PSO Based Feature Optimization for Diabetes Detection using Machine Learning

Shivani Nai[1], Mr.Nagendra Kumar[2], Mr. Prateek Gupta[3]

[1]Reseacrh Scholar, Department of CSE, Shri ram Institute of Science and Technology, Jabalpur, M.P.
[2]Professor, Departmet of CSE, Shri Ram Institute of Science and Technology, Jabalpur, M.P.
[3]Professor, Departmet of CSE, Shri Ram Institute of Science and Technology, Jabalpur, M.P.

---***---

Abstract – Diabetes is a chronic disease that occurs when the body cannot use insulin properly or the pancreas cannot produce enough hormones to control blood sugar. High blood sugar is a symptom of diabetes, a group of metabolic diseases. The two most common types of diabetes are type 1 and type 2, but there are other types, such as gestational diabetes, which occur during pregnancy. There is an increase in the number of type 1 diabetic patients. The genetic disease called type 1 diabetes has a long incubation period and usually manifests itself early in life. People with type 2 diabetes have cells that do not respond to insulin. It changes over time and mostly depends on people's lifestyle. According to the 2022 report of the international diabetes federation, approximately 382 million people worldwide are currently living with diabetes. This number is expected to increase to 592 million in 2035. One of the most common causes of tissue and organ damage and dysfunction, including blindness, kidney failure, heart failure, and stroke, is diabetes. Therefore, early diagnosis of diabetes is important. This project focuses on the use of machine learning such as logistic regression, knn, svc, ann, and random forest to predict diabetes. Each algorithm is calculated to determine the accuracy of the model. Additionally, the highest accuracy of 96.00% was achieved in logistic regression using the pso-based optimization technique.

Keywords: Diabetes Detection, Blood Glucose, Feature Selection, Machine Learning, deep Learning, PSO, Accuracy.

1. INTRODUCTION

Diabetes is a non-communicable disease that affects the control of blood sugar levels in the body. Blood glucose concentration is normally regulated by insulin and glucagon, two hormones secreted by the beta (β) and alpha (α) cells of the pancreas, respectively. The normal release of the two hormones regulates blood sugar levels in the body between 70 - 180 mg/dl (4.0 - 7.8 mmol/l). Insulin lowers blood sugar, while glucagon increases blood sugar. However, abnormality of these hormones can lead to diabetes. There are many types of diabetes, many different types, but the most common types are type 1 diabetes, type 2 diabetes and gestational diabetes (gdm). Type 1 diabetes is more common in children; while type 2 diabetes is more common in adults and the elderly, gdm is more common in women and is diagnosed during pregnancy. While insulin secretion does not work in type 1 diabetes due to the destruction of pancreatic beta cells, there is a disorder in insulin secretion and function in type 2 diabetes. Gdm is a glucose intolerance first diagnosed during pregnancy; it may be mild, but it is also associated with high blood sugar and high insulin levels during pregnancy. All of these types can cause a lack of blood sugar in the body, which can lead to serious diseases in the body. In other words, when blood sugar rises above normal, this condition is called hyperglycemia. On the other hand, when it decreases and falls below normal, the condition is called hypoglycemia [3]-[5]. For example, hyperglycemia can cause chronic problems and lead to kidney disease, retinopathy, heart disease and other tissue damage, while hypoglycemia can cause temporary short-circuits. It can cause kidney disease, retinopathy, heart and heart disease, and other tissue damage can lead to diabetic coma [1] , [3] , [4] . Diabetes has become an important health problem in today's world due to its prevalence in children and adults. According to [6] , [7] , approximately 8.8% of adults worldwide had diabetes in 2015, and this number was approximately 415 million and is expected to reach approximately 642 million by 2040. During this time, more than 500,000 children were killed and approximately 5 million people died. On the other hand, the global economic burden of diabetes was estimated to be approximately 673 billion dollars in 2015, and is expected to reach 802 billion dollars in 2040. Self-monitoring of blood glucose (smbg) using fingertip blood glucose meters is a diabetes treatment method introduced three years ago [8], [9]. In this way, diabetics invasively measure their blood sugar levels three to four times a day by using a thumb glucometer to prick the skin of their fingers. The idea is to provide this: to increase insulin resistance. However, in addition to being laborious and laborious, this method can only be understood if insulin estimation is obtained from small smbg samples. In other words, this may cause the blood sugar in the blood to be higher than normal. To overcome this problem, continuous blood glucose monitoring (cgm) has been introduced, which can provide maximum information about changes in blood sugar within a few days, allowing a good treatment decision to be made for people with diabetes. In this way, blood sugar concentration is constantly monitored thanks to small devices/systems that monitor the glucose level in the blood environment. These systems can be invasive, minimally invasive or noninvasive. Moreover, cgm systems can be divided into two types: retrograde systems and immediate systems [10]. The introduction and availability of new types of cgm devices/machines have brought new opportunities for diabetics to easily manage their diabetes. Most cgm devices today often use a minimally invasive device to calculate and record the patient's current blood sugar every minute by measuring interstitial fluid (isf). These systems/devices have little effect because they damage the skin but not the blood vessels. There are also non-invasive methods to measure blood glucose concentration, such as using electrical current through the skin into blood vessels in the body [11].

