Predicting the Severity of Diabetes Using ECLAT Algorithm in Data Mining

P. Senthil Kumari [1]

[1] Research Supervisor, Alagappa University, Karaikudi, Tamil Nadu 630003, India
senthilmathimca@gmail.com

Abstract. Diabetes is a long-term disease that damages the various parts of the human body. According to the World Health Organization (WHO) report, there was a deficiency of 12.9 million medical care workers estimated in 2035. Various Artificial Intelligence (AI) and Machine Learning (ML) classifiers were used to anticipate and diagnose diabetes. In the case of unsupervised ML classifications, data mining acts an important role in the diagnosis and prediction of the disease. Selecting legitimate classifiers clearly expands the correctness and adeptness of the proposed system. Public awareness of the disease is very poor in India. Deficient healthcare facilities lead to the growth of the disease in families. Apriori, FP growth, and ECLAT are the different types of association rule mining algorithms for the diagnosis and prediction of diabetes. Equivalence Class clustering and bottom-up LAttice Traversal (ECLAT) algorithm is used for the prediction of the severity of the diabetes in the proposed paper. The proposed work will provide a new platform for analyzing the data set for new patients and submitting an accurate prediction. Pregnancy frequency, diastolic blood pressure, diabetes pedigree function, and class distribution outcome, etc. are the parameters considered for the prediction of the severity range of diabetes. This paper aims to develop a model for an Intelligent Diabetes Prediction system using the ECLAT algorithm and reduce medical misdiagnoses by providing proper interpretation and bringing down treatment costs.

Keywords: ECLAT algorithm, prediction, diabetes.

1 Introduction

"Sugar" is the layman's term for Diabetes Mellitus (DM). This disease arises due to endocrine disorders where insulin deficiency or absence emerges. According to the International Diabetes Federation (IDF), the diabetes prevalence in India will be 69.9 million in 2025 [1].

The pancreas possesses 98% exocrine cells and 2% endocrine-type cells. Pancreatic Polypeptide (PP) cells, Alpha cells, Beta cells, and Delta cells are the four types of endocrine cells that are heavily used in metabolic function. B cells release insulin. The inability of GLUcose Transporters (GLUT) to pass glucose in pancreatic B cells minimizes insulin secretion and is reflected in Type 2 DM [2].

N. V. Suresh and P. S. Buvaneswari (eds.), *Proceedings of the International Conference on Digital Transformation in Business: Navigating the New Frontiers Beyond Boundaries (DTBNNF 2024)*, Advances in Economics, Business and Management Research 283,
https://doi.org/10.2991/978-94-6463-433-4_26

Alpha cells (A cells) in pancreas secret glucagon acts an important role in converting glycogen (carbohydrate in skeletal muscles and the liver) into glucose. Insulin is secreted by beta cells that permit glycogenesis and deliver glucose into various body parts like adipose tissue and reduce the blood glucose level [3].

Diabetes patients are suffering from a poor lifestyle through symptoms and signs persistent among them. Audit of Diabetes-Dependent Quality of Life (ADDQoL) is one of the diabetic-specific questionnaires developed by WHO Quality of Life (WHOQoL) [4].

A large amount of information for hospitals, patients, doctors, and clinical instruments available are maintained. Requisite data is extracted from Electronic Health Records (EHR) using data mining techniques. Preliminary stage prediction of the disease saves lives and reduces medical expenses. The association rule mining technique is used for Knowledge Discovery in Database (KDD) [5].

Pattern recognition techniques are used in data mining to conclude decision-making in the prediction of the severity of the disease. Diabetes affects the heart, eyes, foot, and kidneys, and it causes heart disease, blindness, foot ulcers, and kidney transplantation. By getting the various attribute details, the proposed system will give the diabetics level [6].

There are seven steps in the severity range of the diabetes prediction system. They are Data collection, data preprocessing, unsupervised machine learning model of association rule, training set of data, test data, data analysis, and prediction level output [7]. The association rule mining technique is used for generating the rule set. Boolean vectors are used for identifying each attribute as a Boolean variable. One of the association rules is explained as follows:

$$\text{UG student} \Rightarrow \text{Pursuing PG [support} = 2\%; \text{confidence}=60\%] \quad (1)$$

A support of 2% of the association rule implies that 2% of undergraduate (UG) students will pursue Post Graduation (PG). A confidence of 60% signifies that 60% of UG students will undergo PG studies. In the proposed work, the ECLAT algorithm is used for the prediction of the severity range of diabetes which is a backtracking algorithm. It works in the depth-first search traversal method [8].

