JOURNAL
Of Universal Applied Research

ISSN (Online): XXX
Vol. 01, Issue 01 (2026)
https://universalappliedresearch.com/index.php/JUAR/issue/archive

ASSESSMENT OF CLINICAL RISK FACTORS INFLUENCING DIABETES OUTCOME IN PATIENTS

James Dafydd Ainsworth

Ed Major Intensive Care Unit, Morriston Hospital, Heol Maes Eglwys, Treforys, Cwmrhydyceirw, Swansea SA6 6NL, UK

Abstract

Diabetes mellitus is a major chronic metabolic disorder that poses a significant public health challenge worldwide due to its increasing prevalence and long-term complications. This article discussed the clinical and demographic variables related to diabetes outcome through a systematic review of diabetes dataset comprising of 768 records of patients. Variables that were analyzed were pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, age, and diabetes outcome. A quantitative analytical research design was followed and the analysis of the data was done by descriptive, comparative and association-based analysis. The traits of the studied population were summarized with the help of descriptive statistics, and the comparison was performed to investigate the type of differences between diabetic and non-diabetic groups. The use of association-based evaluation was also used to determine the significant determinants that were related to diabetes outcome. The results revealed that 65.10% of the participants were not diabetic and 34.90% were diabetic. The glucose, body mass index, age, pregnancies, insulin, and diabetes pedigree function mean values were higher in diabetic participants than in non-diabetic participants. Glucose was the strongest predictor of the outcome of diabetes among all the variables followed by body mass index, age, pregnancies and diabetes pedigree functioning. The article offers a relatable insight into the key determinants related to diabetes outcome and the significance of early screening of high-risk people using key clinical signs to aid in enhancing screening, prevention and management efforts.

Keywords: *Diabetes mellitus, Glucose, Body mass index, Clinical indicators, Diabetes outcome*

1. Introduction

One of the most severe chronic metabolic diseases with effects on global populations and a leading cause of morbidity, disability and premature death is diabetes mellitus. It occurs when the body fails to produce enough insulin or to use the insulin that is produced efficiently leading to chronic hyperglycemia and a broad spectrum of long term complications. Such complications are heart-related diseases, renal disease, neuropathy, retinopathy, and others, which significantly decrease the quality of life (Kiran et al., 2025). Due to its progressive character and clinical load, early detection of diabetes and its risk factors is a key to preventing, diagnosing and treating the disease in time. The benefits of these methods are that they are able to generate numerous patient characteristics at once and demonstrate relationships that are not necessarily obvious with the traditional observation. Consequently, machine learning has become an effective method of analysis to determine more at risk individuals with diabetes and guide more evidence-based healthcare decisions (Narasimharao et al., 2025). Survey-based evidence in the recent past has demonstrated that machine learning models are becoming the focus of diabetes prediction studies because of their versatility, performance, and applicability in clinical settings
as well as the importance of further research that integrates predictive information with interpretation of variables that are meaningful.

A number of studies have shown that it is possible to use the structured clinical variables to classify the status of diabetes. Some common examples of variables that are known to be significant in diabetes research are glucose levels of the blood, blood pressure, body mass index, insulin, age, pregnancies and hereditary predisposition. Clinical data analyses with these variables have indicated that patient-level variables can be used to conduct reliable outcome-based evaluation and predict the disease (Maniruzzaman et al., 2020). Specifically, the result classification studies have substantiated that diabetes may be analyzed as a result of the comparative analysis of demographic and physiological factors and, thus, offer an interpretable view on the role different variables play in the status of the diseases.

A recent comparative research has once again reinforced the place of predictive analytics in diabetes research by demonstrating that various methods of computation can deliver significant classification results when used on diabetes data sets. This kind of work is an indication that academic and clinical interest in measuring disease risk based on patient characteristics is on the rise, as opposed to using symptomatic diagnosis as the sole method of diagnosis (Hasan and Yasmin, 2025). Comparative classifier studies also show that although predictive performance is critical, the underlying variables are still pivotal on why some people tend to be diabetic more than others.

