



Prediction of Type - I and Type –II Diabetes: A Hybrid Approach using Fuzzy Logic and Machine Learning Algorithms

Geeta Pattun * 

*Corresponding Author, Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University, Hyderabad, Telangana, India-500032. E-mail: geeta781@manuu.edu.in

Khaleda Afroaz 

Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University, Hyderabad, Telangana, India-500032. E-mail: khaleda.afroaz@manuu.edu.in

Ahmad Talha Siddiqui 

Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University, Hyderabad, Telangana, India-500032. E-mail: ahmadtalha2007@gmail.com

Shaheena Ghazala 

Department of Computer Science and Information Technology, School of Technology, Maulana Azad National Urdu University, Hyderabad, Telangana, India-500032. E-mail: shahinaghazala@gmail.com

Abstract

Diseases like diabetes are chronic and require long-term management. Inadequate production of insulin results in high blood sugar levels. Such diseases lead to serious health issues such as heart ailments, blood vessel complaints, eye ailments, kidney function disorders, and nerve ailments. Hence, accurate assessment and management of risk factors are crucial for the onset of diabetes. Our proposed approach combines fuzzy logic & machine learning algorithms for diabetes risk prediction. Three machine learning models were trained to classify patients into two categories of diabetes (Type-I and Type-II) based on their clinical dataset collected from Katihar Medical College & Hospital and Suvadhan Lab. The polynomial regression algorithm achieved a score of 0.947, while the support vector regression algorithm with the rbf kernel achieved a score of 0.954, with a linear kernel achieved a score of 0.73. Our proposed

approach performed well with respect to the conventional approaches with improved accuracy by identifying the patients at diabetes risk. In future work, we further analyze the relationship between other ignored factors which contribute to diabetes risk.

Keywords: Diabetes, Blood Sugar, Machine Learning Algorithm, Fuzzy Logic, Disease Management, Risk Factors, Insulin Resistance, Polynomial Regression, Support Vector Regression.

Journal of Information Technology Management, 2023, Vol. 15, Special Issue, pp. 35- 56

Published by University of Tehran, Faculty of Management

doi: <https://doi.org/10.22059/jitm.2023.95244>

Article Type: Research Paper

© Authors

Received: July 06, 2023

Received in revised form: August 24, 2023

Accepted: November 09, 2023

Published online: December 24, 2023



Introduction

Nowadays, diabetes is a chronic disease that hinders the body's ability to regulate blood sugar level (Geman et al., 2017) (Bressan et al., 2020). Insulin, a hormone produced by the pancreas, plays a crucial role in breaking down the sugars we consume through our daily diet, converting them into glucose to fuel our body for daily activities (Howsalya Devi et al., 2020). However, when the body fails to produce adequate insulin, it can lead to severe complications such as heart disease, blood vessel disorders, eye damage, kidney dysfunction, and nerve disorders (Thakkar et al., 2021).

The two categories of diabetes are type –I and type –II. Type - I diabetes is a perennial autoimmune disease found in infants and adults (Devi & Uma, 2019; Howsalya Devi et al., 2020). This situation occurs when insulin production by pancreatic beta cells is impaired, leading to insulin deficiency and spikes in blood sugar levels. Symptoms to watch out for are excessive hunger, frequent thirst, frequent urination, weight loss, fatigue, allergic reactions, and blurred vision. Premature treatment can lead to serious complications and even death. Insulin therapy is a common treatment given for regulating blood sugar levels of type 1 diabetes.

Type 2 diabetes is a long-term disease that primarily affects the population between the ages of 30 and under 100 (Yahyaoui, Jamil & Rasheed, 2019; Howsalya Devi et al., 2020; Rajeswari et al., 2018; Vijiyakumar et al., 2019). It occurs when the body does not produce enough insulin, causing blood sugar levels to rise. Common symptoms to look out for are excessive hunger, extreme thirst, weakness, distorted vision, itching, excessive urination, slow recovery, and the development of hemorrhoids. These symptoms in any circumstances left unchecked can lead to serious complications to name few heart ailments, kidney functionality

disorders, myopia, and nerve ailments. Several lifestyle factors also contribute to type 2 diabetes.

Fuzzy logic is a mathematical framework that provides a solution for processing data that is inaccurate, uncertain, or ambiguous. It is particularly beneficial in dealing with real-world issues where knowledge is often incomplete or imprecise. By utilizing Fuzzy logic, it becomes easier to develop complex relationships between variables and rules, simplifying the decision-making process and making it easier to comprehend (Bressan et al., 2020; Niswati et al., 2018; Rajeswari et al., 2018; Swain et al., 2013; Thakkar et al., 2021). Fuzzy logic has widespread applications, ranging from control theory to artificial intelligence. Its usefulness lies in its ability to incorporate fuzzy risk factors, such as "slightly obese" or "moderately high" blood sugar levels, into predictive models that cannot be represented by traditional binomial taxonomies (Bressan et al., 2020). This integration of ambiguous risk factors into predictive models can significantly enhance model accuracy and robustness, which is particularly important when predicting the risk of diabetes.

A potent tool for predicting diabetes has emerged: machine learning. These algorithms are able to recognize patterns and relationships that might escape the notice of human experts by training models on vast datasets of patient data. In order to produce precise predictions of a patient's risk for developing diabetes, machine learning models can consider a wide range of variables, such as age, weight, family history, and blood glucose levels. The outcomes of patients may be improved, and healthcare costs may be decreased, if healthcare professionals can intervene earlier with preventative measures. Machine learning algorithms can also continuously increase their predictive accuracy (Lukmanto et al., 2019; Sarwar et al., 2018; Thakkar et al., 2021) as more data is gathered over time, which makes them a valuable tool in the battle against this crippling illness.

Our research introduces a novel approach to diagnosing diabetes. We used fuzzy logic systems to determine boundary values for predicting abnormalities, enabling us to identify individuals who need additional care. By in-depth studying of risk factors associated with type 1 and type 2 diabetes across different age groups, we can contribute valuable insights that inform public health initiatives. Our innovative hybrid approach, combining fuzzy logic with machine learning, enhances the precision of diabetes risk prediction models. This cutting-edge method (Khan et al., 2022; Kumar et al., 2021) effectively handles uncertain data and improves early identification and intervention strategies. Our research has significant implications for healthcare practices and decision support systems, ultimately benefiting individuals living with diabetes.

Literature Review

We have tabulated the previous work done in the context of the prediction of diabetes types. Through Table 1 we attempted to project fuzzy logic adopted for diabetes prediction, in

particular, it describes the methods implemented by researcher enthusiasts & the datasets utilized for training and evaluating the system. In Figure 1 we projected the accuracy of the models developed through a bar plot, in Figure 2 we projected the dataset adopted in the previous study through the count plot, and through the graph we could make out that much of the models developed were based on the Pima Indian diabetes dataset.

Table 1. Methods implemented in previous work for diabetes prediction

S.No	Author	Method	Dataset	Accuracy
1	(Vijiyakumar et al., 2019)	Random Forest classifier	pima Indian diabetes dataset (UCI)	90%
2	(A. Yahyaoui et al., 2019)	CNN, SVM, Random Forest	pima Indian diabetes dataset (UCI)	83.67%
3	(Sarwar et al., 2018)	SVM, KNN, LR, Decision Tree, NB	Pima Indian diabetes dataset (UCI)	77%
4	(Bressan et al., 2020)	Neuro Fuzzy, Decision Tree	Moreira Glycemic index data	70%
5	(Geman et al., 2017)	FNFS Neuro Fuzzy System	Pima Indian diabetes dataset (UCI)	84%
6	(Howsalya Devi et al., 2020)	farthest First & Smo classifier	Pima Indian diabetes dataset (UCI)	99.40%
7	(Khalil & Al-Jumaily, 2017)	SVM, PNN, K-means, Fe mean, Fuzzy e-Mean, probabilistic NN	Black Lion general specialized hospital Ethiopia	96%
8	(Lukmanto et al., 2019)	Fuzzy Support Vector Machine	Pima Indian diabetes dataset (UCI)	89%
9	(Niswati et al., 2018)	Fuzzy System Mamdani Method	6 puskesmas east jakaarte sample data	96%
10	(Raj et al., 2019)	Naïve Baiyes, SVM	sample from multiple hospitals	82%
11	(Rajeswari et al., 2018)	Fuzzy Association classifier	pima Indian diabetes dataset (UCI)	84%
12	(Swain et al., 2013)	NN, Hybrid Adaptive Neuro Fuzzy System	Local inhabitants of Bhubaneshwar, India	66%
13	(Thakkar et al., 2021)	Fuzzy System	Pima Indian diabetes dataset (UCI)	96%
14	(Undre et al., 2015)	Fuzzy Neural Network, cbr approach	Pima Indian diabetes dataset (UCI)	90%
15	(Verma & Mishra (2017)	Naïve Baiyes, SVM, RED Tree, J48	Pima Indian diabetes dataset (UCI)	76.80%