Additionally, e-health in the form of telemedicine not only allows doctors to see patients regularly remotely, but also to send cgm to the hospital's remote database to predict hypoglycemia/hyperglycemia and other complications in diabetes management. One of the challenges of diabetes management is the prevention of hypo/hyperglycemic events; this can be overcome by estimating blood sugar levels based on cgm/smbg and other methods (e.g. Exercise, diet, insulin, etc.). Therefore, it is important to develop tools for processing and interpreting cgm/smbg and blood-related data to obtain future blood glucose results. For this purpose, data mining plays an important role in the development of tools such as diabetes diagnosis and prediction [12], [13]. Data mining is the process of extracting important information from large amounts of data to discover previously unknown patterns, patterns, and relationships that can be used to develop predictive models [14]. In the literature, different data mining-based blood sugar prediction methods and methods have been developed with various models. This technology extracts, analyzes and interprets diabetes data to make medical decisions. A general flow is shown below:

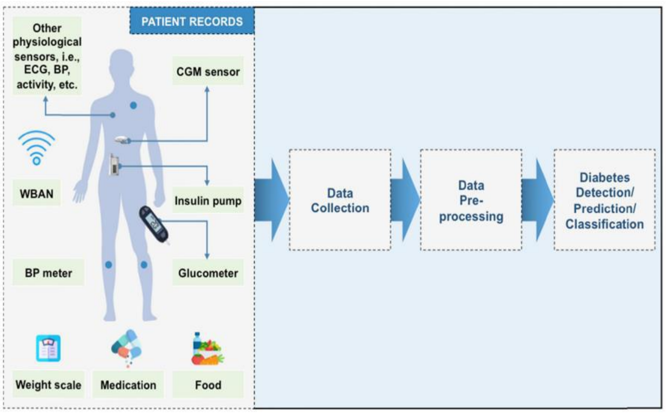

Fig.1.1: A generic flow of diabetes detection and prediction techniques [15].

The motivation for using machine learning for diabetes diagnosis stems from its ability to improve the accuracy and efficiency of diabetes diagnosis, prediction, and management. Diabetes mellitus is a type of metabolic disease caused by high blood sugar levels, and early diagnosis and intervention are crucial for effective treatment and prevention of complications. Here are some motivations for using machine learning for diabetes diagnosis:

1. Early detection and prevention: machine learning models can analyze large amounts of data, including patient data, genetics, and lifestyle data, to identify patterns that indicate diabetes risk. Early detection leads to timely intervention, changes in life, and prevention of infection. Improved accuracy: machine learning algorithms can process different data such as genetic data, medical history and lifestyle. By considering many variables, these models can provide more accurate predictions and help doctors make informed decisions. Personalized medicine: machine learning can create personalized treatment plans based on patient information. This approach takes into account the unique characteristics of each patient, allowing interventions to better control diabetes and improve overall health. It can improve early diagnosis, increase accuracy, advance personalized medicine, and provide better and more effective treatment for people with diabetes.