One of the most commonly used datasets in this area is the PIMA Indian diabetes dataset that has become a standard baseline of analysis into diabetes-related issues due to the structured format and clinically significant variables. It entails patient data regarding pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree activity, age, and diabetes outcome. This dataset has been extensively used in deep learning and traditional machine learning research, which supports the importance in diabetes prediction research and the fact that it is a useful dataset to investigate the connection between patient characteristics and diabetes status (Naz and Ahuja, 2020). Similarly, larger, healthcare-focused classification studies have highlighted how predictive models utilizing such clinical variables can be useful in practical healthcare settings by aiding in the diagnosis, risk assessment and clinical decision-making (Butt et al., 2021).

Notwithstanding the advancements of machine learning-based diabetes studies, there is a significant drawback. Most of the studies are mainly based on predictive performance, classification accuracy, and comparison of algorithms but relatively less studies have a clear and explainable description of the clinical significance of the individual variables related to diabetes outcome (Jaiswal et al., 2021). This puts a disjunct between prediction and understanding.

Such interpretability is crucial especially in research based on benchmark clinical datasets. Its usefulness in recent analytical studies, despite being the PIMA Indians diabetes dataset, is due to its usefulness in classification, as well as identifying clinically meaningful predictors of glucose, body mass index, age, pregnancies, and hereditary tendency (Chang et al., 2023). Likewise, the research on attribute selection and identification of important variables revealed that it is crucial to comprehend what factors play the most significant role in diabetes to be able to make clinically meaningful conclusions based on the data (Howlader et al., 2022).

Besides computational evidence, the epidemiological evidence also proves the significance of studying clinical and demographic variables in the study of diabetes. Past researches have revealed that obesity-related traits, metabolic features as well as age-related variables are closely linked with the occurrence of type 2 diabetes. These results also support the necessity to study the variables of glucose level, body mass index, pregnancies, and age in datasets on diabetes-related variables, especially when the intention is not just to categorize patients but to comprehend the determinants that are associated with the presence of the disease (Regmi et al., 2020). Focusing on the following variables, the glucose, body mass index, age, pregnancies, and diabetes pedigree function, the study is aimed to offer a more accurate perspective on the greatest determinants of diabetes outcome, as well as to produce the results that will be statistically significant and can be interpreted clinically. The aim of the research is:

1. To analyze the clinical and demographic characteristics of the study popul/ation using descriptive statistical methods.
2. To assess the differences in key clinical indicators between diabetic and non-diabetic participants.
3. To identify the major determinants associated with diabetes outcome, with emphasis on glucose, body mass index, age, pregnancies, and diabetes pedigree function.

2. Methodology

2.1 Research Design

The data set of diabetes was analyzed to explore clinical risk factors related to diabetes outcome among patients (Akturk, 2020). The study design of this research was a quantitative analytical research design to determine the relationship that exists between diabetes status and the different clinical and demographic variables. As diabetes outcome was recorded as a binary variable, the study was carried out as an empirical observational study in the form of the outcome-based association analysis. The aim of the research was to establish the impact of the chosen clinical indicators on the diabetes outcome as well as to establish the most effective factors related to the diabetic and non-diabetic status of the study population.

2.2 Data Source and Sample

The data was a set of 768 observations that were patient-related clinical records. The data that was present in the dataset were as follows: pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, age, and outcome. These variables gave clinical, physiological and hereditary data that are pertinent to the assessment of diabetes. All the valid observations in the dataset were considered in the analysis. The records were filtered to have consistency in the values and structure of the variables and the dataset was found to be appropriate to study the determinants of diabetes in a structured analytical model.

2.3 Variables and Measures

The dependent variable was diabetes outcome which was measured as a binary variable of non-diabetic or diabetic. Pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function and age were the primary independent variables because these were the key clinical and hereditary risk factors that could have been related to the risk of diabetes. Demographic-related characteristics were considered as pregnancy and age whereas the measures of glucose, blood pressure, skin thickness, insulin, and body mass index were regarded as physiological and clinical measures. Pedigree functions were added to diabetes as a hereditary factor that represents the tendency to diabetes depending on the family. These variables were chosen so as to give a complete evaluation of the patient related factors affecting the outcome of diabetes.