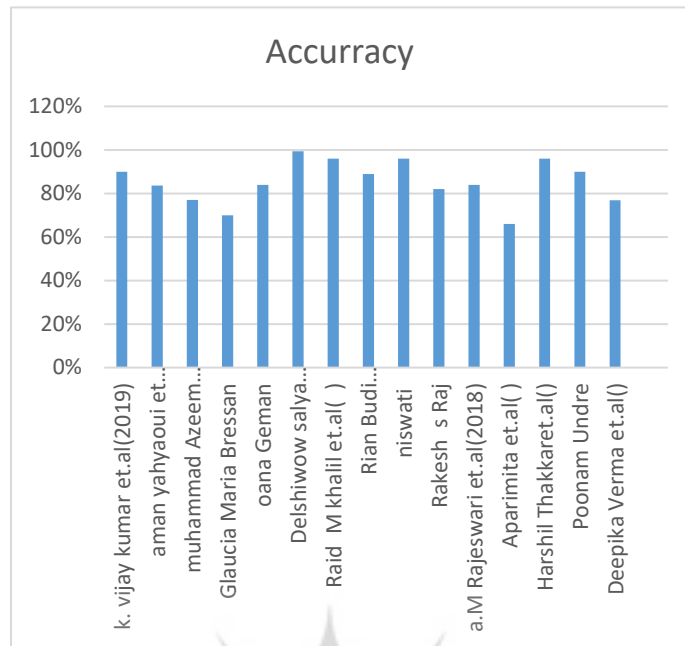


Figure 1. Accuracy obtained in diabetes analysis

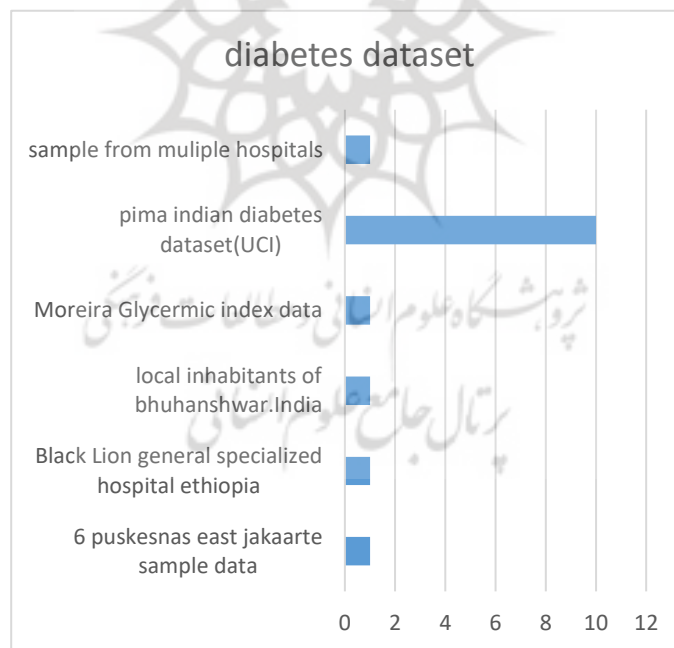


Figure 2. Diabetes dataset used lated work

Methodology

We developed a novel approach for diagnosing diabetes by combining fuzzy logic and machine learning. Our hybrid approach employs a fuzzy system set by experts based on clinical data (Undre et al., 2015). Using this method, we can accurately classify individuals with type I or type II diabetes by extracting pertinent features from diabetic datasets. One of the key strengths of our study lies in the utilization of rule-based feature extraction techniques to enhance the diagnostic precision of the model. By employing these methods, we could effectively extract relevant features from the data, leading to improved accuracy and reliability in the diagnostic process. This model contributes to its overall strength and effectiveness in diagnosing the target condition of the individuals with such a chronic condition (Lukmanto et al., 2019).

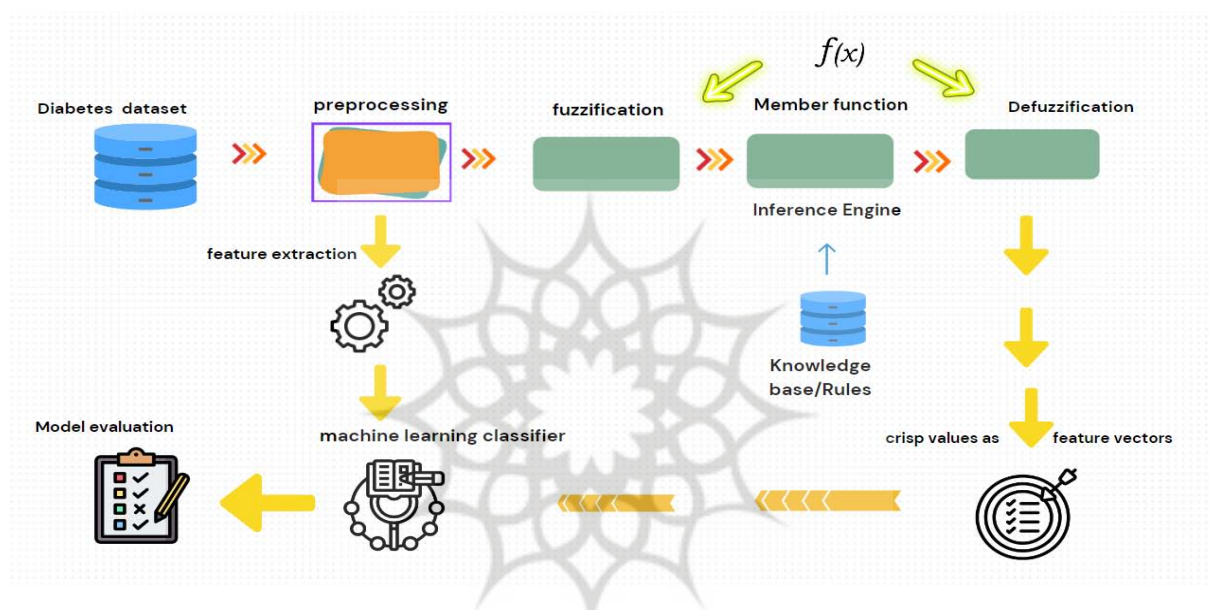


Figure 3. Hybrid model based on machine learning & Fuzzy based system

A mathematical framework called fuzzy logic can be used to reason and make conclusions based on input that is ambiguous or uncertain. It can be used in decision-making systems, engineering, medicine, finance, and economics, among other areas (Bressan et al., 2020; Rajeswari et al., 2018). To convert precise values into degrees of membership using fuzzy logic, membership functions, and linguistic concepts are used (Thakkar et al., 2021). The IF-THEN statements are used to create fuzzy rules and logical operators like AND, OR, and NOT are used in the inference process to create a single fuzzy output result. In our research, we emphasize the critical step of defuzzification, where the fuzzy output value is transformed into a crisp value for decision-making. With the defuzzification process we can ensure clarity and facilitates informed decision-making based on the output of the fuzzy logic system (Niswati et al., 2018; Swain et al., 2013).