The rest of the paper is structured as follows: Section II gives the background details about diabetes and association rule mining. Section III includes the existing literature works related to diabetes mellitus, data mining, machine learning, and ECLAT algorithm. Section IV describes the prediction of diabetes using ECLAT algorithm implementation details. Section V presents the results and discussion of the proposed prediction model. Finally, section VI concludes the paper.

2 Background

In Charaka Samhita, an ancient Sanskrit Ayurveda textbook, diabetes was mentioned as "madhumeha". So the human race knows diabetes in the olden days also. Insulin plays a vital role as the doorkeeper that permits glucose to enter the cell. Cells do not receive glucose in the required amount, because of a deficiency of insulin secretion. To compensate, the body generates gluconeogenesis (more glucose).

2.1 Diabetes Mellitus

Type 1 DM (juvenile-onset DM, Insulin Dependent DM (IDDM), and ketosis-prone DM), Type 2 DM (adult-onset DM, Non-Insulin Dependent DM (NIDDM)), and gestational DM are the three commonly available diabetes. The brain, liver, small intestine, kidneys, pancreas (α-cell and β-cell), skeletal muscle, and adipose tissue are the body parts involved in diabetes [9].

Irresistible blood glucose levels in DM victims may lead to death. So it is necessary to monitor the blood glucose level and provide insulin in the required amount advised by endocrinologists [10].

Etiology.

1. Type 1 DM possesses an epidemic nature.
2. Several viruses are involved as a part of the analysis of DM such as coxsackie and rubella.
3. Genes regulating the immune system, eg. Wolfram syndrome.
4. Pneumonia, mucormycosis and fungal cystitis are the usual infections [11].

Cost of Diabetes.

1. Direct cost – both direct hospital expenses (diagnosis, medication, etc.) and direct non-hospital expenses (social service, transport, etc.).
2. Indirect cost – the cost of defection, loss of capability, and dysfunction.
3. Tangible cost – the cost of social reliance and segregation, low financial situation, loss of quality of life and emotional health [12].

Symptoms.

1. Sudden weight loss
2. Polyuria (Frequent urination)
3. Increased Thirst
4. Nocturia
5. Vulnerability to Infection
6. Vomiting
7. Blurred vision [13].

2.2 Association Rule Mining

This technique splits the larger problem into smaller ones by applying the performance metrics of support and confidence. Support indicates the frequency of data in a data set.

Apriori, ECLAT, and FP-growth are the various association rule mining algorithms used for the prediction of the disease [14].

Relationships or associations between various attributes are generated first in the association rule mining. Depending on the associations, data mining techniques are applied. Association rules have two parts. They are antecedent and consequent. In equation 1, UG student is the antecedent and Pursuing PG is the consequent [15].

A highly important task in knowledge base discovery is identifying the frequent item set. Execution time and memory consumption are the factors used to determine the association rule mining algorithm for the proposed work. ECLAT algorithm works on prefix trees. AprioriTrie, Dynamic Item-set Counting (DIC), and partition are the various association rule mining algorithms [16].

3 Literature Works

The proposed work "Predicting the Severity of Diabetes using ECLAT Algorithm" comprises knowledge regarding Diabetes Mellitus, Data mining, Machine learning algorithm, and ECLAT algorithm of association rule mining. Some of the existing literature works are explained here.

3.1 Diabetes Mellitus

Hallmark explicit features of Diabetes are typical glucose-receiving levels emerging in hyperglycemia, and variations in protein and lipid metabolism [17]. Metformin is used for the treatment of gluconeogenesis. Sulfonylureas oral prescription is used to secret insulin. Severe renal disease, heart-related diseases, critical vision loss, and high pressure are the adverse symptoms of DM [18].

It is a progressive sickness which is otherwise called "slow poison" that worsens the patient's life by affecting several organs. National food policies to provide healthy, nutritious food, health policies to reduce the misbehavior of citizens, prevention policies to improve disease awareness, and policies to reduce medical expenses should be implemented by various ministries to provide a healthy society by Indian government [19].

3.2 Data Mining

Data mining methods are nowadays used in every aspect of the medical field to make a final decision. In a country like India with high population density and restricted resource facilities, mining approaches can be utilized to manage healthcare problems. In remote areas of the country, online healthcare counseling can be utilized to replace the presence of physicians [20].

Mining concerns processing known data to conclude a judgment. Social Security firms provide the details for the affected persons. Number of clinical tests conducted

for patients with no disease is higher than the diseased patient [21]. Prediction and control of diabetes is important to improve the patient's lifestyle. Pima Indian Diabetes data set is used for data analysis which contains 768 women data with 9 attributes. Different classification algorithms give different accuracy levels [22].