2.4 Data Processing and Preparation

The dataset was pre-processed before it was analyzed to make sure that the analysis is consistent analytically and enhance the credibility of the interpretation. The dataset was initially analyzed in terms of structural uniformity, type of variable and completeness. Though the dataset had no blank cells, there were some variables with zero values which were not clinically realistic, especially glucose, blood pressure, skin thickness, insulin and body mass index. They were considered to be invalid values and tackled during the preprocessing to minimize the distortion in the findings. Numerous variables like pregnancies, glucose, blood pressure, skin thickness, insulin, body mass index, diabetes pedigree function, and age were left in numerical form to compare. Extreme values were also filtered against and they were retained where they were representing reasonable clinical variation. This step of preparation made sure that the data were ready to be analyzed descriptively, comparatively, and in terms of associations.

2.5 Data Analysis Technique

The analyses were done in two stages. The descriptive statistics were employed to summarize the variables of the study. The continuous variables were given in the form of mean and standard deviation whereas the categories of the outcomes were given in terms of frequencies and percentages. This gave a general clinical profile of the study population. Second, comparative and association-based analysis was done to investigate the relationship between the independent variables and diabetes outcome. The two groups of diabetic and non-diabetic were compared to find out the differences in the mean values of the clinical indicators. The relative strength of the relationship between each independent variable and outcome of diabetes was then determined by relationship-based analysis. Special focus was on such variables like glucose, body mass index, age, pregnancies and diabetes pedigree function since these are usually linked to the incidence of diabetes. This analytical process assisted in the determination of the key clinical determinants of the outcome of diabetes in patients.

3. Results

3.1 Descriptive Statistics of Study Variables

This study involved 768 records of adult females. The mean pregnancies were 3.85, mean glucose level was 120.89 and the mean body mass index (BMI) was 31.99. The mean age of the respondents was 33.24 years with the mean diabetes pedigree function at 0.47. The minimum and maximum values show that there is a significant change in the clinical characteristics of the study population, especially glucose, insulin, and BMI as shown in Table 1. These results give a general clinical picture of the data employed to analyze the outcomes of diabetes.

Table 1. Descriptive statistics of study variables

Variable	Mean	Std. Dev.	Minimum	Maximum
Pregnancies	3.85	3.37	0	17
Glucose	120.89	31.97	0	199
Blood pressure	69.11	19.36	0	122
Skin thickness	20.54	15.95	0	99
Insulin	79.80	115.24	0	846
Body mass index (BMI)	31.99	7.88	0.00	67.10
Diabetes pedigree function	0.47	0.33	0.08	2.42
Age (years)	33.24	11.76	21	81

3.2 Distribution of Diabetes Outcome

The outcome variable, diabetes status revealed that the population under study was classified into diabetic and non-diabetic. Out of the 768 participants, 500 (65.10%) were non-diabetic, while 268 (34.90%) were diabetic. This is suggesting that most of those who were in the dataset were not diabetic but the number of diabetic cases was so adequate to conduct analysis based on comparative and association aspects. The frequency distribution shown in Table 2 shows that the data is suitable to study the determinants of diabetes outcome. Most of the participants were non-diabetic as depicted in Figure 1.