2. Related Work

2.1 Diabetic treatments using AI Tools

Smart devices have the potential to revolutionize the management and treatment of diabetes by providing a personalized, data-driven approach [16]. Below are a few applications of wisdom in diabetes management:

1. It provides insight into different patterns and predicts the next stage of diabetes. This allows patients and doctors to make better decisions. This helps diabetics and healthcare professionals manage diabetes. Effective use of insulin [17]

-closed-loop system: artificial intelligence-supported closed-loop insulin delivery system instantly automatically adjusts the insulin dose according to blood sugar. This system is designed to control blood sugar in a single target without having to constantly deal with it manually. Recommends the best diet for insulin sensitivity. Medication compliance and personal care planning

-artificial intelligence supported applications: mobile applications equipped with artificial intelligence can remind patients to take medications, monitor compliance, and provide recommendations for lifestyle changes based on personal health information. -Personalized treatment plan: artificial intelligence can analyze various patient data, including genetic information, to create a personalized treatment plan. This will help improve drug selection and dosage to achieve good results [18]. Early detection of problems - retinal scanning: artificial intelligence algorithms can analyze retinal images to detect early signs of diabetic retinopathy, a complication of diabetes. Early diagnosis allows for timely intervention and prevention of blindness [19]. This leads to early intervention to prevent the problem. Patient education and support

-chatbots and virtual assistants: ai-powered chatbots can provide ongoing support to diabetic patients, provide information, answer questions, and help manage lifestyle. Patient behavior data to identify patterns to help doctors tailor education and support to meet specific needs and problems. This should be done in collaboration with healthcare providers to ensure a balance between technology and personal care. Regulatory and ethical issues also need to be considered to ensure the safety and effectiveness of ai applications in healthcare.

2.2 Detection of diabetic behaviour using machine learning

Machine learning algorithms can be used to diagnose diabetes by analyzing different types of data, including clinical data, diagnoses, and even non-traditional data such as image data [20] [21]. Here are some ways to use machine learning to diagnose diabetes:

1. Medical data analysis:

- predictive models: machine learning models can be trained based on medical history data to predict diabetes risk. These models often take into account factors such as age, body mass index (bmi), family history and lifestyle. . This will include information about blood sugar levels, hba1c values and other clinical signs. Lab results:

- blood tests and biomarkers: machine learning can analyze blood test results and diabetes biomarkers such as increased blood sugar, hba1c levels, and lipid profile. Algorithms identify patterns and patterns that may indicate diabetes or diabetes. Imaging data:

- retinal imaging: machine learning algorithms can analyze retinal images to detect signs of diabetic retinopathy, an eye disease associated with diabetes. Early diagnosis of retinopathy is important to prevent vision loss.

4. Genetic information:

- Genomic analysis: machine learning can analyze genetic information to identify genetic markers or patterns associated with diabetes risk. This method is part of the development of precision medicine. Mobile health (mHealth) data:

-Wearable: machine learning algorithms can analyze data from wearable devices that track physical activity, sleep patterns, and other life factors. This information will be helpful in assessing diabetes risk. Behavior analysis:

- Health behavior monitoring: machine learning can analyze behavioral data such as food intake, physical activity levels, and medication adherence. Changes in behavior may indicate the need for further evaluation. Federated data models:

- Integration of multiple data sources: many machine learning models benefit from combining data from multiple sources. Comprehensive models that take into account clinical data, genetics, and lifestyle factors can provide more accurate prediction of diabetes risk. It is important to remember that machine learning can be a valuable tool in diagnosing diabetes, but it must be supported by clinical information. Interpreting patient care and making decisions is best accomplished through the collaboration of healthcare professionals and intelligent tools. Additionally, ethical considerations and patient identity should be prioritized when developing and implementing machine learning models for clinical use. The collaboration was carried out as shown in [22], which investigated and analyzed methods based on diabetes diagnosis, prediction and classification. Various data mining techniques for diabetes diagnosis are reviewed and discussed in [23]. Similarly, a review of the use of data mining and related data, methods, software, and methods in diabetes treatment was conducted in [24] . Based on this review, it was determined that data mining plays an important and scientifically proven role in glycemic control. The data is used only to extract important information from blood sugar data, ultimately helping diabetics control glycemic control. Similarly, in [25], the use of different data mining methods, including neural networks (ann), for the assessment and classification of diabetes was investigated. The study found that ann outperformed other methods with 89% prediction accuracy. And support vector machines (svm) to evaluate diabetes metrics using noise-free and noise-free data from the university of california, irvine (uci) machine learning data repository [27] . From the comparison of these technologies, it can be seen that the j48 classifier performs better when the data is noisy, with 73.82% accuracy. In noisy data, knn (k=1) and random forest are better than the other two methods with 100% accuracy. Also, in [18], a comparison of nine different methods based on diabetes prediction from the pimaindian diabetes dataset (pidd) from the uci machine learning repository was made with the help of data paper search tools such as weka, tanagra and Matlab [27]. According to the performance analysis, the best products in weka, Tanagra, and Matlab were j48graft, nb, and adaptive neuro-fuzzy inference system (anfis) with 81.33%, 100%, and 78.79% accuracy, respectively. Similarly, comparison and performance evaluation of various materials are presented in [28].

