International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Predicting Diabetes Mellitus Using Logistic Regression On Clinical And Demographic Data

[1]Ayobami Ibukun Okegbade, [2]Abiola Opeyemi Egbewumi
[1,2]Ladoke Akintola University of Technology, Ogbomosho, Nigeria
[1]aiokegbade@lautech.edu.ng, [2]egbewumiabiola1@gmail.com

Correspodence Email: aiokegbade@lautech.edu.ng

Abstract: *Diabetes mellitus is a long term metabolic condition characterized by elevated blood glucose levels due to impaired insulin production, insulin action or both. The global rise in diabetes prevalence presents a major public health concern. This study utilized a dataset of 768 Diabetes Cases (No Diabetes (Type 1) = 500 Cases while Yes (Type 2) Diabetes Cases = 268 Cases) Obtained from Kaggle.com to explore the clinical and demographic predictors of diabetes mellitus using logistic regression analysis. Results revealed that glucose concentration, body mass index (BMI), diabetes pedigree function and number of pregnancies were the most significant predictors of diabetes. Elevated glucose emerged as he strongest predictor while obesity and hereditary risk substantially increased the likelihood of diabetes. The model demonstrated a good fit and moderate explanatory power, correctly classifying 78.3% of cases, though it performed better at identifying non-diabetic than diabetic individuals. Receiver Operating Characteristic (ROC) analysis confirmed glucose as the most discriminative variable followed by BMI and age whereas insulin, skin thickness and blood pressure contributed minimally. These findings reinforce the multifactorial etiology of diabetes emphasizing the combined influence of clinical, genetic and demographic factors in disease prediction. Clinically, the results suggest that regular monitoring of glucose levels, BMI and family history could enhance early detection and preventive management of diabetes in at risk populations.*

Keywords: *Diabetes Mellitus, Logistic Regression, Clinical and Demographic Predictors*

INTRODUCTION

Diabetes mellitus is a long-term metabolic condition marked by elevated blood glucose levels due to impaired insulin production, action, or both. The global rise in diabetes prevalence presents a major public health issue. This review explores the patterns and global implications of diabetes, discussing prevalence trends, contributing factors, associated health risks, and the economic burden it imposes. Diabetes mellitus is a long-term metabolic condition marked by elevated blood glucose levels, resulting from impaired insulin production, insulin action, or both. Its prevalence has grown significantly worldwide. According to the International Diabetes Federation, around 537 million adults were affected in 2021, a number projected to escalate to 783

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

million by 2045. This review examines current strategies for managing and treating diabetes, emphasizing lifestyle changes, medical therapies, and novel treatment technologies. Despite advances in medical diagnosis and management early identification o individuals at risk for diabetes remains a major concern in low and middle income countries particularly due to limited access to advanced diagnostic tools and inconsistent clinical record systems (Owolabi et al., 2020; Bello et al., 2022). Existing studies have explored numerous clinical, lifestyle and demographic determinants of diabetes, leading to the development of predictive models aimed at identifying high risk individuals. Logistic regression remains a preferred method for predicting diabetes mellitus due to its interpretability, ease of implementation and ability to handle binary outcomes efficiently. Unlike random forests or neutral networks, logistic regression provides clear insights into the direction and magnitude of associations between predictors (such as age, BMI, and family history) and diabetes risk. This particularly valuable for clinical decision making where understanding the influence of individual risk factors is as crucial as accurate prediction.

Literature Review

Diabetes mellitus is a long-term metabolic condition marked by elevated blood glucose levels due to impaired insulin production, action, or both. The global rise in diabetes prevalence presents a major public health issue. This review explores the patterns and global implications of diabetes, discussing prevalence trends, contributing factors, associated health risks, and the economic burden it imposes. Extensive research has shed light on diabetes' global spread, revealing concerning increases in both prevalence and incidence. The International Diabetes Federation (IDF) reported that in 2021, around 537 million adults (aged 20–79) were living with diabetes. Projections suggest this number could reach 643 million by 2030 and 783 million by 2045 (IDF, 2021). Prevalence and Incidence: Diabetes rates differ by region, with the Western Pacific and Southeast Asia recording the highest figures. These disparities are influenced by factors such as lifestyle habits, genetic background, and the quality of healthcare access (Zheng et al., 2018). Types of Diabetes: Type 2 diabetes (T2D) is the most common form, accounting for 90–95% of all cases. Type 1 diabetes (T1D), although less common, is becoming more frequent, especially among young individuals (Patterson et al., 2019). Risk Factors: Major risk contributors include obesity, sedentary lifestyles, poor dietary habits, and hereditary influences. The global obesity epidemic, particularly in low- and middle-income nations, is a major driver behind the diabetes crisis (Ng et al., 2014).