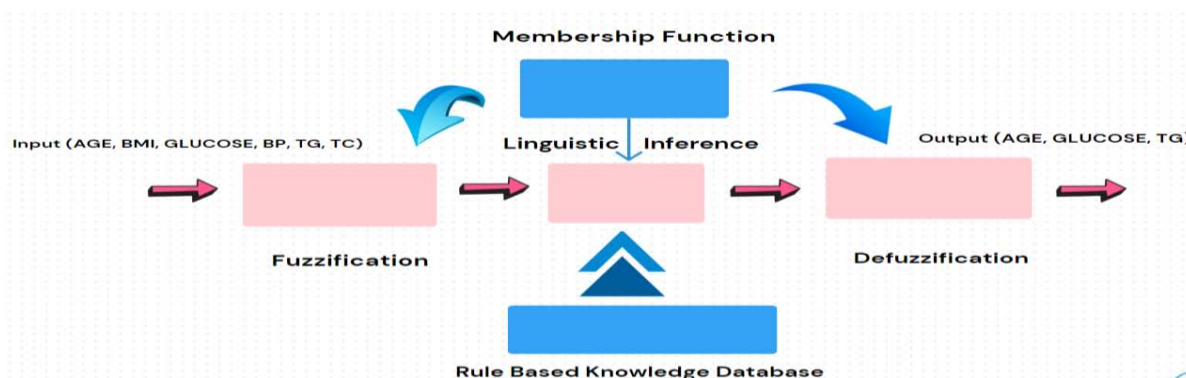


Figure 4. Fuzzy Expert system

Machine learning algorithms

Several algorithms have been applied to train the dataset for diabetes patient prediction in a hybrid environment that integrates fuzzy logic-based systems with machine-learning approaches (Howsalya Devi et al., 2020; Swain et al., 2013)

Linear Regression is a simple algorithm that models the linear relationship between input and output variables, which can be used to predict blood glucose levels based on factors such as age, BMI, and blood pressure. Each input value or column receives one scale factor from the linear equation, known as a coefficient, and is denoted by the capital Greek letter Beta (B). This second coefficient, often known as the intercept or bias coefficient, offers the line an extra degree of freedom, allowing it to move up and down on a two-dimensional plot. As an illustration, the model form for a simple regression problem (one x and one y) might be as follows:

$$\text{Where, } y = B_0 + B_1 * X$$

B_0 stands for the Intercept. The independent variable is denoted by the symbol B_1 , which represents the coefficient X . y stands for the dependent variable or outcome.

Support Vector Regression (SVR) is one such technique that converts input data into a high-dimensional feature space and locates a hyperplane that maximizes the separation between the data points. For predicting continuous variables like blood glucose levels, this is helpful (Yahyaoui, Jamil & Rasheed, 2019; Khalil & Al-Jumaily, 2017) The SVR seeks to fit the best line within a threshold value (Distance across the hyperplane and boundary line) a , in contrast to other regression models that aim to minimize the deviation between the real and predicted value. As a result, we can say that the SVR model tries to meet

$$\text{Condition } -a < y - wx + b < a$$

To forecast the value, it uses the locations along this boundary. Because of its adaptability, SVR discovers a suitable line (or hyperplane in greater dimensions) to fit the

data and lets us select how much variation in the model is acceptable. (Howsalya Devi et al., 2020; Lukmanto et al., 2019)

Polynomial Regression, this is a variant of linear regression that uses a polynomial function to describe non-linear correlations between input and output variables. This variation can help anticipate intricate relationships between patient parameters and blood glucose levels. However, it is vulnerable to overfitting and might not work well when used with fresh data.

Data Collection

This research is an experiment on primary diabetes disease on the original dataset. To conduct the study we selected participants from clinical records of individuals diagnosed with type 1 or type 2 diabetes, who fall into different age groups.

We used data from 836 diabetic patients of different ages who were treated at Katihar Medical College & Hospital and Suvadhan Laboratory. We only included patients with type 1 or type 2 diabetes. We collected the data using a convenient method.

The data analysis graph describes a dataset of diabetic patients who received treatment at Katihar Medical College & Hospital and Suvadhan Laboratory. The dataset includes 836 rows and 16 columns, representing various attributes of the patients such as patient id, patient name, gender, age, blood pressure, fasting plasma, two hours plasma, clinical history, Hemoglobin, BMI, family history, blood sugar, total cholesterol, Triglycerides, insulin and target column diabetic type. The patients in the dataset have different ages, and they were diagnosed with either type 1 or type 2 diabetes.

Table 2. Describes the variables of the diabetes dataset

S.no	Attribute	Units/measurements
1	Patient Id	Required
2	Patient Name	Required
3	Gender(F/M)	Category (male, female)
4	Age	Years(1-100)
5	Blood Pressure	(1-130) mmHg
6	Fasting Plasma Glucose	126 mg/dl
7	Two Hours After Eating	Mg/dl
8	Clinical History	If any
9	Hemoglobin A1c	HbA1c(4.0-14.0)
10	Body Mass Index	Kg/height (0-80)
11	Family history	If any
12	Blood Sugar	(mg/dl)
13	Total Cholesterol	(mg/dl)
14	Triglycerides	(mg/dl)
15	Insulin	μIU/mL
16	Diabetes Status	Type –I, Type –II

Data Preprocessing

Family clinical history has the highest percentage of missing results when compared to fasting plasma glucose, two hours after eating, triglycerides, and HbA1c. The locations of missing values in columns were located using a heatmap visualization of the correlation matrix, and a distribution plot was generated to illustrate the distribution of missing data.

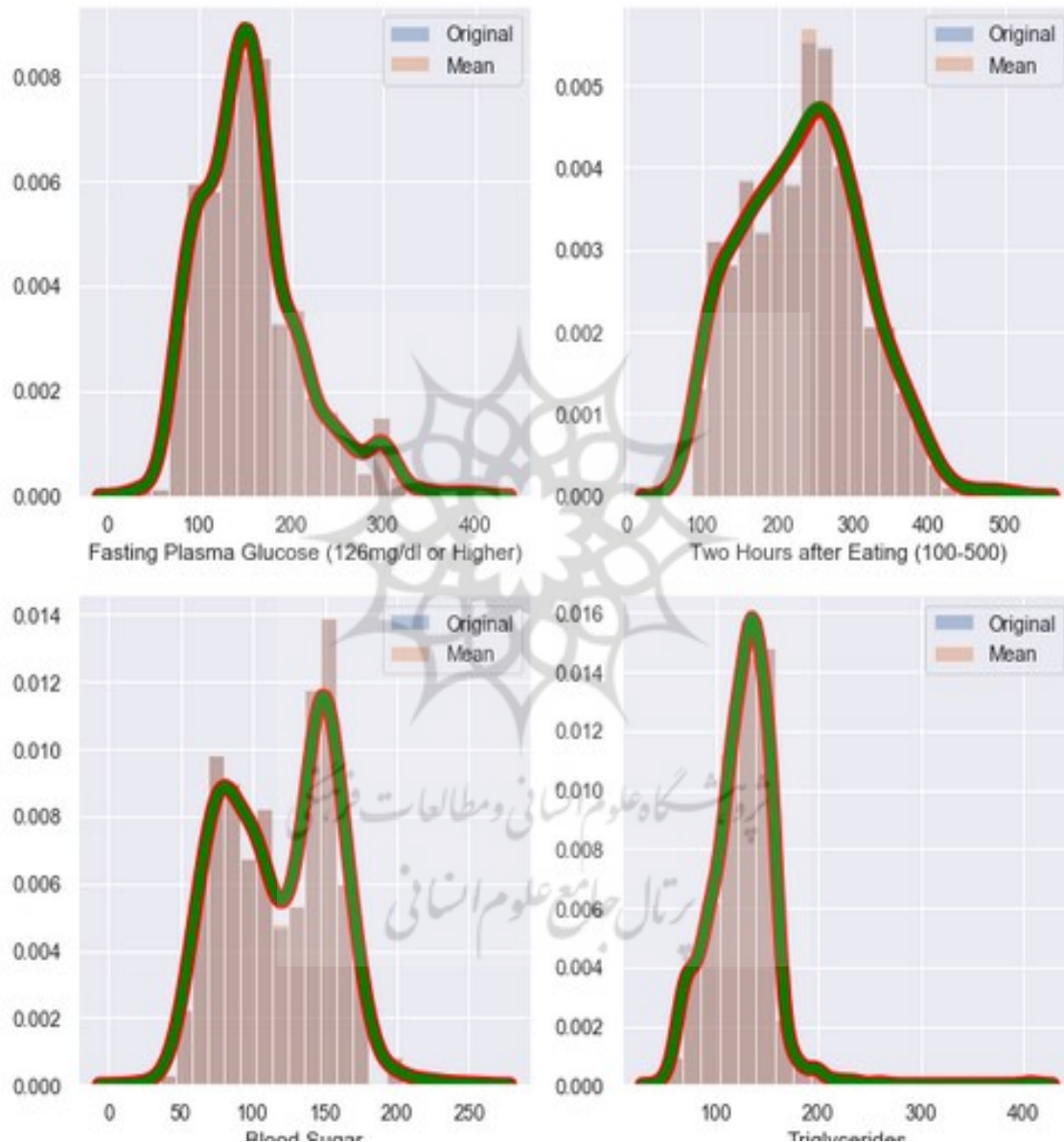


Figure 5. Distribution plot for missing values in dataset

We employed one-hot encoding and label encoding approaches to handle the missing data and transform the family clinical history categorical variable into a numerical form for analysis. Each categorical variable was handled similarly and we used the results as features to enhance our predictions.