3.3 Machine Learning Algorithms

Diabetes is predominantly irrecoverable and the only approach to avoid is premature diagnosis and detection. Machine learning (ML) algorithms and Medicare data analysis are the need of the hour. They can provide suggestions for doctors. Digital diabetes clinics can be used to help diabetes patients [23].

The predictive analysis technique encompasses a collection of data mining algorithms, machine learning approaches, and statistical procedures that utilize past and current data to extract useful patterns and predict future data. This technique can be applied to healthcare data to arrive at a decision-making in a critical situation [24].

ML techniques improve patient care. Enhance the allocation of resources to patients, maximize emergency readiness, and encourage novel techniques such as remote monitoring and telemedicine [25].

3.4 ECLAT Algorithm

In this proposed Diabetes Severity Prediction System, the ECLAT algorithm is used for prediction which comes under the unsupervised ML technique. Tree-like structure Tidset is used. ECLAT algorithm reduces memory usage during processing. It produces a large number of rule sets [8].

There are n number of transactions and n number of item sets. In a transactional database, the ECLAT algorithm identifies frequent item sets. ECLAT data mining identifies minimum support transactions from a set of transactions [26].

The depth-first traversal is used by the ECLAT algorithm whereas the Apriori algorithm uses breadth-first traversal. Memory usage is lesser than the Apriori algorithm. Pattern generation time is lesser than the Apriori algorithm for tiny data sets [27].

4 Predicting the Severity of Diabetes using ECLAT Algorithm Implementation

The proposed work was implemented in a dual-core processor, 4 GB RAM, and 512 GB Hard Disk Drive. The front-end tool is Microsoft Visual C#. The back-end tool is SQL Server 2005 and the platform used is ASP .Net Framework 3.5. The Diabetes data set is fetched from the Repository of UCI Machine Learning. It includes 8 attributes and 1 binary class variable outcome data for 768 women. The partial data set is shown in Figure 1.

Eight features are diastolic blood pressure, body mass index, plasma glucose concentration, two-hour serum insulin, age in years, number of times pregnant, Triceps skinfold thickness, and Diabetes Pedigree function.

	A	B	C	D	E	F	G	H	I	J
1	preg	plas	pres	skin	insu	mass	pedi	age	class	
2	6	148	72	35	0	33.6	0.627	50	tested_positive	
3	1	85	66	29	0	26.6	0.351	31	tested_negative	
4	8	183	64	0	0	23.3	0.672	32	tested_positive	
5	1	89	66	23	94	28.1	0.167	21	tested_negative	
6	0	137	40	35	168	43.1	2.288	33	tested_positive	
7	5	116	74	0	0	25.6	0.201	30	tested_negative	
8	3	78	50	32	88	31	0.248	26	tested_positive	
9	10	115	0	0	0	35.3	0.134	29	tested_negative	
10	2	197	70	45	543	30.5	0.158	53	tested_positive	
11	8	125	96	0	0	0	0.232	54	tested_positive	
12	4	110	92	0	0	37.6	0.191	30	tested_negative	
13	10	168	74	0	0	38	0.537	34	tested_positive	
14	10	139	80	0	0	27.1	1.441	57	tested_negative	
15	1	189	60	23	846	30.1	0.398	59	tested_positive	
16	5	166	72	19	175	25.8	0.587	51	tested_positive	

Fig. 1. UCI Diabetes Data set

Data cleaning, preprocessing, and filtering of data are executed. Data is encoded for the training data set and sample data set. Training data is executed ECLAT algorithm to generate the association rules. Pruning is accomplished to compress the size of the decision tree. An association rule database is created. Test data entered by the doctor are considered for rule matching in the database. Depending upon the matching level for test data, the severity level of diabetes was predicted and the results are displayed. The entire Diabetes prediction model is shown in figure 2.

Four modules are used in the Diabetes prediction decision support system. They are Admin, Input Dataset, Users, and Disease Prediction. Admin module can add the doctor's details and training data sets. UCI Diabetes dataset is given as the input. In the user module, the clinical test results are entered to identify the diabetes severity range. In the disease prediction module, the results are displayed to the intended users who have entered their details for diabetes checking. The entire steps are summarized in figure 3.