The effects of diabetes are far-reaching, impacting individuals, health systems, and economies worldwide. Health Impacts: Diabetes can lead to numerous complications, such as heart disease, nerve damage, kidney failure, and vision problems. These outcomes greatly elevate the risk of illness and death. The World Health Organization (WHO) noted that diabetes was responsible for 1.5 million deaths in 2019 and plays a major role in the global disease burden (Saeedi et al., 2019). Economic Consequences: The financial toll of diabetes includes both direct healthcare spending and indirect costs, such as productivity losses. In 2019, global healthcare spending on diabetes reached an estimated USD 760 billion (IDF, 2019), underlining the pressing need for better prevention and treatment strategies. Healthcare System Strain: The growing diabetes population strains healthcare systems, especially in under-resourced regions with limited access to essential treatments and medications. Robust, comprehensive care strategies are essential to ease this burden (Gonzalez et al., 2020). The current trends in diabetes incidence and its widespread consequences highlight the urgent need for coordinated public health action. Prevention, early diagnosis, and efficient management must be prioritized. Collaboration between governments, health professionals, and local communities is key to tackling the diabetes epidemic. Future research should aim to develop innovative care approaches, raise public awareness, and encourage healthier lifestyle choices to counter the rise of this chronic condition.

Type 2 diabetes mellitus (T2D) is a multifaceted metabolic condition marked by persistent high blood sugar levels, primarily caused by insulin resistance and inadequate insulin production. Its global incidence continues to surge, posing substantial public health challenges. Gaining insight into how genetic makeup and environmental exposures contribute to this condition is crucial for crafting targeted prevention and treatment strategies. This review consolidates current scientific findings on the genetic and environmental contributors to diabetes onset and progression.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Genetic susceptibility is a significant component in the development of diabetes, particularly T2D. Multiple lines of research have pinpointed specific gene variants that increase an individual's likelihood of developing the condition. Genome-Wide Association Studies (GWAS): Through GWAS, researchers have identified more than 400 genomic regions associated with T2D. These include genes that influence insulin production, glucose regulation, and inflammatory pathways (Mahajan et al., 2018). Genes such as TCF7L2, FTO, and KCNJ11 have shown strong and consistent associations with higher diabetes risk. Family and Twin Studies: Hereditary factors are underscored by studies involving families and twins, which suggest that genetics may account for 30–70% of the risk for developing T2D (Lyssenko et al., 2009). The likelihood of developing diabetes is notably higher among individuals with diabetic first-degree relatives. Mendelian Randomization: This method helps clarify the causal links between genetic markers and disease. For instance, variants in the HNF1A and GCK genes have been shown to directly influence insulin secretion and glucose regulation (Chatterjee et al., 2016).

Though genetics set the foundation, environmental elements play an equally pivotal role in diabetes development. These external influences can trigger or exacerbate genetic predispositions. Lifestyle Choices: Lack of physical activity, unhealthy eating patterns, and obesity are key modifiable risk factors for T2D. Diets rich in sugar and saturated fats contribute to weight gain and decreased insulin sensitivity, while regular exercise enhances glucose utilization and metabolic function (Hu, 2003). Socioeconomic Influences: Social determinants such as income, education level, and healthcare access significantly impact diabetes prevalence. Populations with lower socioeconomic status are more prone to diabetes, often due to limited access to nutritious foods and preventive care services (Marmot et al., 2008). Exposure to Pollutants: There is growing evidence that environmental contaminants particularly endocrine-disrupting chemicals (EDCs) and airborne pollutants may interfere with insulin function and glucose metabolism. These toxins can alter hormonal pathways, raising the risk of metabolic diseases (Rudolph et al., 2018). Psychological Stress and Mental Health: Chronic psychological stress and mental health disorders such as anxiety and depression are linked to an elevated risk of diabetes. Persistent stress may

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com
disrupt hormonal balance and encourage behaviors like overeating and physical inactivity, which heighten insulin resistance (Gonzalez et al., 2016).