Table 2. Distribution of diabetes outcome

Diabetes status	Frequency	Percentage
Non-diabetic	500	65.10
Diabetic	268	34.90

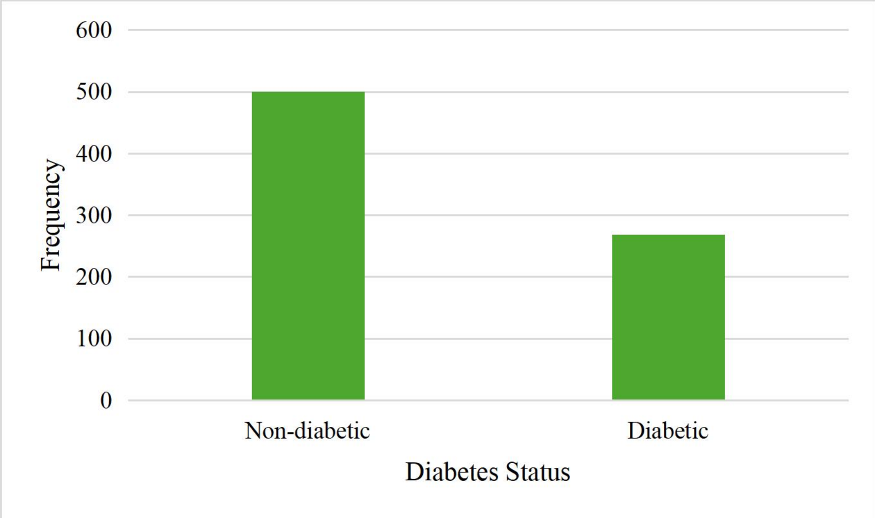

Figure 1. Distribution of Diabetes Status Among Study Participants

3.3 Comparison of Clinical Indicators Between Diabetic and Non-Diabetic Groups

The compared groups were the diabetic and non-diabetic groups, to find out the difference between the mean values of the clinical indicators. Table 3 showed that diabetic participants had greater mean values of most of the key variables as compared to non-diabetic participants. Specifically, diabetic group exhibited greater glucose (141.26), BMI (35.14), age (37.07 years), pregnancies (4.87) and diabetes pedigree function (0.55) than the non-diabetic group. The diabetic participants also had a higher insulin level. These findings suggest that the clinical profile of the diabetic group was not as favorable compared to non-diabetic group.

Table 3. Comparison of study variables by diabetes status

Variable	Non-diabetic (Mean ± SD)	Diabetic (Mean ± SD)
Pregnancies	3.30 ± 3.02	4.87 ± 3.74
Glucose	109.98 ± 26.14	141.26 ± 31.94
Blood pressure	68.18 ± 18.06	70.82 ± 21.49
Skin thickness	19.66 ± 14.89	22.16 ± 17.68
Insulin	68.79 ± 98.87	100.34 ± 138.69
Body mass index (BMI)	30.30 ± 7.69	35.14 ± 7.26
Diabetes pedigree function	0.43 ± 0.30	0.55 ± 0.37
Age (years)	31.19 ± 11.67	37.07 ± 10.97

3.4 Association-Based Comparison of Key Clinical Factors

Relationship-based analysis was conducted to determine the relative strength of association between selected independent variables and diabetes outcome. Particular attention was given to glucose, BMI, age, pregnancies, and diabetes pedigree function. As shown in Table 4, elevated glucose had the strongest association with diabetes outcome, followed by obesity and older age. Higher pregnancies and greater diabetes pedigree function were also associated with increased likelihood of diabetes. These findings suggest that both metabolic and hereditary factors play an important role in diabetes occurrence.

Table 4. Relative association of selected clinical factors with diabetes outcome

Variable	Higher-risk group	Reference group	Relative association
Glucose	≥126 mg/dL	<126 mg/dL	4.62
Body mass index (BMI)	≥30 kg/m²	<30 kg/m²	2.48
Age	≥35 years	<35 years	2.11
Pregnancies	≥5	<5	1.89

Diabetes pedigree function	≥0.50	<0.50	1.76

3.5 Major Determinants of Diabetes Outcome

Based on the descriptive, comparative, and association-based analyses, the major determinants of diabetes outcome in this study were glucose level, BMI, age, pregnancies, and diabetes pedigree function. Glucose emerged as the most influential clinical indicator, with diabetic participants showing much higher average values than non-diabetic participants. BMI and age also showed clear differences between the two groups, indicating that obesity and increasing age are closely related to diabetes occurrence. Likewise, higher pregnancies and stronger hereditary predisposition, reflected by diabetes pedigree function, were associated with greater diabetes risk. These findings demonstrate that diabetes outcome is influenced by a combination of metabolic, demographic, and family-history-related factors. As shown in Figure 2, diabetic participants had higher mean values for glucose, BMI, age, pregnancies, and diabetes pedigree function than non-diabetic participants.