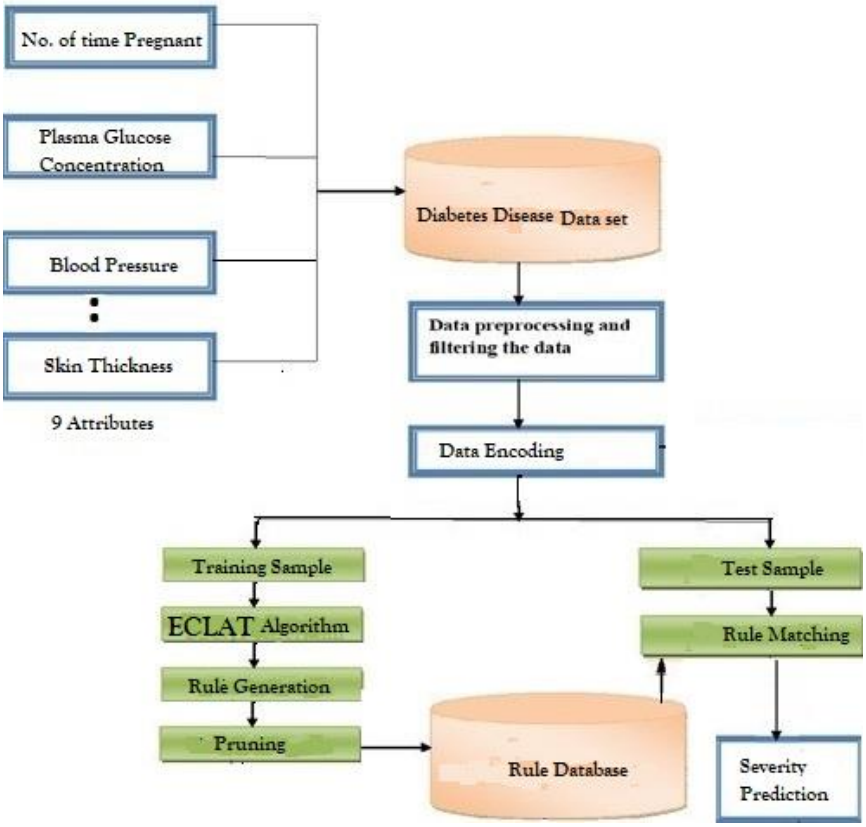


Fig. 2. Proposed Diabetes Prediction Model

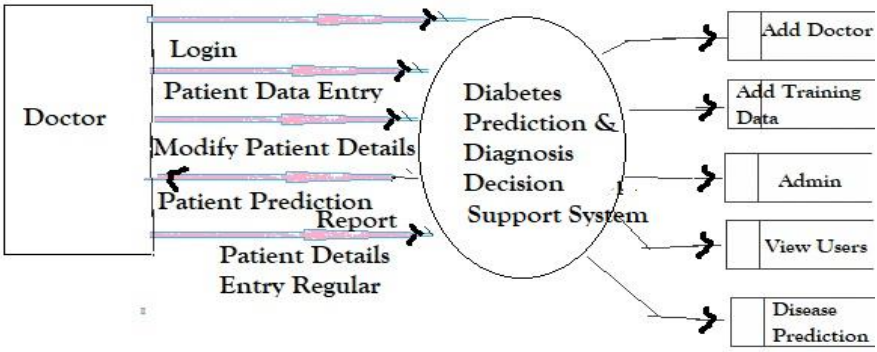


Fig. 3. Decision Support System for Diabetes Prediction

5 Results and Discussion

Details of the patients/ users are registered in the Decision Support system. ID number automatically generated, user name, password, email ID, Gender, date of birth, and address are needed for the registration process. It is shown in Figure 4.

Initially system administrator will add the doctor's details into the system. It is shown in Figure 5.