The interplay between genetic makeup and environmental exposures is intricate and mutually influential. Environmental factors can lead to epigenetic changes chemical modifications that affect gene expression without altering DNA sequences. For example, maternal health conditions like obesity or gestational diabetes can induce epigenetic alterations in the developing fetus, thereby increasing the child's future diabetes risk (Szyf, 2015). With diabetes rates continuing to rise, understanding its multifactorial origins is more critical than ever. Both genetic factors and environmental exposures contribute substantially to the development of T2D. Future research should prioritize uncovering the dynamic relationships between genes and environment, which could pave the way for personalized prevention strategies. Public health initiatives must consider these dual influences to effectively reduce the global diabetes burden.

Diabetes mellitus is a long-term metabolic condition marked by elevated blood glucose levels, resulting from impaired insulin production, insulin action, or both. Its prevalence has grown significantly worldwide. According to the International Diabetes Federation, around 537 million adults were affected in 2021, a number projected to escalate to 783 million by 2045. This review examines current strategies for managing and treating diabetes, emphasizing lifestyle changes, medical therapies, and novel treatment technologies. Dietary management plays a vital role in diabetes control. Research supports the use of low-carbohydrate diets in enhancing blood sugar regulation and promoting weight loss (Tay et al., 2018). Additionally, adherence to a Mediterranean-style diet featuring whole grains, fruits, vegetables, and healthy fats has been linked to improved metabolic health and reduced diabetes risk (Schwingshackl & Hoffmann, 2014). Incorporating regular physical activity is essential for individuals living with diabetes. Evidence from Colberg et al. (2016) demonstrates that both aerobic and resistance exercises improve blood glucose control, cardiovascular function, and general well-being in diabetic patients. Psychological and behavioral approaches, such as blood glucose self-monitoring and motivational counseling, are effective in encouraging lasting lifestyle changes (Fisher et al., 2016). Structured

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com
education programs also help patients build confidence and skills to manage their condition more effectively.

For those with Type 1 diabetes and some with advanced Type 2 diabetes insulin therapy is indispensable. Innovations in insulin types, including ultra-fast-acting and extended-release versions, have enhanced treatment flexibility and glucose control (Buse et al., 2020).

Metformin remains the preferred initial treatment for Type 2 diabetes, helping to lower HbA1c and support weight reduction (Inzucchi et al., 2015). Other drug categories such as sulfonylureas, DPP-4 inhibitors, and SGLT2 inhibitors offer additional treatment options based on individual patient profiles (American Diabetes Association, 2021).

These injectable agents promote insulin release and contribute to weight loss, making them particularly beneficial for Type 2 diabetes management (Davies et al., 2018). Continuous Glucose Monitoring (CGM) technology provides real-time blood sugar data, enabling timely and precise therapy adjustments. Studies have shown that CGM use improves glycemic control and minimizes episodes of hypoglycemia (Bergenstal et al., 2018). Closed-loop systems that automatically adjust insulin delivery based on continuous glucose data are a breakthrough in diabetes care. These systems ease the burden of daily management while improving glucose stability (Kovacs et al., 2021). Advances in gene editing and regenerative medicine offer the potential to restore insulin production. Current research is exploring techniques to regenerate or replace pancreatic beta cells, though these therapies remain largely in experimental stages (Gonzalez et al., 2019). Effectively managing diabetes requires a comprehensive, multifactorial approach. Combining lifestyle changes with pharmacological treatments and technological advancements offers the best outcomes for individuals with diabetes. Ongoing research is key to refining these methods and tailoring them to individual needs. As diabetes continues to pose a global health threat, prevention and public education must remain top priorities in addressing its widespread impact.

Diabetes mellitus is a long-term metabolic condition marked by elevated blood glucose levels. As diabetes becomes increasingly common worldwide, a variety of innovative technologies have emerged to support its effective management. This review examines major technological breakthroughs in diabetes care, with an emphasis on glucose monitoring devices, insulin administration systems, mobile health (mHealth) tools, and remote healthcare (telemedicine).

Frequent monitoring of blood sugar is vital in managing diabetes effectively. Traditional fingerstick methods are often inconvenient and uncomfortable. Recent progress in continuous glucose monitoring (CGM) has transformed self-monitoring practices. Thabit et al. (2018) report that CGM systems provide continuous, real-time glucose readings, empowering users to make timely decisions about diet and insulin dosage. The integration of CGM data with mobile apps has further boosted user interaction and treatment adherence (Bergenstal et al., 2018).

