening for Prediabetes Using Machine Learning Models”, The global prevalence of diabetes is rapidly increasing. Studies support the necessity of screening and interventions for

pre-diabetes, which could result in serious complications and diabetes. This study aimed at developing an intelligence-based screening model for pre-diabetes. Data from the Korean National Health and Nutrition Examination Survey (KNHANES) were used, excluding subjects with diabetes. The KNHANES 2010 data were used for training and internal validation, while data from KNHANES 2011 were used for external validation. We developed two models to screen for prediabetes using an artificial neural network (ANN) and support vector machine (SVM) and performed a systematic evaluation of the models using internal and external validation. We compared the performance of our models with that of a screening score model based on logistic regression analysis for prediabetes that had been developed previously. The SVM model showed the areas under the curve of 0.731 in the external datasets, which is higher than those of the ANN model (0.729) and the screening score model (0.712), respectively[11].

M. Veluet. al, “Visual data mining techniques for classification of diabetic patients”, [2013] Clustering is a data mining technique for finding important patterns in unorganized and huge data collections. The likelihood approach of clustering technique is quite often used by many researchers for classifications due to its' being simple and easy to implement. It uses Expectation-Maximization (EM) algorithm for sampling. The study of classification of diabetic patients was main focus of this research work. Diabetic patients were classified by data mining techniques for medical data obtained from Pima Indian Diabetes (PID) data set. This research was based on three techniques of EM Algorithm, h-means+ clustering and Genetic Algorithm (GA). These techniques were employed to form clusters with similar symptoms. Result analyses proved that h-means+ and double crossover genetics process based techniques were better on performance comparison scale. The simulation tests were performed on WEKA software tool for three models used to test classification. The hypothesis of similar patterns of diabetes case among PID and local hospital data was tested and found positive with correlation coefficient of 0.96 for two types of the data sets. About 35% of a total of 768 test samples were found with diabetes presence[12].

Kalliopi V. Dalakleidi et. al, [2013] “A hybrid genetic algorithm for the selection of the critical features for risk prediction of cardiovascular complications in Type 2 Diabetes patients”, The purpose of this study is to present a hybrid approach based on the combined use of a genetic algorithm (GA) and a nearest neighbours classifier for the selection of the critical clinical features which are strongly related with the incidence of fatal and non fatal Cardiovascular Disease (CVD) in patients with Type 2 Diabetes Mellitus (T2DM). For the development and the evaluation of the proposed algorithm, data from the medical records of 560 patients with T2DM are used. The best subsets of features proposed by the implemented algorithm include the most common risk factors, such as

age at diagnosis, duration of diagnosed diabetes, glycosylated haemoglobin (HbA1c), cholesterol concentration, and smoking habit, but also factors related to the presence of other diabetes complications and the use of antihypertensive and diabetes treatment drugs (i.e. proteinuria, calcium antagonists, b-blockers, diguanides and insulin). The obtained results demonstrate that the best performance was achieved when the weighted k-nearest neighbours' classifier was applied to the CVD dataset with the best subset of features selected by the GA, which resulted in high levels of accuracy (0.96), sensitivity (0.80) and specificity (0.98). [13]

Abdulfatai B. Olokoba et. al, [2012] “Type 2 Diabetes Mellitus: A Review of Current Trends” Type 2 diabetes mellitus (DM) is a chronic metabolic disorder in which prevalence has been increasing steadily all over the world. As a result of this trend, it is fast becoming an epidemic in some countries of the world with the number of people affected expected to double in the next decade due to increase in ageing population, thereby adding to the already existing burden for healthcare providers, especially in poorly developed countries. This review is based on a search of Medline, the Cochrane Database of Systemic Reviews, and citation lists of relevant publications. Subject heading and key words used include type 2 diabetes mellitus, prevalence, current diagnosis, and current treatment. Only articles in English were included. Screening and diagnosis is still based on World Health Organization (WHO) and American Diabetes Association (ADA) criteria which include both clinical and laboratory parameters. [14]

David B. Sacks et. al, [2011] “Guidelines and Recommendations for Laboratory Analysis in the Diagnosis and Management of Diabetes Mellitus”, Multiple laboratory tests are used in the diagnosis and management of patients with diabetes mellitus. The quality of the scientific evidence supporting the use of these assays varies substantially. An expert committee compiled evidence-based recommendations for the use of laboratory analysis in patients with diabetes. A new system was developed to grade the overall quality of the evidence and the strength of the recommendations. [16]

Ping Zhang et.al, [2010] “Global healthcare expenditure on diabetes for 2010 and 2030”, Country-by-country expenditures for 193 countries, expressed in United States Dollars (USD) and in International Dollars (ID), were estimated based on the country's age- sex specific diabetes prevalence and population estimates, per capita health expenditures, and health expenditure ratios per person with and without diabetes. Diabetes prevalence was estimated from studies in 91 countries. Population estimates and health expenditures were from the United Nations and the World Health Organization. The health expenditure ratios were estimated based on utilization and cost data of a large health plan in the U.S. Diabetes expenditures for the year 2030 were projected by

considering future changes in demographics and urbanization. Results: [17]