3. Proposed Work

In this section the details are given about the proposed idea based on the feature engineering for detection of diabetic patients. The thesis is being developed in python language. All the steps of machine learning like data preprocessing, feature analysis, feature extraction and model building is used in this work. The steps used for the proposed model is detailed below along with the architecture.

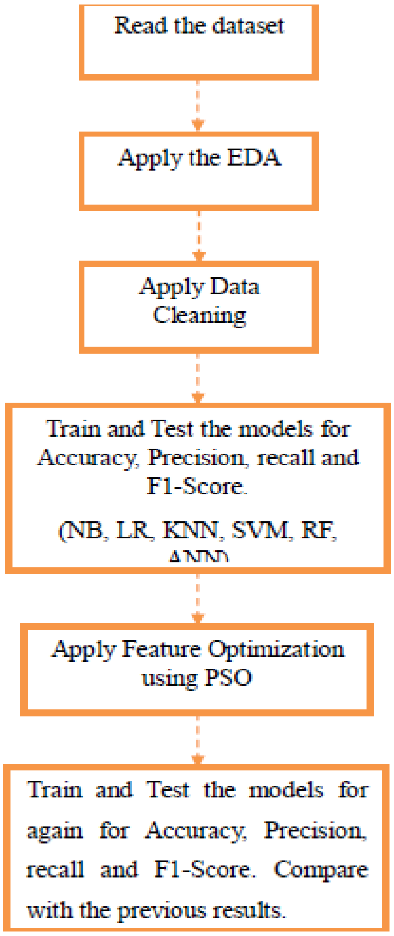

Fig. 3.1: Proposed model or diabetes detection.

Following data cleaning steps are applied:

1. Replace drinker to 1 and non-drinker to 0.

2. Replace Type of Diabetes T1DM to 1 and T2DM to 0.

3. Replace in Acute Diabetic Complications from none to 0 and diabetic ketoacidosis to 1.

4. Apply Label Encoding to column Alcohol drinking history by replacing 1 with "drinker" and 0 with "non-drinker".

5. Apply Label Encoding to column "Type of Diabetes", by replacing 1 with "T1DM" and 2 with "T2DM".

6. Apply Label Encoding to column "Acute Diabetic Complications", by replacing diabetic ketoacidosis with 1 and none with 0.

7. Apply one hot encoding on Diabetic Microvascular Complications.

3.1 Proposed Algorithm

After the EDA and the feature engineering is completed, training and testing will start.

Start.

1. Put all columns inx except "Type of Diabetes".

2. Divide the data into train set and test set with size=0.2.

3. Apply Gaussian Naive Bayes and print the results.

4. Apply KNN and print the results.

5. Apply Logistic Regression and print the results.

6. Apply Random Forest and print the results.

7. Apply ANN and print the results.

8. Apply Feature Optimization using PSO algorithm.

9. Repeat step 3 to 7.

10. Compare the Results.

11. Exit.

End.

4. Result Analysis

After the EDA, now we will build our model using machine learning classifiers. The snapshots are show below:

Algorithm Used	Using all 217 features Without Optimization (accuracy in %)	Using Optimized Features with PSO (Accuracy in %)
Gaussian Naive Bayes	48	55
KNN	84	91
Logistic Regression	88	96
Random forest	84	89
SVC	88	88
ANN	84	78

Fig. 4.1: comparisons of Results in terms of accuracy.