Insulin remains a critical treatment, especially for individuals with Type 1 diabetes. Conventional injections may result in inconsistent absorption and pose a burden to patients. The advancement of insulin pumps and automated insulin delivery systems has enhanced diabetes care. According to Kovatchev et al. (2017), hybrid closed-loop systems also known as artificial pancreas devices regulate insulin delivery automatically based on CGM inputs. These systems have proven to lower HbA1c levels and improve glycemic stability, while reducing the risk of low blood sugar events (Bode et al., 2016).

The widespread use of smartphones has facilitated the growth of mobile health apps tailored for diabetes self-care. These apps offer features such as glucose tracking, meal planning, medication alerts, and more. Goyal et al. (2020) found that mobile-based interventions significantly enhance blood sugar control and patient involvement in their treatment. Apps with built-in social networks or support communities tend to yield better results, encouraging users to stay motivated and engaged (Gonzalez et al., 2018).

Telemedicine has become a key player in healthcare delivery, especially during the COVID-19 pandemic. It allows people with diabetes to receive medical attention remotely, reducing the need for in-person consultations. Sinha et al. (2021) concluded that telemedicine services effectively improve blood glucose control and patient satisfaction. Real-time communication with healthcare providers helps tailor treatment recommendations and maintain continuity of care, particularly during emergencies.

International Journal of
Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

While these technologies offer promising solutions, several challenges must be addressed. Concerns about data security, technological literacy, and access disparities remain significant. The "digital divide" may limit the reach of such innovations, particularly in low-resource settings. Gunter et al. (2019) emphasize the need for user-friendly platforms and thorough patient training to ensure these tools are used effectively. Additionally, the long-term health implications of relying heavily on technology are still being investigated. Technological advancements are reshaping how diabetes is managed, with tools such as CGM systems, advanced insulin pumps, smartphone apps, and virtual healthcare services enhancing patient care. However, maximizing the potential of these innovations depends on overcoming implementation barriers. Future research should prioritize improving user accessibility, safeguarding personal data, and assessing the sustained impact of digital tools on diabetes outcomes.

Diabetes is a long-term metabolic disorder marked by elevated blood sugar levels resulting from either insufficient insulin production or the body's inability to use insulin effectively. With its growing global prevalence, diabetes presents major public health concerns, including increased disease burden, mortality, and healthcare expenditure. Effective public health actions and policy frameworks are essential for prevention, control, and education around diabetes. The worldwide burden of diabetes is increasing at an alarming rate. According to the International Diabetes Federation (IDF), around 537 million adults were living with diabetes in 2021 a number projected to climb to 783 million by 2045. Type 2 diabetes dominates, largely influenced by lifestyle-related risk factors such as obesity, physical inactivity, and poor dietary habits. Lifestyle-Based Programs: Programs like the Diabetes Prevention Program (DPP) have shown that behavioral modifications especially weight loss and physical activity can lower the incidence of type 2 diabetes by as much as 58% in high-risk populations (Ackermann et al., 2008).

Community-focused initiatives have proven effective across various groups (Patterson et al., 2015). Early Detection and Screening: Targeted screening for at-risk populations helps identify undiagnosed diabetes cases, allowing for earlier and more effective treatment (Saeedi et al., 2019). Diabetes Self-Management Education (DSME): Educating individuals on managing their

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com
condition enhances their confidence and improves glucose control (Powers et al., 2015). Incorporating ongoing support and digital tools like mobile apps has further improved outcomes (Pal et al., 2020). Nutrition-Focused Policies: Public health regulations such as taxes on sugary drinks and restrictions on unhealthy food marketing can reduce the consumption of diabetes-linked foods (Pratt et al., 2019). U.S. initiatives like the Healthy Food Financing Initiative aim to increase access to nutritious food in underserved communities (Bower et al., 2017). Community-based programs that encourage exercise such as public fitness classes, walking clubs, and accessible parks have been linked to increased physical activity and lower diabetes risk (Morrato et al., 2007).