M. Patilet. al, [2010] “Association rule for classification of type -2 diabetic patients”, The discovery of knowledge from medical databases is important in order to make effective medical diagnosis. The aim of data mining is extract the information from database and generate clear and understandable description of patterns. In this study we have introduced a new approach to generate association rules on numeric data. We propose a modified equal width binning interval approach to discretizing continuous valued attributes. The approximate width of the desired intervals is chosen based on the opinion of medical expert and is provided as an input parameter to the model. First we have converted numeric attributes into categorical form based on above techniques. Apriori algorithm is usually used for the market basket analysis was used to generate rules on Pima Indian diabetes data. The data set was taken from UCI machine learning repository containing total instances 768 and 8 numeric attributes..[18]

HelenaW. Rodbardet. al, [2009] “Trends in Method of Diagnosis of Type 2 Diabetes Mellitus: Results from SHIELD”, Aims. This study assessed whether recent screening recommendation have led to increased diagnosis of type 2 diabetes mellitus (T2DM) through routine screening. Methods. Respondents to the 2006 US SHIELD survey reported whether a physician told them they had T2DM, age at diagnosis, specialty of the physician who made the diagnosis, and whether the diagnosis was made after having symptoms, during routine screening, or when being treated for another health problem. Results. Of 3 022 T2DM respondents, 36% of respondents reported that T2DM diagnosis was made during routine screening alone, 20% after having symptoms alone, and 6% when being treated for another health problem alone. [19]

Rajeeb Dey et. al, [2008] “Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus” The present work is a classification problem applied to diagnosis of diabetes mellitus using back propagation algorithm of artificial neural network (ANN). The data base used for training and testing the ANN has been collected from Sikkim Manipal Institute of Medical Sciences Hospital, Gangtok, Sikkim for the diabetic patients of the state of Sikkim. This work is validated by comparing the network response with the actual hospital data of Sikkim Manipal Institute of Medical Sciences for the first time. This work is a step towards development of prototype system for knowledge classification in this field. [20]

SarojiniBalakrishnanet. al, [2008] “SVM ranking with backward search for feature selection in type II diabetes databases”, Clinical databases have accumulated large quantities of information about patients and their clinical history. Data mining is the search for relationships and patterns within this data that could provide useful

knowledge for effective decision-making. Classification analysis is one of the widely adopted data mining techniques for healthcare applications to support medical diagnosis, improving quality of patient care, etc. Usually medical databases are high dimensional in nature. If a training dataset contains irrelevant features (i.e., attributes), classification analysis may produce less accurate results. Data pre-processing is required to prepare the data for data mining and machine learning to increase the predictive accuracy. Feature selection is a pre-processing technique commonly used on high-dimensional data and its purposes include reducing dimensionality, removing irrelevant and redundant features, reducing the amount of data needed for learning, improving algorithms' predictive accuracy, and increasing the constructed models' comprehensibility. Much research work in data mining has gone into improving the predictive accuracy of the classifiers by applying the techniques of feature selection. [21]

IV. CONCLUSIONS

In this survey paper discuss the different method of diabetes patient analysis. There are number of different methods available for the analysis of diabetes patients on the basis of patient physical properties such as weight, body mass index, blood pressure and other parameters. There are different classification analysis method available diabetes patient such J-45, support vector machine, multi support vector, navbais and neural network. In this survey paper discuss the different methods of clustering and classification of diabetes patients.

REFERENCES

- [1] Kopitar, Leon, Primoz Kocbek, Leona Cilar, Aziz Sheikh, and Gregor Stiglic. "Early detection of type 2 diabetes mellitus using machine learning-based prediction models." Scientific reports 10, no. 1 (2020): 1-12.
- [2] Krishnan, Devi R., Gayathri P. Menakath, Anagha Radhakrishnan, Yarrangangu Himavarshini, A. Aparna, Kaveri Mukundan, Rahul Krishnan Pathinarupothi, Bithin Alangot, Sirisha Mahankali, and Chakravarthy Maddipati. "Evaluation of predisposing factors of Diabetes Mellitus post Gestational Diabetes Mellitus using Machine Learning Techniques." In 2019 IEEE Student Conference on Research and Development (SCOREd), pp. 81-85. IEEE, 2019.
- [3] D. Vigneswari1 , N. Komal Kumar , V. Ganesh Raj , A. Gungan , S. R.Vikash, “Machine Learning Tree Classifiers in Predicting Diabetes Mellitus”, 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 2019.
- [4] Alassaf, Reem A., Khawla A. Alsulaim, Noura Y. Alroomi, Nouf S. Alsharif, Mishael F. Aljubeir, Sunday O. Olatunji, Alaa Y. Alahmadi, Mohammed Imran, Rahma A. Alzahrani, and Nora S. Alturayef. "Preemptive Diagnosis of Diabetes Mellitus Using