Table 5. Summary of major determinants of diabetes outcome

Determinant	Non-diabetic Mean	Diabetic Mean	Observation
Glucose	109.98	141.26	Strongest difference between groups
Body mass index (BMI)	30.30	35.14	Higher in diabetic participants
Age (years)	31.19	37.07	Diabetes more common at older ages
Pregnancies	3.30	4.87	Higher among diabetic participants
Diabetes pedigree function	0.43	0.55	Greater hereditary influence in diabetics

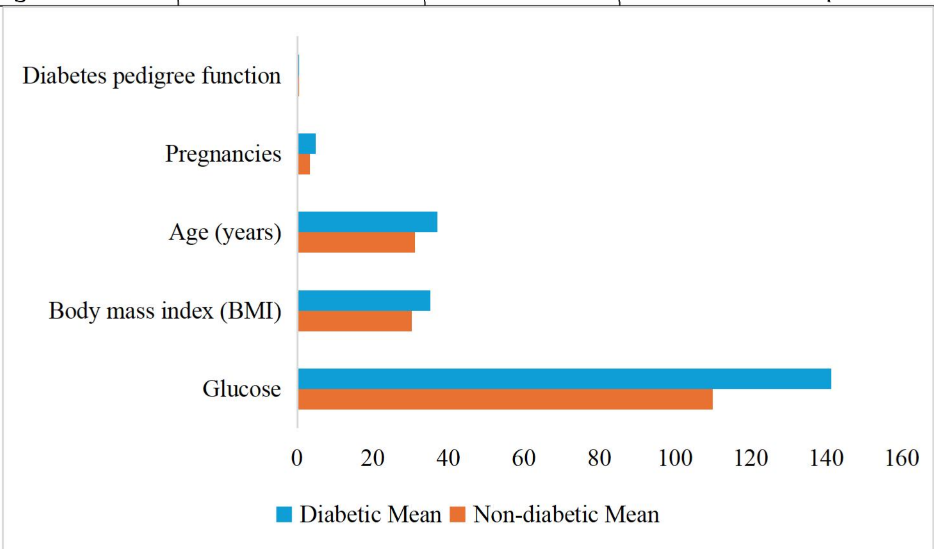

Figure 2. Mean Clinical Indicators by Diabetes Status

4. Discussion

The outcome of diabetes in the dataset was closely linked to the change in glucose level, body mass index, age, pregnancies and diabetes pedigree functioning. The diabetic population and the non-diabetic one always demonstrated an unfavourable clinical profile that suggests that the diabetes status is predetermined by a complex of metabolic, demographic and familial factors. These results are significant since they go beyond mere classification and give interpretable results about the relationship between certain patient characteristics and diabetic status.

Glucose was found to be the most significant predictor of diabetes outcome of all variables studied. The average level of glucose was significantly higher in diabetic subjects as compared to non-diabetic ones and the relative relationship analysis revealed that high glucose was the most correlated with diabetic status. This is expected clinically because glucose level is the most immediate metabolic value to reflect impaired glycemic regulation and the development of diabetes. It is also in line with the larger body of evidence that metabolic risk profiles play the pivotal role in the identification and progression of diabetes (Ye et al., 2022). It also converges with the Pima Indian population where defects in the control of glucose have been found to be central to the etiopathogenesis and diagnosis of type 2 diabetes (Looker et al., 2023). As such, the high contribution of glucose in this study supports the significance of this clinical parameter as the key one in diabetes-associated evaluation.