Addressing mental health is crucial in diabetes care. Integrating psychological support into treatment plans has shown to improve treatment adherence and overall well-being (Nouwen et al., 2016). The policy level interventions of healthcare system improvements universal access to care: Making essential diabetes services and medications affordable and accessible is key to equitable healthcare delivery (Burgess et al., 2018). Chronic Care Models: Adopting the Chronic Care Model improves healthcare delivery by fostering collaboration between providers and patients and improving care coordination (Wagner et al., 2001). Public campaigns educating people about diabetes prevention, risk factors, and healthy behaviors are instrumental. Programs like the U.S. National Diabetes Prevention Program have successfully promoted awareness and lifestyle changes. Building robust health data systems allows policymakers to monitor diabetes trends, guide resource allocation, and design targeted interventions (IDF, 2021). Legislation aimed at improving healthcare access, regulating food and beverage marketing, and promoting physical activity-friendly environments can significantly reduce diabetes risk over time (Bleich et al., 2018).

Effectively addressing the diabetes crisis requires a combination of public health initiatives and policy reforms. Integrating individual-level actions with systemic strategies backed by stakeholder collaboration is critical. Future interventions must prioritize equity, embrace technological innovations, and consider the broader social determinants of health to combat the diabetes epidemic successfully. Diabetes has become a global public health issue, with its

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com
prevalence surging, especially in low- and middle-income nations. This trend aggravates existing financial and healthcare challenges in these areas. The condition disproportionately affects older populations and certain ethnic groups, resulting in unequal social and economic outcomes. Managing diabetes incurs significant costs, including spending on medications, insulin, glucose monitoring tools, and hospital care. These expenses often impose a financial strain on patients and their families. On a broader scale, the condition escalates national healthcare costs, driving up insurance premiums and public health expenditures. Indirect effects include productivity loss due to illness-related absenteeism and reduced performance at work (presenteeism). People with diabetes are more prone to job-related disability and diminished work capacity.

Complications such as cardiovascular diseases and nerve damage amplify these costs, contributing to higher unemployment rates and economic instability among affected groups. Diabetes adversely affects individuals' mental and emotional well-being, often leading to depression and anxiety, which hinder social and professional engagement. The caregiving burden also extends to families, increasing their emotional stress and financial obligations, which may lead to overall household socioeconomic decline. Socioeconomic background significantly influences access to diabetes treatment and education. Those in poorer communities often face obstacles such as inadequate insurance, poor access to healthy foods, and limited health literacy. These disparities result in worse health outcomes and higher complication rates in economically disadvantaged populations. The literature highlights that targeted public health strategies such as local education and prevention initiatives can help reduce the socioeconomic toll of diabetes. Cost-effectiveness studies show that preventive healthcare interventions can generate substantial long-term savings for both individuals and health systems. Diabetes imposes a deep and complex socioeconomic burden on individuals and society. Ongoing research, equitable healthcare access, and well-crafted public policies are essential to reduce this burden. By tackling the socioeconomic dimensions of diabetes, stakeholders can improve health outcomes and reduce long-term costs.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

METHOD

The data used in this study were obtained from the publicly available Kaggle.com Diabetes originally derived from National Institutes of Diabetes and Digestive and Kidney Diseases (NIDDK). The dataset comprises 768 patient records, each containing clinical and demographic attributes related to diabetes diagnosis. The dependent variable indicates whether or not a patient is diagnosed with diabetes based on specific diagnostic criteria. To ensure data quality and model reliability, several data preprocessing phase involved data cleaning, handling of missing values, normalization and outlier detection. Although the dataset contained no explicit missing values, certain variables (Glucose, Blood Pressure, Skin, Thickness, Insulin, and BMI) contained zero entries that are physiologically implausible and thus treated by extreme values and provides a robust measure of central tendency. All continuous variables were normalized using min-max scaling to bring their values within a standard range of 0 and 1. This process ensured that features with larger numerical scale did not disproportionally influence the model coefficients. The study employed a binary. Data Analysis was conducted using Microsoft Excel, SPSS 25 and Gretl

RESULT AND DISCUSSION

Result

Descriptive Statistics on Factors Associated to Diabetes Mellitus and Outcome Of Diabetes Mellitus

	N	Mean	Std. Deviation
Pregnancies	768	3.85	3.370
Glucose	768	120.89	31.973
Blood Pressure	768	69.11	19.356
Skin Thickness	768	20.54	15.952
Insulin	768	79.80	115.244
BMI	768	31.99	7.884
Diabetes Pedigree Function	768	.47	.331
Age	768	33.24	11.760
Valid N (listwise)	768		