- Machine Learning." In 2018 21st Saudi Computer Society National Computer Conference (NCC), pp. 1-5. IEEE, 2018.
- [5] Kavakiotis, Ioannis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, and Ioanna Chouvarda. "Machine learning and data mining methods in diabetes research." Computational and structural biotechnology journal (2017).
- [6] Songthung, Phattharat, and Kunwadee Sripanidkulchai. "Improving type 2 diabetes mellitus risk prediction using classification." In Computer Science and Software Engineering (JCSSE), 2016 13th International Joint Conference on, pp. 1-6. IEEE, 2016.
- [7] Panwar, Madhuri, Amit Acharyya, Rishad A. Shafik, and Dwaipayan Biswas. "K-nearest neighbor based methodology for accurate diagnosis of diabetes mellitus." In Embedded Computing and System Design (ISED), 2016 Sixth International Symposium on, pp. 132-136. IEEE, 2016.
- [8] Saxena, Kanak, and Richa Sharma. "Diabetes mellitus prediction system evaluation using c4. 5 rules and partial tree." In Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions), 2015 4th International Conference on, pp. 1-6. IEEE, 2015.
- [9] Zahiria Kiarash, Mehdi Teimouri, Rohallah Rahmani, and Amin Salaq. "Diagnosis of type 2 diabetes using cost-sensitive learning." In Computer and Knowledge Engineering (ICCKE), 2015 5th International Conference on, pp. 158-163. IEEE, 2015.
- [10] Lee, Bum Ju, Boncho Ku, Jiho Nam, Duong Duc Pham, and Jong Yeol Kim. "Prediction of fasting plasma glucose status using anthropometric measures for diagnosing type 2 diabetes." IEEE journal of biomedical and health informatics 18, no. 2 (2014).
- [11] Choi, SooBeom, Won Jae Kim, Tae Keun Yoo, JeeSoo Park, Jai Won Chung, Yong-ho Lee, EunSeok Kang, and Deok Won Kim. "Screening for prediabetes using machine learning models." Computational and mathematical methods in medicine 2014 (2014).
- [12] Velu, C. M., and K. R. Kashwan. "Visual data mining techniques for classification of diabetic patients." In Advance Computing Conference (IACC), 2013 IEEE 3rd International, pp. 1070-1075. IEEE, 2013.
- [13] Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, 2003, pp. 459-469.
- [14] B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in Mobile computing, ed: Springer, 1996, pp. 153-181.
- [15] S. Ananth, R. Nithya "Extemporize Data Transmission Through Estimating Queue Length And Energy Efficient Clustering Method" ISSN: 2395-5325 Volume 1, Issue 2 (May'2015)
- [16] Sacks, David B., Mark Arnold, George L. Bakris, David E. Bruns, Andrea Rita Horvath, M. Sue Kirkman, Ake Lernmark, Boyd E. Metzger, and David M. Nathan. "Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus." Clinical chemistry 57, (2011).
- [17] Zhang, Ping, Xinzhi Zhang, Jonathan Brown, Dorte Vistisen, Richard Sicree, Jonathan Shaw, and Gregory Nichols. "Global healthcare expenditure on diabetes for 2010 and 2030." Diabetes research and clinical practice 87, no. 3 (2010): 293-301.
- [18] B. M. Patil, R. C. Joshi and Durga Toshniwal, "Association rule for classification of type -2 diabetic patients", Second International Conference on Machine Learning and Computing, 2010.
- [19] Rodbard, Helena W., Andrew J. Green, Kathleen M. Fox, and Susan Grandy. "Trends in method of diagnosis of type 2 diabetes mellitus: results from SHIELD." International journal of endocrinology 2009 (2009).
- [20] Rajeeb Dey, Vaibhav Bajpai, Gagan Gandhi and Barnali Dey, "Application of Artificial Neural Network (ANN) technique for Diagnosing Diabetes Mellitus", IEEE Region 10 and the Third international Conference on Industrial and Information Systems, Dec 2008.
- [21] Balakrishnan Sarojini, Ramaraj Narayanaswamy, Nickolas Savarimuthu, and Rita Samikannu. "SVM ranking with backward search for feature selection in type II diabetes databases." In Systems, Man and Cybernetics, 2008. SMC 2008. IEEE International Conference on, pp. 2628-2633. IEEE, 2008.
- [22] Liang Lily, Vinay Mandal, Yi Lu, and Deepak Kumar. "Multi-dimensional cluster misclassification test for pathway differential analysis of diabetes." In Computer and Computational Sciences, 2007. Second International Multi-Symposiums on, pp. 84-91. IEEE, 2007.
- [23] Aekplakorn, Wichai, Pongamorn Bunnag, Mark Woodward, Piyamitr Sritara, Sayan Cheepudomwitt, Sukit Yamwong, Tada Yipintsoi, and Rajata Rajatanavin. "A risk score for predicting incident diabetes in the Thai population." Diabetes care 29, no. 8 (2006): 1872-1877.
- [24] Lindström, Jaana, and Jaakko Tuomilehto. "The diabetes risk score." Diabetes care 26, no. 3 (2003): 725-731.
- [25] Tuomilehto, Jaakko, Jaana Lindström, Johan G. Eriksson, Timo T. Valle, Helena Hämäläinen, Pirjo Ilanne-Parikka, Sirkka Keinänen-Kiukaanniemi et al. "Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance." New England Journal of Medicine 344, no. 18 (2001): 1343-1350.