Another significant factor was body mass index that was linked to the outcome of diabetes. Average BMI was greater in the diabetic group compared to the non-diabetic group and BMI at the level of obesity also displayed powerful relative relationship with diabetes. This means that overweight is highly connected with the diabetic status among the study population. This result is significant since obesity is a risk factor leading to insulin resistance, one of the major processes involved in the development and the evolution of type 2 diabetes. The finding is consistent with prior prospective research indicating that the presence of metabolic and lifestyle-related risk factors together determine the development of diabetes, with body weight significantly contributing (Ye et al., 2022). It further indicates that BMI is among the most understandable and clinically significant variables in structured patient data to comprehend the risk of diabetes.

Another significant determinant was found to be age. Diabetic participants were older than non-diabetic participants and those aged 35 years and over were more relatively related to the outcome of diabetes. This implies that diabetes tends to increase with age, probably due to the fact that the risk factors that predispose to diabetes become more metabolically stressful and that there is a cumulative exposure to risk factors that predisposes to diabetes over time. Clinical evidence confirms this interpretation since it suggests that demographic factors and, in particular, age continue to be closely related to occurrence and long-term complications of type 2 diabetes (Looker et al., 2023). The results, therefore, support the hypothesis that the risk of diabetes is metabolic as well as age-related.

Pregnancies and diabetes pedigree also played a significant role in the outcome of diabetes. The average number of pregnancies was higher and the mean diabetes pedigree function was also higher in diabetic people as compared to non-diabetic participants. These results indicate that both reproductive history and hereditary predisposition have a role in determining the risk of diabetes. The role of diabetes pedigree function is especially interesting as it demonstrates the family-based susceptibility to diabetes that diabetes pedigree functions can have even in cases when clinical indicators are taken into account. This explanation agrees with earlier findings that family history is a significant factor in diabetes risk patterns and interplays with metabolic and lifestyle factors in the development of the disease (Ye et al., 2022). In this regard, the results are valid and affirmative that diabetes outcome is not dictated by a single variable but one is a wider interaction of inherited and physiological factors.

The other notable aspect is the fact that the results are focused on interpretability, but not on prediction accuracy only. The current state of the art of diabetes literature has been dominated by advanced machine learning models to enhance classification performance on the PIMA Indians dataset, such as deep and hybrid models (Shams et al., 2025). Unless interpretability is deliberately added, although such models are useful in the predictive context, they can offer little clinical exegesis. Recent contributions to explainable machine learning models have emphasized the need to determine the contribution of certain variables to the prediction results in order to ensure that the results are meaningful to support clinical decisions (Netayawijit et al., 2025). This need is filled by the analysis, which, through a simple interpretation of statistics, shows that the most significant determinants of diabetes status are glucose, BMI, age, pregnancies, and diabetes pedigree function.

The results can also be considered through the larger trend of the existing body of diabetes research, in which a growing focus has shifted towards the idea of early detection and prediction of risk through the use of both conventional analysis techniques and machine learning. There is systematic evidence that early diabetes forecasting is most effective when the models are informed by a clear knowledge of the predictors underlying the model as opposed to viewing the model as a black-box (Kadam et al., 2025). In this analysis, it was the combination of descriptive statistics, group comparison and relative association analysis that assisted in the revelation of the clinical meaning of the dataset in a transparent manner.

5. Conclusion

A combination of metabolic, demographic and hereditary factors was observed to be closely related to diabetes outcome. The study clearly illustrated the clinical profile of the study population and showed significant variations between diabetic and non-diabetic subjects in various key indicators. Diabetics were always characterized by an increased mean of glucose, body mass index, age, pregnancies, insulin, and diabetes pedigree functionality, which means a relatively worse clinical state. Glucose was found to be the most influential variable in defining the outcome of diabetes, as compared to the other variables examined, and it was followed by body mass index, age, pregnancies and diabetes pedigree function. These results indicate that high blood glucose is the most significant indicator of diabetic condition, whereas obesity, age, reproductive history and family history are also significant in occurrence of diabetes. The outcome of diabetes is, however, not caused by one risk factor, but it can be seen as a reflection of the effect of the combination of the many patient-related characteristics. The results present a comprehensible insight into the key determinants that relate to diabetes outcome instead of emphasizing the predictive classification. This renders the results useful in clinical interpretation and future research on diabetes based on data. The analysis facilitates the relevance of early risk identification in the form of regular clinical and demographic evaluation by determining the most important indicators associated with the diabetic status.