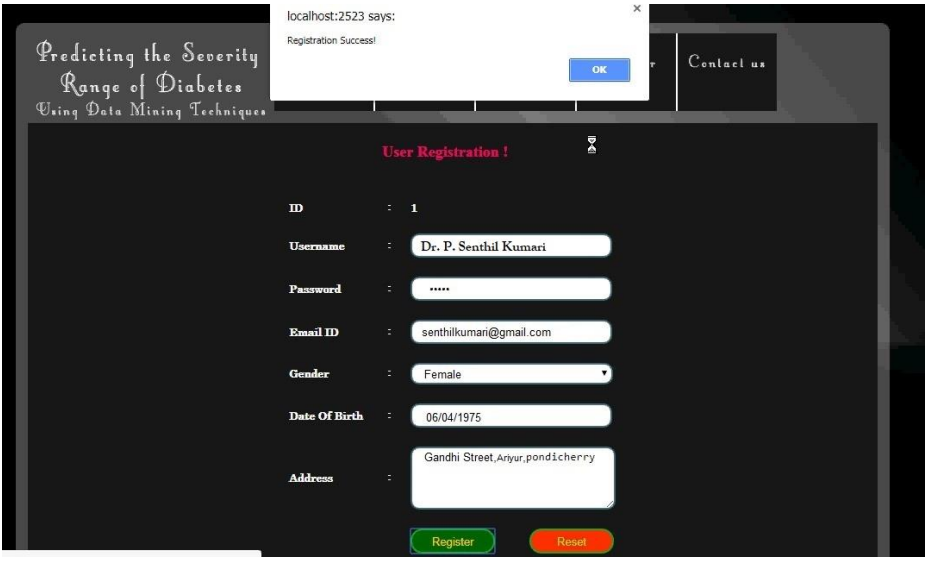


Fig. 4. User Registration

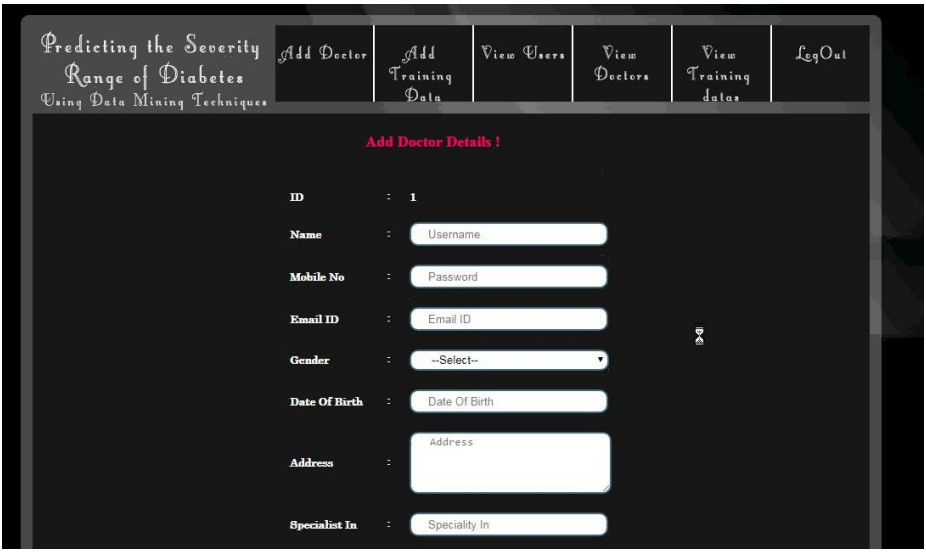


Fig. 5. Adding Doctor Details by Admin

The intended user for the diabetes prediction enters test data. It is shown in Figure 6. The sample training data set is shown in Figure 7.