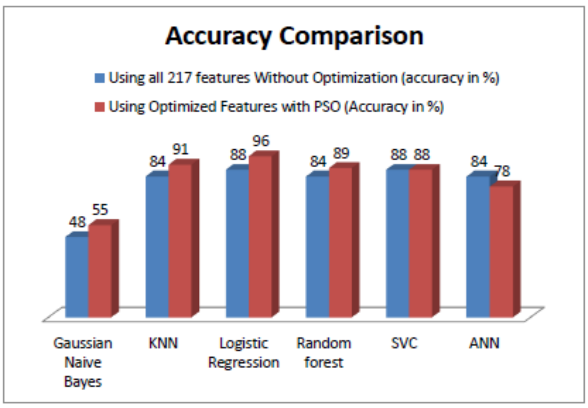

Fig. 4.2: Accuracy comparison chart.

5. Conclusion

In conclusion, leveraging machine learning for the detection of Type-2 diabetes presents a promising avenue in the field of healthcare. Through the analysis of various patient parameters such as demographic information, medical history, and biomarkers, machine learning algorithms can efficiently identify individuals at risk or already affected by Type-2 diabetes.

By utilizing robust datasets and employing diverse machine learning techniques such as logistic regression, decision trees, support vector machines, and neural networks, researchers and healthcare professionals can develop accurate and reliable predictive models for early detection and intervention.

Moreover, the integration of advanced technologies like wearable devices and remote monitoring systems can further enhance the effectiveness of these models by providing real-time data streams for continuous monitoring and analysis.

However, it's important to acknowledge the challenges associated with the implementation of machine learning in clinical settings, including data privacy concerns, model interpretability, and the need for rigorous validation and regulatory approval.

6. References

[1] P. Dua, F. J. Doyle, and E. N. Pistikopoulos, “Model-based blood glucose control for type 1 diabetes via parametric programming,'' IEEE Trans.Biomed. Eng., vol. 53, no. 8, pp. 1478_1491, Aug. 2006.

[2] American Diabetes Association, “2. Classification and diagnosis of diabetes: Standards of medical care in diabetes_2020,'' Diabetes Care, vol. 43, no. 1, pp. S14_S31, Jan. 2020.

[3] G. Sparacino, F. Zanderigo, S. Corazza, A. Maran, A. Facchinetti, and C. Cobelli, “Glucose concentration can be predicted ahead in time from continuous glucose monitoring sensor time-series,'' IEEE Trans. Biomed.Eng., vol. 54, no. 5, pp. 931_937, May 2007.

[4] S. Guerra, A. Facchinetti, G. Sparacino, G. D. Nicolao, and C. Cobelli, “Enhancing the accuracy of subcutaneous glucose sensors: A real-time deconvolution-based approach,'' IEEE Trans. Biomed. Eng., vol. 59, no. 6, pp. 1658_1669, Jun. 2012.

[5] J. M. Norris, R. K. Johnson, and L. C. Stene, “Type 1 diabetes_Early life origins and changing epidemiology,'' Lancet Diabetes Endocrinol., vol. 8, no. 3, pp. 226_238, Mar. 2020.

[6] National Diabetes Statistics Report, 2020. Accessed: Jan. 15, 2021. [Online]. Available: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

[7] ID Federation. IDF DIABETES ATLAS 9th Edition 2019. Accessed: Jan. 15, 2021. [Online]. Available: https://diabetesatlas.org/en/.

[8] L. Olansky and L. Kennedy, “Finger-stick glucose monitoring: Issues of accuracy and specificity,'' Diabetes Care, vol. 33, no. 4, pp. 948_949, Apr. 2010.

[9] J. B. Buse, D. J.Wexler, A. Tsapas, P. Rossing, G. Mingrone, C. Mathieu, D. A. D'Alessio, and M. J. Davies, “2019 update to: Management of hyperglycaemia in type 2 diabetes, 2018. A consensus report by the American diabetes association (ADA) and the European association for the study of diabetes (EASD),'' Diabetologia, vol. 63, no. 2, pp. 221_228, Feb. 2020.

[10] M. Langendam, Y. M. Luijf, L. Hooft, J. H. D. Vries, A. H. Mudde, and R. J. Scholten, “Continuous glucose monitoring systems for type 1 diabetes mellitus,'' Cochrane Database Syst. Rev., vol. 2012, no. 1, pp. 1_144, 2012, Art. No. CD008101.