We were able to determine the important features of a continuous variable in the diabetic dataset through histograms. We discovered, in particular, that fasting plasma glucose is right-skewed with a few outliers, insulin data is bimodal with some outliers, triglycerides are right-skewed and contain outliers, blood sugar is symmetrically distributed but have more outliers than other variables, and the data for glucose two hours after eating has multiple peaks and a few outliers that need further investigation. Figure 6 & Figure 7 describes the distribution of data using the histogram.

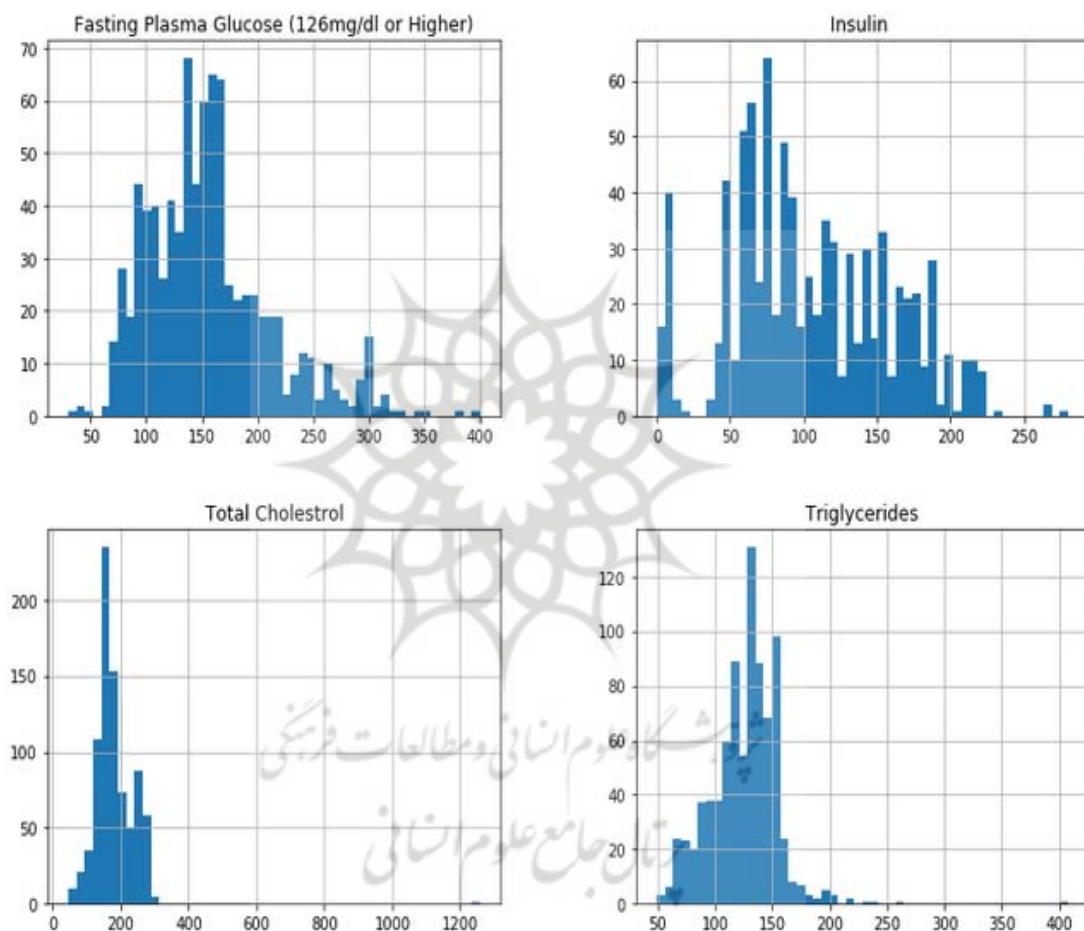


Figure 6. Histogram for fasting plasma glucose, insulin, total cholesterol, triglycerides

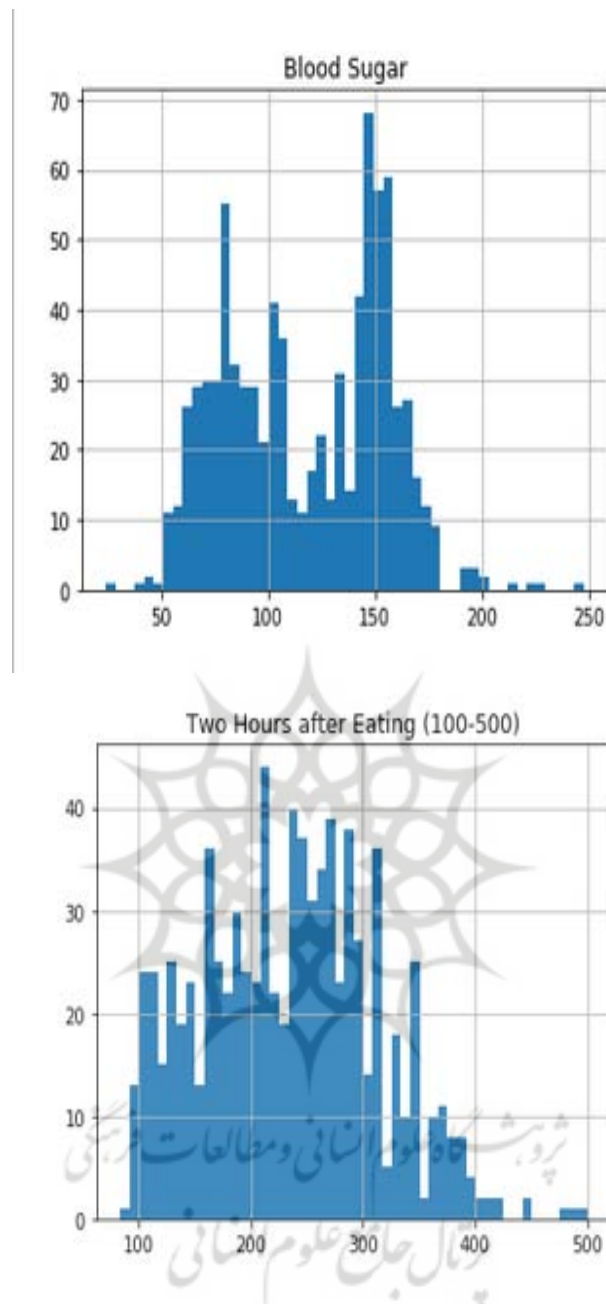


Figure 7. histogram for blood sugar and two hours after eating

It is important to note that outliers can significantly affect statistical analyses, so it's crucial to handle them appropriately.