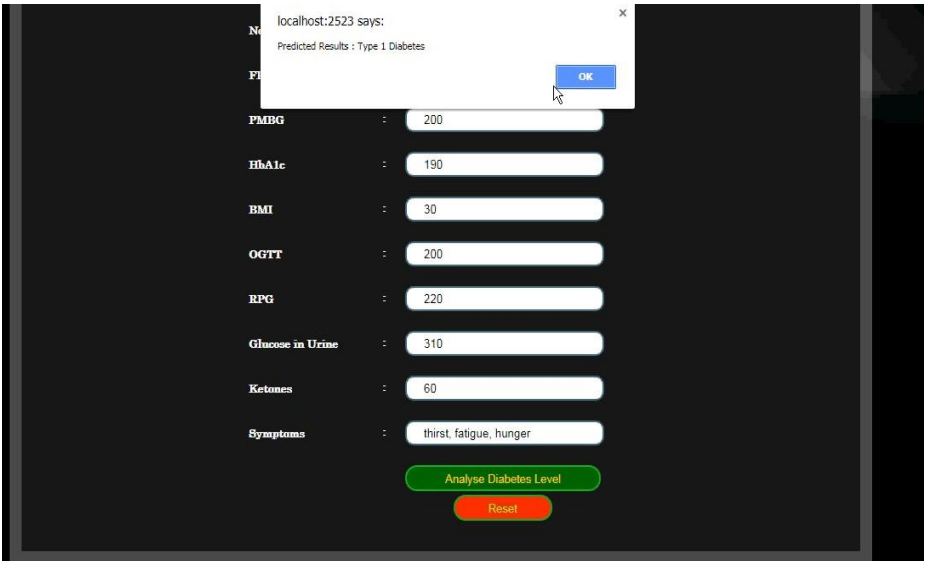
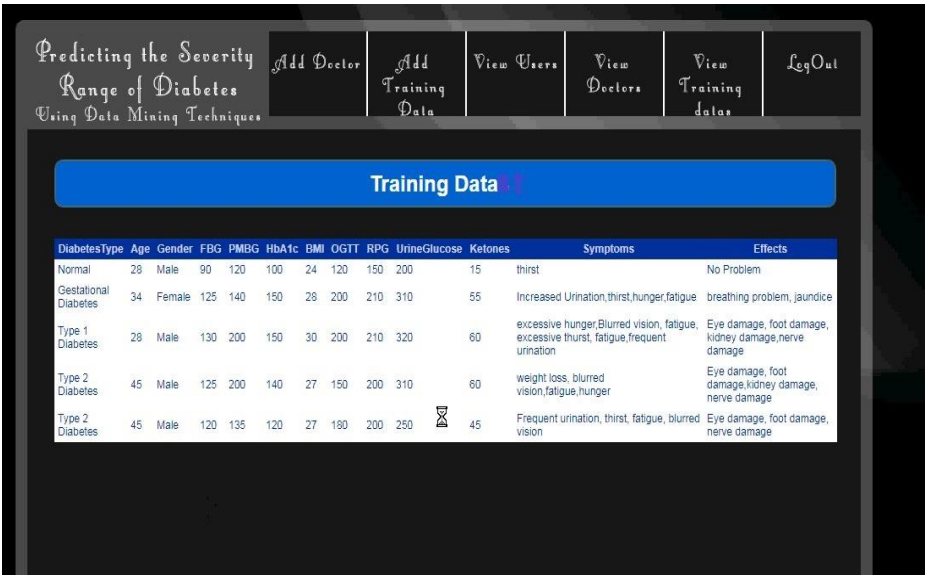


Fig. 6. Test Data



DiabetesType	Age	Gender	FBG	PMBG	HbA1c	BMI	OGTT	RPG	UrineGlucose	Ketones	Symptoms	Effects
Normal	28	Male	90	120	100	24	120	150	200	15	thirst	No Problem
Gestational Diabetes	34	Female	125	140	150	28	200	210	310	55	Increased Urination,thirst,hunger,fatigue	breathing problem, jaundice
Type 1 Diabetes	28	Male	130	200	150	30	200	210	320	60	excessive hunger,Blurred vision, fatigue, excessive thurst, fatigue,frequent urination	Eye damage, foot damage, kidney damage,nerve damage
Type 2 Diabetes	45	Male	125	200	140	27	150	200	310	60	weight loss, blurred vision,fatigue,hunger	Eye damage, foot damage,kidney damage, nerve damage
Type 2 Diabetes	45	Male	120	135	120	27	180	200	250	45	Frequent urination, thirst, fatigue, blurred vision	Eye damage, foot damage, nerve damage

Fig. 7. Training Data for the Diabetes Prediction Model

Test data are analyzed for the diabetes prediction and the results are shown in figure 8. Gender, number of times pregnant, Fasting Blood Glucose (FBG), Post Meal Blood

Glucose (PMBG), Glycated hemoglobin (HbA1c), body mass index, Oral Glucose Tolerance Test (OGTT), Random Plasma Glucose (RPG), Glucose in Urine, and ketones are taken for consideration of the diabetes severity level prediction.