[11] C. Choleau, J. C. Klein, G. Reach, B. Aussedat, V. Demaria-Pesce, G. S.Wilson, R. Gifford, and W. K.Ward, “Calibration of a subcutaneous amperometric glucose sensor: Part 1. Effect of measurement uncertainties on the determination of sensor sensitivity and background current,'' Biosensors Bioelectronics, vol. 17, no. 8, pp. 641_646, Aug. 2002.

[12] D. Sisodia and D. S. Sisodia, “Prediction of diabetes using classification algorithms,'' Procedia Comput. Sci., vol. 132, pp. 1578_1585, Jun. 2018.

[13] H. Kaur and V. Kumar, “Predictive modelling and analytics for diabetes using a machine learning approach,'' Appl. Comput. Inform.vol. 16, pp. 1_11, Jul. 2020.

[14] K. Kincade, “Data mining: Digging for healthcare gold,'' Insurance Technol., vol. 23, no. 2, no. 2, pp. 2_7, 1998.

[15] F. A. Khan, K. Zeb, M. Al-Rakhami, A. Derhab and S. A. C. Bukhari, "Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review," in IEEE Access, vol. 9, pp. 43711-43735, 2021, doi: 10.1109/ACCESS.2021.3059343.

[16] P. Maiti and C. Chakravarty, "AI Based Automated Detection & Classification of Diabetic Retinopathy," 2023 7th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/CSITSS60515.2023.10334199.

[17] T. Pala and A. Y. Camurcu, "Evaluation of data mining classification and clustering techniques for diabetes,'' Malaysian J. Comput., vol. 2, no. 1, pp. 1_9, 2014.

[18] R. M. Rahman and F. Afroz, "Comparison of various classification techniques using different data mining tools for diabetes diagnosis,'' J. Softw. Eng. Appl., vol. 6, no. 3, p. 85, 2013.

[19] V. Karthikeyan, I. P. Begum, K. Tajudin, and I. S. Begam, "Comparative of data mining classification algorithm (CDMCA) in diabetes disease prediction,'' *Int. J. Comput.* Appl., vol. 60, no. 12, pp. 26_31, Dec. 2012.

[20] P. Thirumal and N. Nagarajan, "Utilization of data mining techniques for diagnosis of diabetes mellitus: A case study,'' ARPN J. Eng. Appl. Sci., vol. 10, no. 1, pp. 8_13, Jan. 2015.

[21] G. Visalatchi, S. J. Gnanasoundhari, and M. Balamurugan, "A survey on data mining methods and techniques for diabetes mellitus,'' Int. J. Comput. Sci. Mobile Appl., vol. 2, no. 2, pp. 100_105, 2014.

[22] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research,'' Comput. Struct. Biotechnology. J., vol. 15, pp. 104_116, Jan. 2017.

[23] D. Tomar and S. Agarwal, "A survey on data mining approaches for healthcare,'' Int. J. Bio-Sci. Bio-Technol., vol. 5, no. 5, pp. 241_266, Oct. 2013.

[24] M. Marinov, A. S.M. Mosa, I. Yoo, and S. A. Boren, "Data-mining technologies for diabetes: A systematic review,'' J. Diabetes Sci. Technol., vol. 5, no. 6, pp. 1549_1556, Nov. 2011.

[25] M. Durairaj and K. Priya, "Breast cancer prediction using soft computing techniques a survey,'' Int. J. Comput. Sci. Eng., vol. 6, no. 8, pp. 135_145, Aug. 2018.

[26] J. P. Kandhasamy and S. Balamurali, "Performance analysis of classifier models to predict diabetes mellitus,'' Procedia Comput. Sci., vol. 47, pp. 45_51, May 2015.

[27] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools,'' Energy Buildings, vol. 49, pp. 560_567, Jun. 2012.

[28] L. Tapak, H. Mahjub, O. Hamidi, and J. Poorolajal, "Real-data comparison ofdata mining methods in prediction of diabetes in Iran,'' Healthcare Informat. Res., vol. 19, no. 3, no. 3, pp. 177_185, 2013.