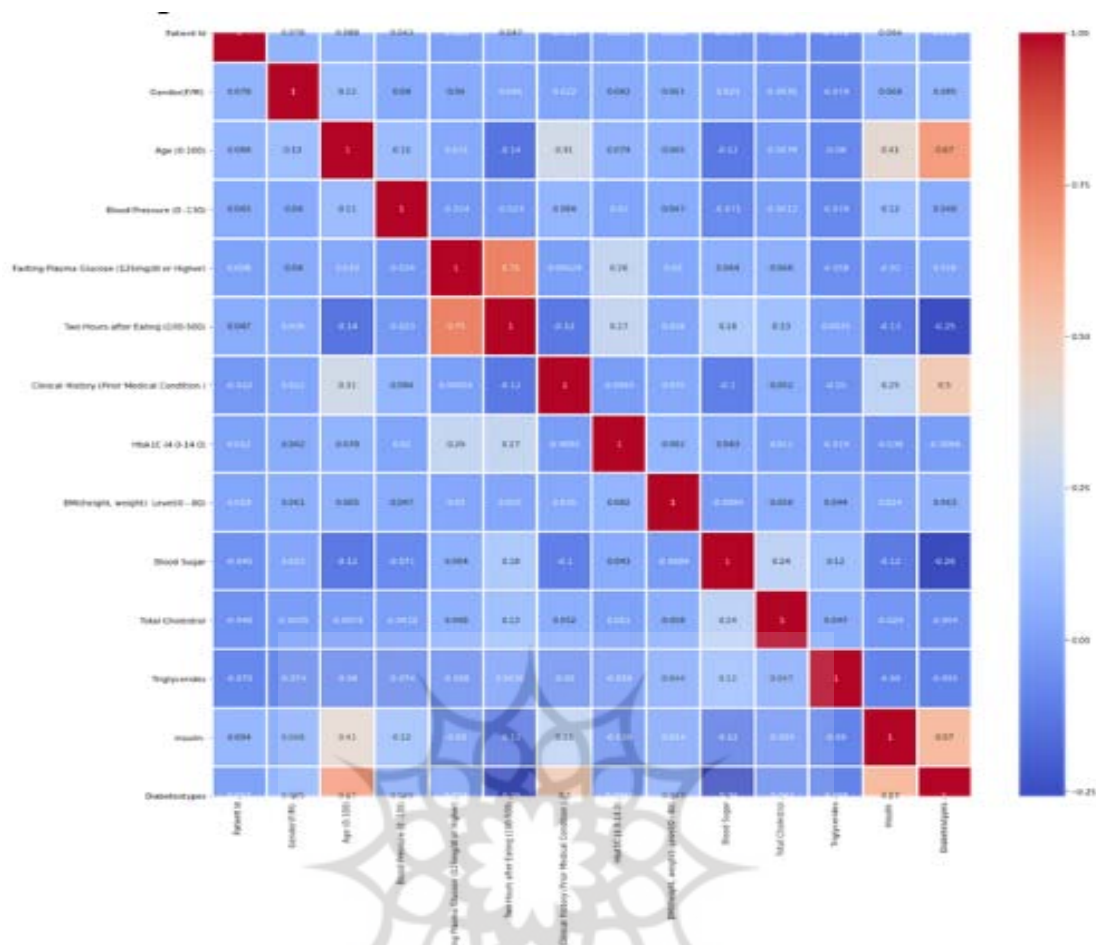


Figure 8. Correlation matrix describing the variables relation with respect to type 1 diabetes patients

RQ1: What are the factors that contributes to the risk of type 1 diabetes among the varied age groups?

Age, insulin, and clinical history were revealed to be the risk factors for Type-1 diabetes by utilizing a heat map of the data. We can build prevention and treatment strategies using this information to better understand the causes of Type-1 diabetes.

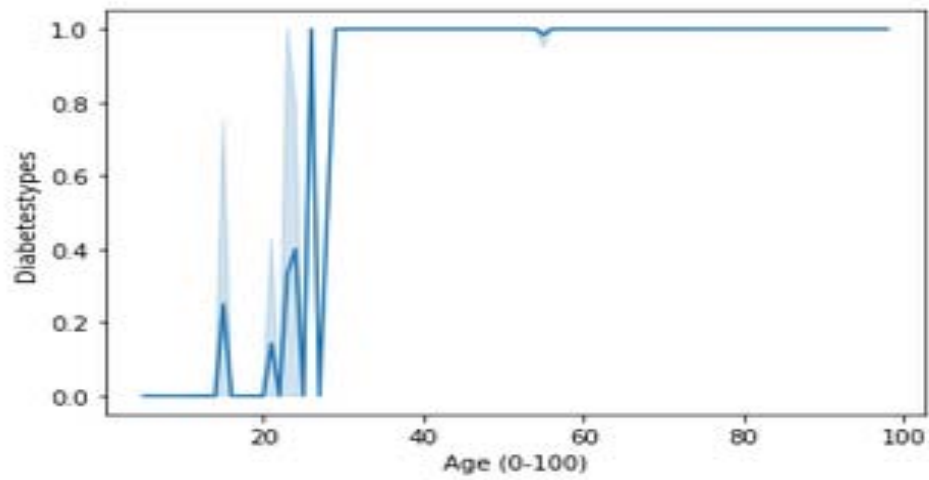


Figure 9. Line plot for Age Vs Diabetes Type

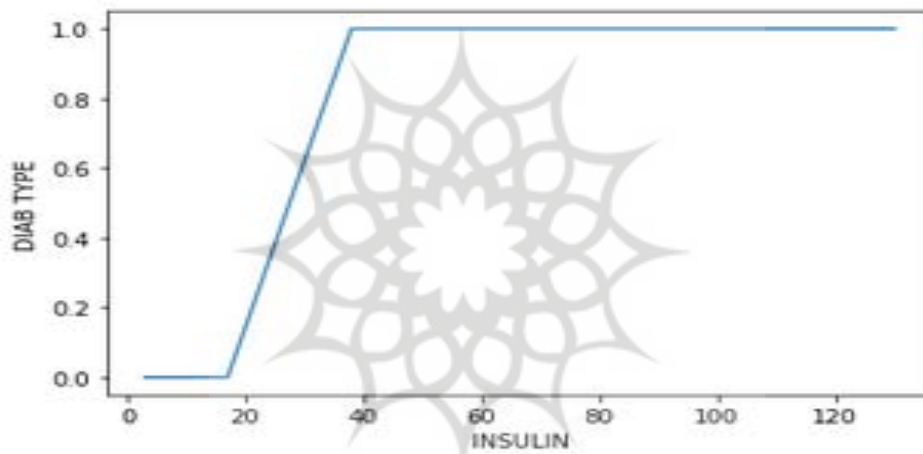


Figure 10. Insulin Vs Diabetes Type

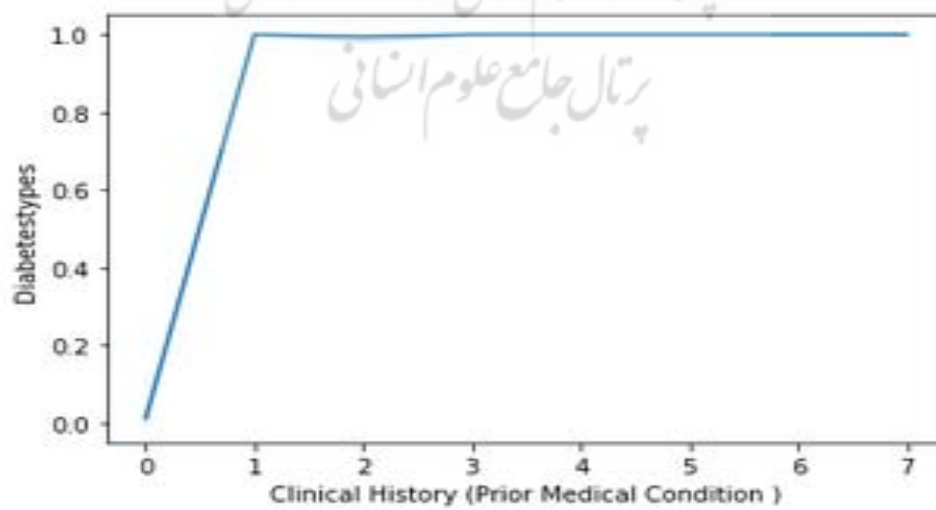


Figure 11. Clinical History Vs Diabetes Types

A deeper analysis was conducted on the three variables - age, insulin, and clinical history against the target variable. The results showed that Type-1 diabetes is more common in lower age groups and the amount of insulin is negligible in Type-1 patients. The plot also revealed that individuals who have an autoimmune disorder are more likely to develop Type-1 diabetes. This information can help healthcare professionals better diagnose and treat Type-1 diabetes, as well as potentially develop preventative measures for individuals at risk (Howsalya Devi et al., 2020).

RQ2: What are the factors that contribute to the risk of type 2 diabetes among the varied age groups?

The analysis showed that age, insulin, and clinical history are important factors for identifying type-II diabetes patients. Glucose and triglyceride levels after eating are also important. Type-II diabetes is more common with increasing age and can lead to hypertension due to unstable blood glucose levels (Bressan et al., 2020).