References

1. Akturk, M. (2020). *Diabetes dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/mathchi/diabetes-data-set

2. Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine learning based diabetes classification and prediction for healthcare applications. *Journal of healthcare engineering*, *2021*(1), 9930985.
3. Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Computing and Applications*, *35*(22), 16157-16173.
4. Hasan, M., & Yasmin, F. (2025). Predicting diabetes using machine learning: A comparative study of classifiers. *arXiv preprint arXiv:2505.07036*.
5. Howlader, K. C., Satu, M. S., Awal, M. A., Islam, M. R., Islam, S. M. S., Quinn, J. M., & Moni, M. A. (2022). Machine learning models for classification and identification of significant attributes to detect type 2 diabetes. *Health information science and systems*, *10*(1), 2.
6. Jaiswal, V., Negi, A., & Pal, T. (2021). A review on current advances in machine learning based diabetes prediction. *Primary Care Diabetes*, *15*(3), 435-443.
7. Kadam, P., Godse, S., & Mahalle, P. (2025, October). Machine learning in the early detection and prediction of diabetes: A systematic review. In *AIP Conference Proceedings* (Vol. 3325, No. 1, p. 070036). AIP Publishing LLC.
8. Kiran, M., Xie, Y., Anjum, N., Ball, G., Pierscionek, B., & Russell, D. (2025). Machine learning and artificial intelligence in type 2 diabetes prediction: a comprehensive 33-year bibliometric and literature analysis. *Frontiers in digital health*, *7*, 1557467.
9. Looker, H. C., Chang, D. C., Baier, L. J., Hanson, R. L., & Nelson, R. G. (2023). Diagnostic criteria and etiopathogenesis of type 2 diabetes and its complications: lessons from the Pima Indians. *La Presse Médicale*, *52*(1), 104176.
10. Maniruzzaman, M., Rahman, M. J., Ahammed, B., & Abedin, M. M. (2020). Classification and prediction of diabetes disease using machine learning paradigm. *Health information science and systems*, *8*(1), 7.
11. Narasimharao, M., Swain, B., Priyadarshi, R., Nayak, P. P., & Bhuyan, S. (2025). A survey of machine learning techniques for diabetes prediction: current trends and future directions. *Archives of Computational Methods in Engineering*, 1-36.
12. Naz, H., & Ahuja, S. (2020). Deep learning approach for diabetes prediction using PIMA Indian dataset. *Journal of Diabetes & Metabolic Disorders*, *19*(1), 391-403.
13. Netayawijit, P., Chansanam, W., & Sorn-In, K. (2025, October). Interpretable Machine Learning Framework for Diabetes Prediction: Integrating SMOTE Balancing with SHAP Explainability for Clinical Decision Support. In *Healthcare* (Vol. 13, No. 20, p. 2588). MDPI.
14. Regmi, D., Al-Shamsi, S., Govender, R. D., & Al Kaabi, J. (2020). Incidence and risk factors of type 2 diabetes mellitus in an overweight and obese population: a long-term retrospective cohort study from a Gulf state. *BMJ open*, *10*(7), e035813.
15. Shams, M. Y., Tarek, Z., & Elshewey, A. M. (2025). A novel RFE-GRU model for diabetes classification using PIMA Indian dataset. *Scientific reports*, *15*(1), 982.
16. Ye, C., Wang, Y., Kong, L., Zhao, Z., Li, M., Xu, Y., ... & Wang, T. (2022). Comprehensive risk profiles of family history and lifestyle and metabolic risk factors in relation to diabetes: A prospective cohort study. *Journal of Diabetes*, *14*(6), 414-424.