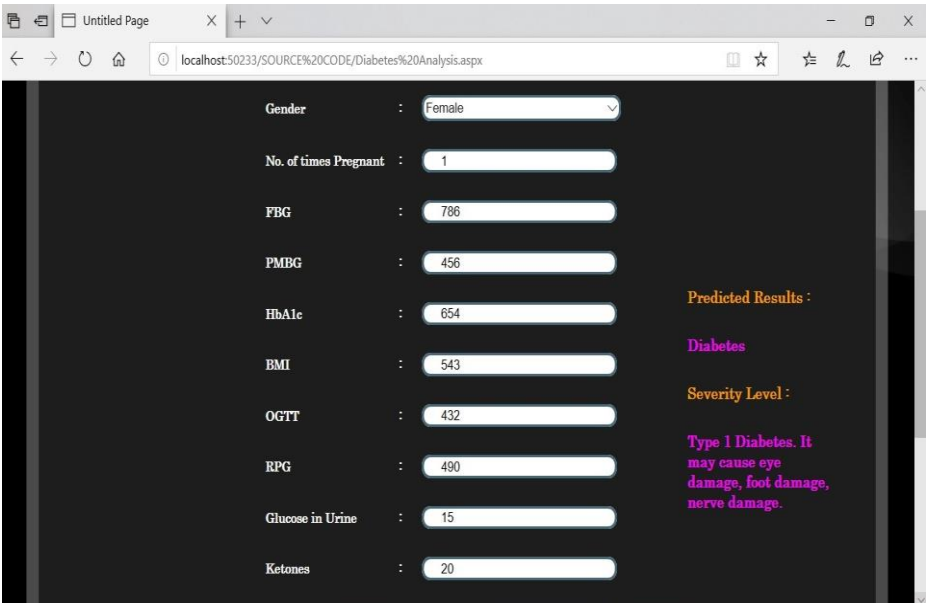


Fig. 8. Diabetes Prediction Results

6 Conclusion

Many digital apps are introduced on Android phones for diabetes patients to remind them about clinical tests, diet, and regular activity. So, their quality of life is improved [23]. Wearable devices are introduced in the market to monitor blood glucose levels and send a warning message to patients if the value exceeds a certain threshold level. Warning messages to remind them to intake tablets, inject insulin, and request assistance in case of emergency [28].

A thorough investigation is needed for the analysis of acute diseases. Computer-assisted decision support system for diabetes prediction helps doctors with the prediction of severity and further treatment proceedings. Diabetes can be prevented if it is diagnosed in the early stage. Many awareness programs should be inculcated for human welfare to impart knowledge about diabetes. In the future, the research can be extended with different prediction algorithms.

References

1. Bacchetta R, Passerini L, Gambineri E, Dai M, Allan SE. Defective regulatory and effector T cell functions in patients with FOXP3 mutations, J Clin Invest. 2006; 116:1713-1722.
2. University of Pretoria, Diabetes: Literature Review, Chapter 1, 14-84.

3. Sandhya Ranee Bhoi, Kaustubh Singh, Manmeet Kaur, and Rojina Swayamsiddha Sahu. A Short Review on Diabetes Mellitus. International Journal for Modern Trends in Science and Technology 2021, 7 pp. 284-288. https://doi.org/10.46501/IJMTST0712054
4. Ramasamy Aarthy, Antonina Mikocka-Walus, Rajendra Pradeepa, Ranjit Mohan Anjana, Viswanathan Mohan, Kathryn Aston-Mourney. Quality of Life and Diabetes in India: A Scoping Review, Indian Journal of Endocrinology and Metabolism, 2021, 25, 5, 365-380: 10.4103/ijem.ijem_336_21.
5. M. Sornalakshmi, S. Balamurali, M. Venkatesulu, M. Navaneetha Krishnan, Lakshmana Kumar Ramasamy, Seifedine Kadry, Sangsoon Lim. An efficient apriori algorithm for frequent pattern mining using map reduce in healthcare data, Bulletin of Electrical Engineering and Informatics Vol. 10, No. 1, February 2021, pp. 390~403.
6. Sandeep Gupta, Ashish Gupta, Nilima Bhoir. Severity Estimation & Food Recommendation for Diabetic Patient using Data Mining, IJSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 12, 2019, pp. 220-222.
7. Rohini Patil, Leena Majumder, Manisha Jain, Vedanti Patil. Diabetes Disease Prediction Using Machine Learning, International Journal of Research in Engineering, Science and Management Volume-3, Issue-6, June-2020, pp. 292-295.
8. Ms. J. Omana, Ms. S. Monika, Ms. B. Deepika. Survey on Efficiency of Association Rule Mining Techniques, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.4, April- 2017, pg. 5-8.
9. Unai Galicia-Garcia, Asier Benito-Vicente, Shifa Jebari, Asier Larrea-Sebal, Haziq Siddiqi, Kepa B. Uribe, Helena Ostolaza and César Martín. Pathophysiology of Type 2 Diabetes Mellitus, Int. J. Mol. Sci. 2020, 21, 6275; doi:10.3390/ijms21176275.
10. Prajoona Valsalan, Najam Ul Hasan, Umer Farooq, Manaf Zghaibeh, and Imran Baig. IoT Based Expert System for Diabetes Diagnosis and Insulin Dosage Calculation, Healthcare 2023, 11, 12. https://doi.org/10.3390/healthcare11010012
11. Nishita Singh, Roohi Kesherwani, Arun Kumar Tiwari and Dilip Kumar Patel. A review on diabetes mellitus, The Pharma Innovation Journal 2016; 5(7): 36-40.
12. Sumit Oberoi & Pooja Kansra. Economic menace of diabetes in India: a systematic review, International Journal of Diabetes in Developing Countries, 17 June 2020, 12. https://doi.org/10.1007/s13410-020-00838-z
13. Dereje Abebe, Yayehirad Tassachew, Jemal Adem, Nejmudin Reshad, and Sintayehu Delelegn. Diabetes Mellitus For the Ethiopian Health Center Team, Debub University, 2005, pp. 1-132.
14. R. Karthiyayini. Enhanced Association Rule Mining Algorithm for Medical Analysis (EARMAM), International Journal of Pure and Applied Mathematics, Volume 119 No. 12 2018, 16241-16248.
15. M. Kivrak, F.B. Akcesme, and C. Colak. Evaluation of Association Rules Based on Certainty Factor: An Application on Diabetes Data Set, The Journal of Cognitive Systems, Vol. 5, No. 1, 2020, pp. 19-22.
16. E. Ramaraj, N. Venkatesan. An Efficient Pattern Mining Analysis in Health Care Database, Journal of Theoretical and Applied Information Technology, pp. 549-555.
17. Kroon LA, Assemi M, Carlisle BA. Diabetes mellitus, in: applied therapeutics: the clinical use of drugs. KodaKimble, Mary Anne, Young Lloyd Yee, et al. (Eds). Lippincott Williams & Wilkins (2009).
18. Gudisa Bereda. Brief overview of diabetes mellitus, Diabetes Management (2021) S1: 21-27.
19. Rajendra Pradeepa, Viswanathan Mohan. Epidemiology of type 2 diabetes in India, Indian Journal of Ophthalmology, November 2021; 69: pp. 2932-2938.