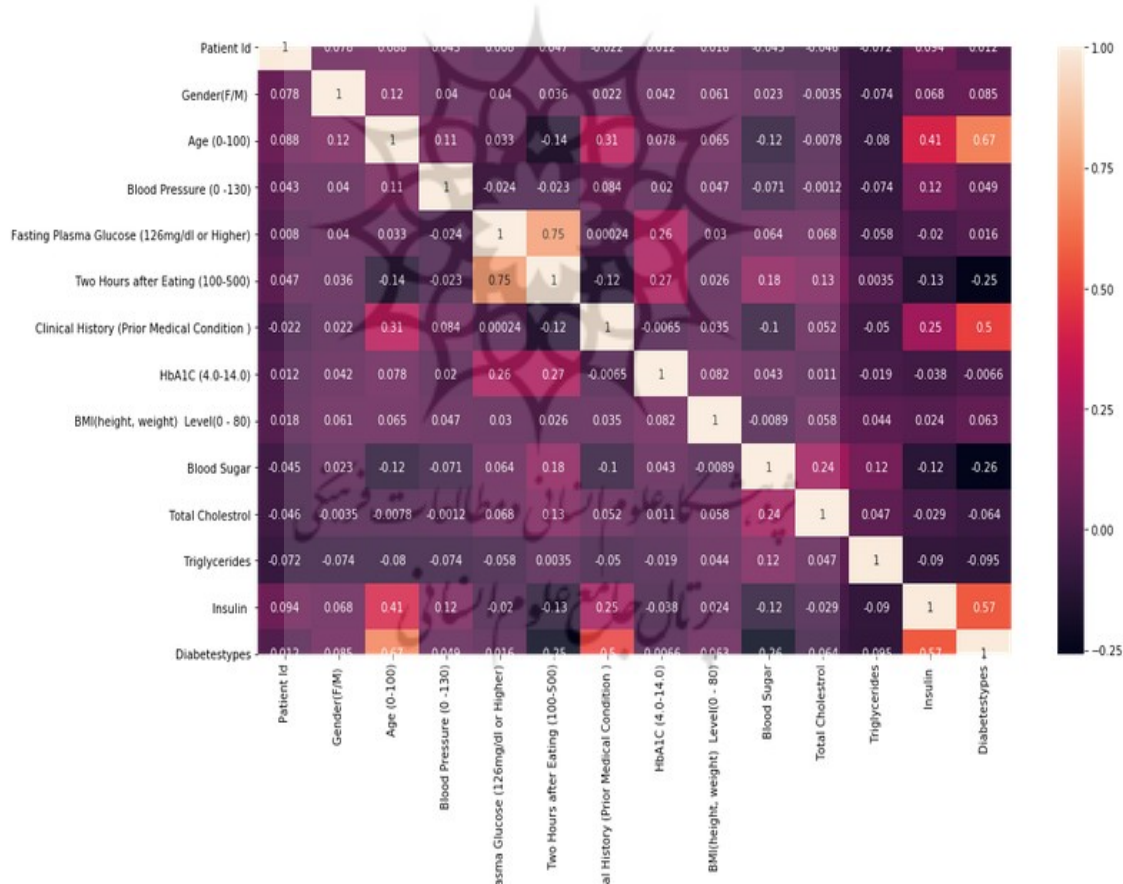


Figure 12. Correlation matrix for diabetes dataset variables with type –II diabetes patients

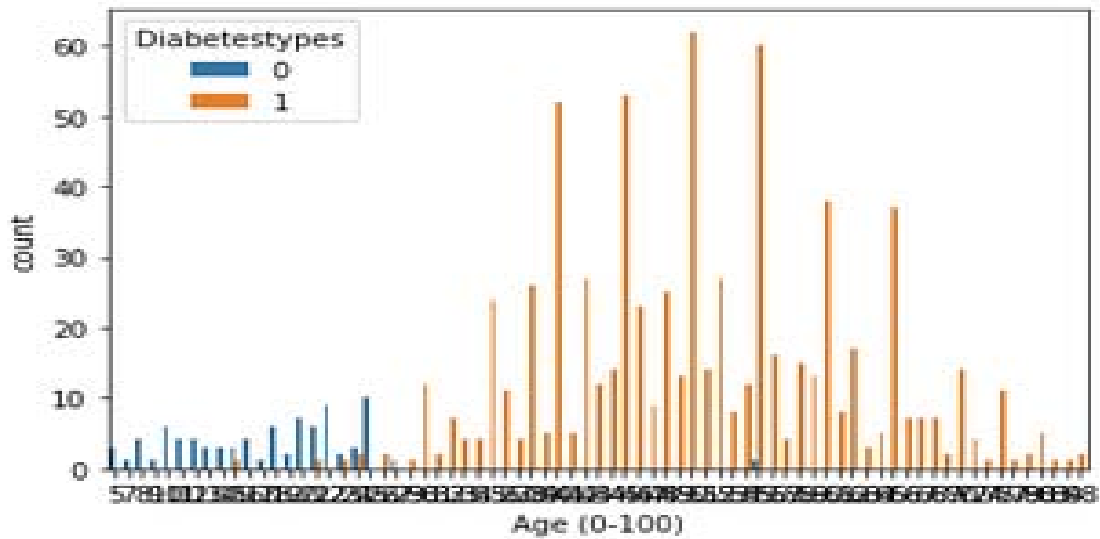


Figure 13. Describing the number of patients identified as diabetes patients for various age groups

The count plot projects the distribution of glucose levels among the type – II diabetes sufferers, which covers the complete graph, while the distribution of glucose levels for patients with type I diabetes has only a little overlap covering the smaller section comparatively, which suggests that glucose levels are a good predictor for distinguishing between the two classes. Since type II diabetes diseases patients are characterized by insulin resistance and high blood glucose levels, patients with type II diabetes patients often have higher glucose levels than those with type-I diabetes individuals.

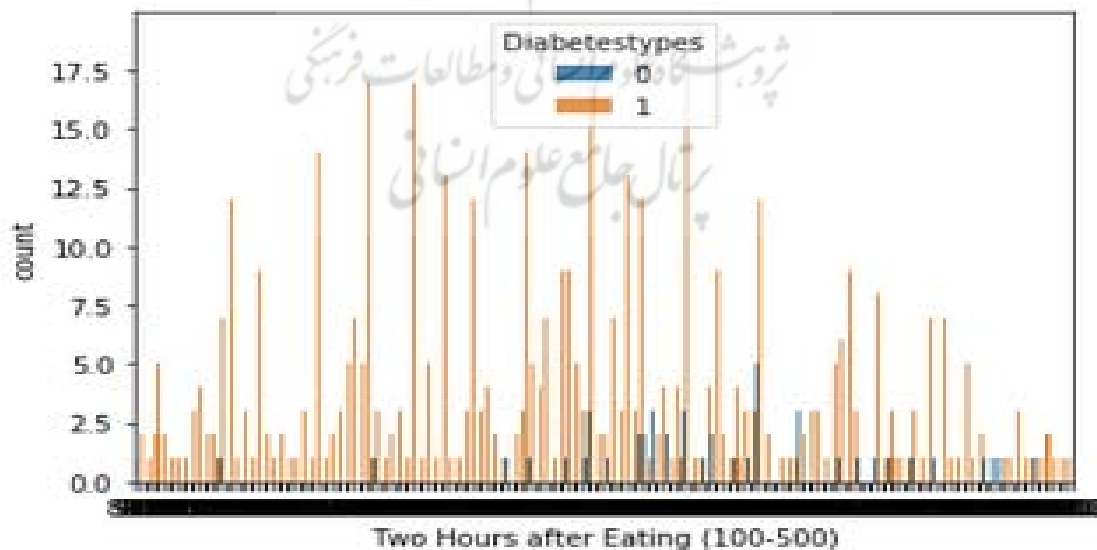


Figure 14. Describes the glucoses fluctuations after two hours among diabetes patients

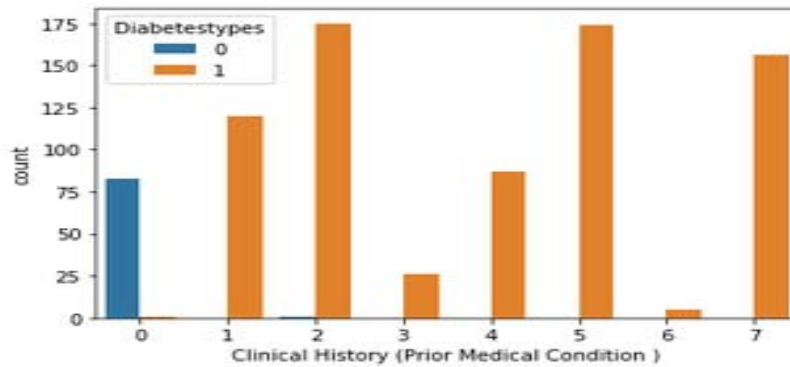


Figure 15. Describing the patients diagnosed as diabetes based on clinical history

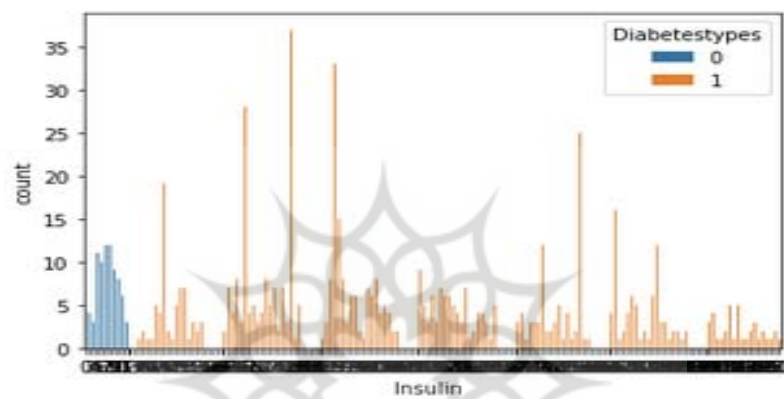


Figure 16. Describing the patients insulin for 2 classes of diabetes

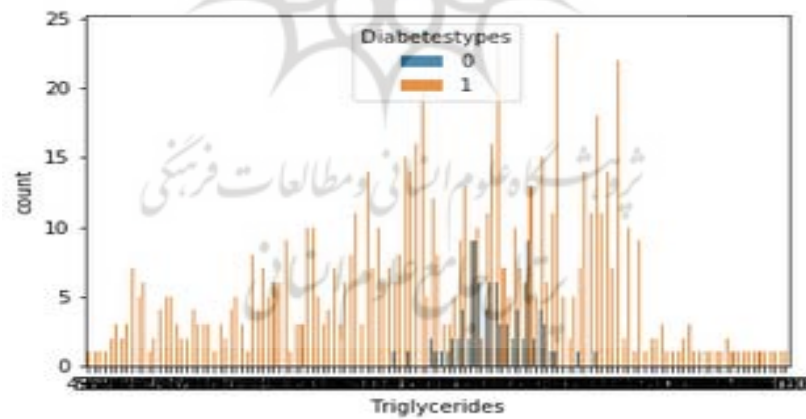


Figure 17. Describing the triglycerides reading for diabetes patients

According to our study, high amounts of total cholesterol and triglycerides accumulate in Type-2 diabetic individuals who have a clinical history of obesity. Furthermore, the findings show that normal to moderate insulin levels are often present in Type-2 diabetic patients.

RQ 3: What are the boundary values for prediction of outliers in risk factors for diabetes using a fuzzy logic system?

The clinical data of the diabetic dataset revealed the presence of 15 outliers. These outliers were observed in variables such as age, insulin, triglycerides, glucose, and blood pressure. We employed the center of gravity defuzzification method to detect uncertain variables in the diabetic dataset. Further analysis showed that some patients had clinical values that were lower than the range set by member functions (Swain et al., 2013), while others had values that exceeded the range. For example, the defuzzification plot of Triglycerides shows more overlap than the plots of other variables like blood pressure and glucose, it may indicate that the Triglycerides variable has more ambiguity or uncertainty associated with it than the other variables. This could potentially lead to ambiguity in the decision-making process of the fuzzy system. It is important to consider the overall performance of the fuzzy logic system; Therefore, it is necessary to explore alternative techniques to identify outliers in the any other diabetes dataset for better predictions.

```
# Defining Constants
#Age (0-100)
age_low =35
age_med = (30,50)
age_high =45
#Blood Pressure (0 -130)
BP_low = 105
BP_med = (95,135)
BP_high = 125
#Fasting Plasma Glucose (126mg/dl or Higher)
Glucose_low = 110
Glucose_med = (100,350)
Glucose_high = 340
#BMI (height, weight) Level(0 - 80)
BMI_low = 55
BMI_med = (50,80)
BMI_high = 75
#Insulin
Insulin_low =25
Insulin_med = (15,330)
Insulin_high = 320
#Total Cholestrol
TC_low =210
TC_med = (200,300)
TC_high = 190
# Triglycerides
TG_low = 130
TG_med = (120,220)
TG_high =210
```

Figure 18. linguistic variable for fuzzy system

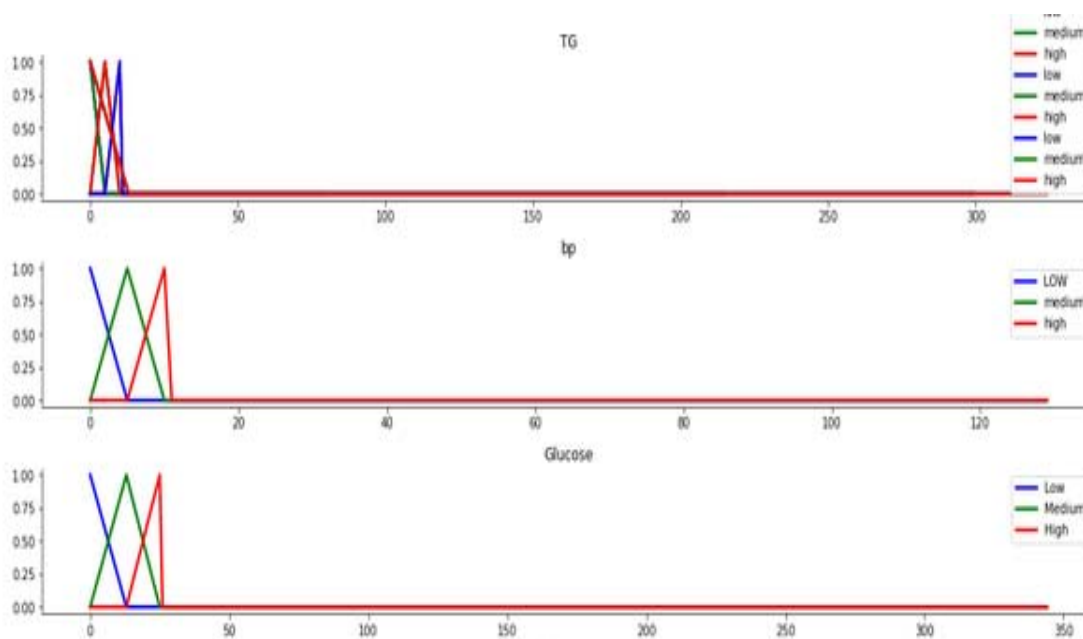


Figure 19. Defuzzification plots for TG, BP,Glucose

RQ4: How can a hybrid approach incorporating fuzzy logic and machine learning algorithms enhance the accuracy of diabetes risk prediction models?

The diabetic dataset describes a binary classification problem with two classes, type-I and type-II diabetic patients, with the dataset being class-imbalanced. To address this class imbalance, the under-sampling technique was adopted, but this approach suffered from the loss of data.



Figure 20. Graphical representation of target parameter

Feature scaling

Standardization and min-max scaling were used in the feature scaling on the diabetes datasets numerical variables. Some rows were removed to enhance the quality of the data for the training and testing procedure.

Table 3. Descriptive analysis of the diabetes dataset

	Patient Id	Gender(F/M)	Age (0-100)	Blood Pressure (0 -130)	Fasting Plasma Glucose (126mg/dl or Higher)	Two Hours after Eating (100-500)	Clinical History (Prior Medical Condition)	HbA1C (4.0-14.0)	BMI(height, weight) Level(0 - 80)	Blood Sugar	Total Cholesterol	Triglycerides	Insulin
count	662.00	662.00	662.00	662.00	662.00	662.00	662.00	662.00	662.00	662.00	662.00	662.00	662.00
mean	0.03	0.59	0.46	0.23	0.34	0.36	0.49	0.35	0.63	0.42	0.11	0.21	0.36
std	0.07	0.49	0.16	0.05	0.15	0.19	0.33	0.16	0.14	0.17	0.05	0.08	0.20
min	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
25%	0.00	0.00	0.38	0.20	0.23	0.21	0.29	0.25	0.54	0.27	0.08	0.16	0.22
50%	0.02	1.00	0.47	0.22	0.32	0.37	0.57	0.38	0.64	0.44	0.10	0.22	0.32
75%	0.03	1.00	0.55	0.26	0.39	0.49	0.71	0.38	0.72	0.56	0.14	0.26	0.51
max	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Feature extraction

Regression and correlation coefficient analyses identified the age, fasting plasma glucose, clinical history, total cholesterol, and triglycerides as influential features. As per literature, the rules in the expert system are used to extract features from the training data, which are then used as input to the machine learning algorithm. Even the defuzzification output values can also be used as a feature or input variable in your machine learning algorithm. For example, you can use the defuzzification output value as a predictor variable in a logistic regression, decision tree, or support vector machine algorithm to predict the likelihood of a patient being type I or type II diabetic. The crisp output labels can be used as the ground truth or target variable in a supervised machine learning algorithm. Before training the machine learning algorithms on the KMC diabetic dataset collected from the Suvadhan laboratory, preliminary activities such as handling missing values, transforming data, scaling features, and feature extraction were done from fuzzy rules. Three different machine learning models were trained to classify patients into two categories of disease (Type-I and Type-II) based on their clinical data.

Performance Evaluation

The three algorithms used for training were support vector regression (SVR), linear regression, and polynomial regression. These algorithms were given fuzzy rules as features to tune the training process with respect to the dataset. The linear regression model achieved a score of 0.73, with a mean square error (MSE) of 0.033 and a root mean square error (RMSE) of 0.1833. The polynomial regression algorithm achieved a score of 0.947, while the support

vector regression algorithm with 'rbf' kernel achieved a score of 0.954, and the one with a 'linear' kernel achieved a score of 0.73 (Lukmanto et al., 2019).