20. Aditya Kumar Tripathi and Anu Sharma. Techniques for Data Mining Prediction in the Health Care Sector, International Journal of Innovative Research in Computer Science and Technology (IJIRCST) ISSN (online): 2347-5552, Volume-11, Issue-3, May 2023 https://doi.org/10.55524/ijircst.2023.11.3.6 Article id- IRP1380, Pages 32-37.
21. Rashi Rastogi, Mamta Bansal. Diabetes prediction model using data mining techniques, Measurement Sensors, Vol. 25, Feb 2023, 100605, https://doi.org/ 10.1016/j.measen. 2022. 100605.
22. Dr. Manpreet Singh, Dr. Pankaj Bhambri, Dr. Inderjit Singh, Dr. Amit Jain, Er. Kirandeep Kaur. Data Mining Classifier for Predicting Diabetics, Annals of R.S.C.B., Vol. 25, Issue 4, 2021, Pages. 6702 – 6712.
23. Cahn, A., Akirov, A., & Raz, I. (2018). Digital health technology and diabetes management. Journal of Diabetes, 10(1), 10–17. https://doi.org/10.1111/1753-0407.12606
24. Gauri D. Kalyankar, Shivananda R. Poojara and Nagaraj V. Dharwadkar. Predictive Analysis of Diabetic Patient Data Using Machine Learning and Hadoop, International Conference On I-SMAC, 978-1-5090-3243-3, 2017.
25. Sand Mohammad Salhout. Machine learning in healthcare strategic management: a systematic literature review, Arab Gulf Journal of Scientific Research, November 2023, DOI 10.1108/AGJSR-06-2023-0252.
26. Suganya, V., & Suresh, N. V. (2024). Potential Mental and Physical Health Impacts of Spending Extended Periods in the Metaverse: An Analysis. In Creator's Economy in Metaverse Platforms: Empowering Stakeholders Through Omnichannel Approach (pp. 225-232). IGI Global.
27. Suresh, N. V., & Rexy, V. A. M. (2024, February). An Empirical Study on Empowering Women through Self Help Groups. In 3rd International Conference on Reinventing Business Practices, Start-ups and Sustainability (ICRBSS 2023) (pp. 957-964). Atlantis Press.
28. Jason Reynaldo, and David Boy Tonar. Data Mining Application using Association Rule Mining ECLAT Algorithm Based on SPMF, MATEC Web of Conferences 164, 01019 (2018) https://doi.org/10.1051/matecconf/201816401019.
29. Vrushali R. Balpande, and Rakhi D. Wajgi. Prediction and Severity Estimation of Diabetes Using Data Mining Technique, International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2017), 978-1-5090-5960-7/17.
30. Fleming, G. A., Petrie, J. R., Bergenstal, R. M., Holl, R. W., Peters, A. L., & Heinemann, L. (2020). Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetologia, 63(2), 229–241. https://doi.org/10.1007/s00125-019-05034-1