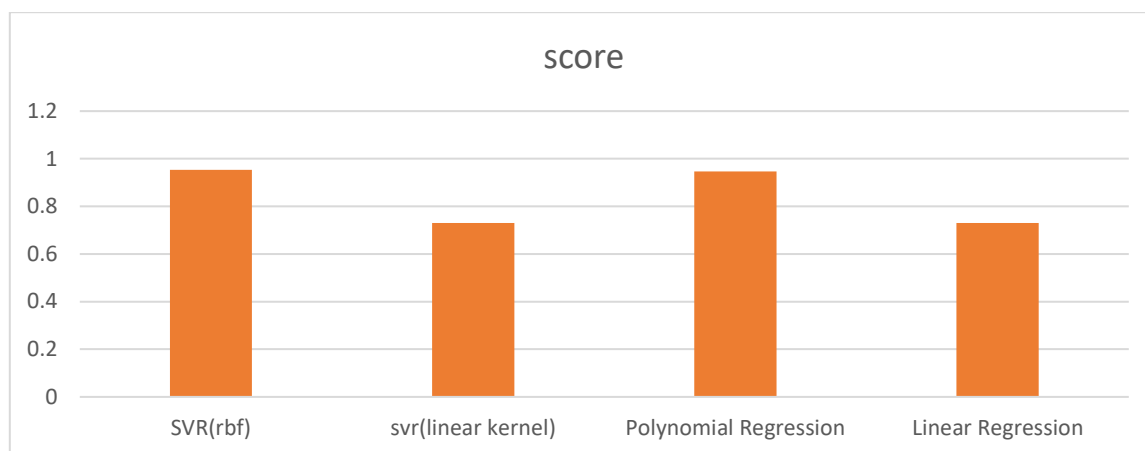


Figure 21. Describing the scores obtained by the machine learning algorithm

Conclusion

Our Machine learning models have performed well and very well utilized the ground truth of the fuzzy system. However, it's important to note that glucose levels can vary widely among individuals, and other factors such as age, lifestyle, and genetics can also contribute to the risk of developing diabetes. Therefore, additional analysis may be necessary to fully understand the relationship between glucose levels and diabetes risk. In future work we may involve investigating into new risk factors, enhancing forecasting techniques, creating more precise and user-friendly diagnostic tools, and assessing the success of remedies. We can work towards reducing the effects of diabetes takes on individuals as well as the community at large by expanding our understanding and putting evidence-based practices into action.

Conflict of interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- A. Yahyaoui, A. Jamil, J. Rasheed, M. Y. (2019). 1st International Informatics and Software Engineering Conference (IISEC-2019): "Innovative Technologies for Digital Transformation": proceedings book : 6-7 November 2019, Ankara/Turkey. 2, 1–4.
- Bressan, G. M., de Azevedo, B. C. F., & de Souza, R. M. (2020). A fuzzy approach for diabetes mellitus type 2 classification. *Brazilian Archives of Biology and Technology*, 63. <https://doi.org/10.1590/1678-4324-2020180742>
- Devi, U. R., & Uma, K. (2019). A Study on Fuzzy Expert System for Diagnosis of Diabetes Mellitus. *International Journal of Applied Engineering Research (IJAER)*, 14(4), 129–139. https://www.ripublication.com/ijaerspl2019/ijaerv14n4spl_16.pdf
- Geman, O., Chiuchisan, I., & Todorean, R. (2017). Application of Adaptive Neuro-Fuzzy Inference System for diabetes classification and prediction. 2017 E-Health and Bioengineering Conference, EHB 2017, Dm, 639–642. <https://doi.org/10.1109/EHB.2017.7995505>
- Howsalya Devi, R. D., Bai, A., & Nagarajan, N. (2020). A novel hybrid approach for diagnosing diabetes mellitus using farthest first and support vector machine algorithms. *Obesity Medicine*, 17, 100152. <https://doi.org/10.1016/j.obmed.2019.100152>
- Khalil, R. M., & Al-Jumaily, A. (2017). Machine learning based prediction of depression among type 2 diabetic patients. *Proceedings of the 2017 12th International Conference on Intelligent Systems and Knowledge Engineering, ISKE 2017, 2018-Janua*, 1–5. <https://doi.org/10.1109/ISKE.2017.8258766>
- Khan, T., Singh, K., Manjul, M., Ahmad, M. N., Zain, A. M., & Ahmadian, A. (2022). A Temperature-Aware Trusted Routing Scheme for Sensor Networks: Security Approach. *Computers & Electrical Engineering*, 98, 107735.
- Kumar, A., Singh, K., & Khan, T. (2021). L-RTAM: Logarithm based reliable trust assessment model for WBSNs. *Journal of Discrete Mathematical Sciences and Cryptography*, 24(6), 1701-1716.
- Kumar, A., Singh, K., Khan, T., Ahmadian, A., Saad, M. H. M., & Manjul, M. (2021). ETAS: an efficient trust assessment scheme for BANs. *IEEE Access*, 9, 83214-83233.
- Lukmanto, R. B., Suharjito, Nugroho, A., & Akbar, H. (2019). Early detection of diabetes mellitus using feature selection and fuzzy support vector machine. *Procedia Computer Science*, 157, 46–54. <https://doi.org/10.1016/j.procs.2019.08.140>
- Niswati, Z., Mustika, F. A., & Paramita, A. (2018). Fuzzy logic implementation for diagnosis of Diabetes Mellitus disease at Puskesmas in East Jakarta. *Journal of Physics: Conference Series*, 1114(1). <https://doi.org/10.1088/1742-6596/1114/1/012107>
- Raj, R. S., Sanjay, D. S., Kusuma, M., & Sampath, S. (2019). Comparison of Support Vector Machine and Naïve Bayes Classifiers for Predicting Diabetes. 1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2019, 41–45. <https://doi.org/10.1109/ICATIECE45860.2019.9063792>
- Rajeswari, A. M., Sidhika, M. S., Kalaivani, M., & Deisy, C. (2018). Prediction of Prediabetes using Fuzzy Logic based Association Classification. *Proceedings of the*

- International Conference on Inventive Communication and Computational Technologies, ICICCT 2018, Icticct, 782–787. <https://doi.org/10.1109/ICICCT.2018.8473159>
- Sarwar, M. A., Kamal, N., Hamid, W., & Shah, M. A. (2018). Prediction of diabetes using machine learning algorithms in healthcare. ICAC 2018 - 2018 24th IEEE International Conference on Automation and Computing: Improving Productivity through Automation and Computing, September, 1–6. <https://doi.org/10.23919/IConAC.2018.8748992>
- Swain, A., Mohanty, S., & Das, A. (2013). COMPARATIVE RISK ANALYSIS ON PREDICTION OF DIABETES MELLITUS USING MACHINE LEARNING APPROACH. 1, 3312–3317. International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) - 2016
- Thakkar, H., Shah, V., Yagnik, H., & Shah, M. (2021). Comparative anatomization of data mining and fuzzy logic techniques used in diabetes prognosis. *Clinical EHealth*, 4, 12–23. <https://doi.org/10.1016/j.ceh.2020.11.001>
- Undre, P., Kaur, H., & Patil, P. (2015). Improvement in prediction rate and accuracy of diabetic diagnosis system using fuzzy logic hybrid combination. 2015 International Conference on Pervasive Computing: Advance Communication Technology and Application for Society, ICPC 2015, 00(c). <https://doi.org/10.1109/PERVASIVE.2015.7087029>
- Verma, D, M. N. (2017). using Data mining classification Techniques. 2017 International Conference on Intelligent Sustainable Systems (ICISS), Iciss, 533–538.
- Vijiyakumar, K., Lavanya, B., Nirmala, I., & Sofia Caroline, S. (2019). Random forest algorithm for the prediction of diabetes. 2019 IEEE International Conference on System, Computation, Automation and Networking, ICSCAN 2019, 1–5. <https://doi.org/10.1109/ICSCAN.2019.8878802>

Bibliographic information of this paper for citing:

Pattun, Geeta; Afroaz, Khaleda; Siddiqui, Ahmad Talha & Ghazala, Shaheena (2023). Prediction of Type - I and Type –II Diabetes: A Hybrid Approach using Fuzzy Logic and Machine Learning Algorithms. *Journal of Information Technology Management*, 15 (Special Issue), 35-56. <https://doi.org/10.22059/jitm.2023.95244>