Table 1: *Descriptive Statistics on factors associated to Diabetes*

Table 1 revealed the descriptive statistics of the clinical and demographic variables for the 768 patients included in the study. The mean number of pregnancies was 3.85 with standard

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

deviation of 3.37, indicating that, on average, female participants had approximately four pregnancies, with considerable variability across the sample. The mean fasting blood glucose level was 120.89mg/dl with standard deviation of 31.97, which is above the normal threshold of <100mg/dl, suggesting that a substantial proportion of the participants may have impaired glucose regulation or diabetes. The average diastolic blood pressure was 69.11mmHg with standard deviation of 19.36, which falls within the normal range; however, the relatively high standard deviation indicates notable variation, with some patients exhibiting either hypotension or hypertension. The mean skinfold thickness was 20.54 mm with standard deviation of 15.95, reflecting differences in subcutaneous fat levels among participants. Insulin levels exhibited a mean of 79.80µU/ml with standard deviation of 115.24, with the large variability suggesting the presence of extreme values, possibly indicating significant differences in insulin resistance status within the cohort. Body Mass Index (BMI) had a mean value of 31.99kg/m^2 with standard deviation of 7.88, placing the average participant the obese category, which is a recognized risk factor for diabetes pedigree function (DPF), which estimates the hereditary risk of developing diabetes, had a mean of 0.47 with standard deviation of 0.33, indicating moderate familial risk across the sample. The mean age of the participants was 33.24 years with standard deviation of 11.76, suggesting that the study population consisted predominantly of young to middle aged adults. Overall, the descriptive statistics highlights that several key risk factors for diabetes mellitus, including elevated fasting glucose, obesity and high insulin variability are prevalent in the study population. These variables will be further examined in the logistic regression analysis to determine their predictive significance in the occurrence of diabetes mellitus.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	No	500	65.1	65.1	65.1
	Yes	268	34.9	34.9	100.0
	Total	768	100.0	100.0	

Table 2: *Descriptive Statistics on the Outcome of Diabetes*

Table 2 revealed the distribution of diabetes mellitus types. The analysis revealed that among the 768 patients in the study, 500 individuals 65.1% were diagnosed with No 1 diabetes mellitus

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

while 268 individuals 34.9% were diagnosed with diabetes mellitus. This indicates that type 1 diabetes was nearly twice as prevalent as type 2 in the sampled population. The cumulative percentage shows that all case were accounted for, with no missing data.

Correlation Matrix to Examines the Linear Relationships among the Study Variables

		Constant	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Predigree Function	Age
Step 1	Constant	1.000	-.063	-.522	-.182	.005	.166	-.584	-.215	-.227
	Pregnancies	-.063	1.000	.104	-.075	.004	.033	.068	.069	-.486
	Glucose	-.522	.104	1.000	-.109	.156	-.366	-.025	.058	-.180
	Blood Pressure	-.182	-.075	-.109	1.000	-.175	.048	-.187	.004	-.185
	Skin Thickness	.005	.004	.156	-.175	1.000	-.429	-.286	-.089	.096
	Insulin	.166	.033	-.366	.048	-.429	1.000	.009	-.077	.058
	BMI	-.584	.068	-.025	-.187	-.286	.009	1.000	.010	.082
	Diabetes Predigree Function	-.215	.069	.058	.004	-.089	-.077	.010	1.000	-.041
	Age	-.227	-.486	-.180	-.185	.096	.058	.082	-.041	1.000

Table 3. *Correlation Matrix to Examines the Linear Relationships among the Study Variables*

Table 3 revealed the correlation matrix revealed generally weak to moderate associations among the predictor variables, indicating no evidence of severe multicollinearity that could compromise the logistic regression model. The strongest relationships observed were a moderate negative correlation between pregnancies and age (r = -0.486), suggesting that higher pregnancy counts were more common among younger women in the dataset, and a moderate negative correlation between skin thickness and insulin values among individuals with higher skinfold thickness. Glucose was negatively correlated with possible insulin deficiency in diabetic patients and showed only weak positive correlations with skin thickness and BMI, Blood pressure, BMI, diabetes pedigree function and age demonstrated largely weak or negligible associations with other variables. Overall, the pattern of correlations suggests that while a few moderate relationships exist, most predictors are relatively independent supporting their simultaneous inclusion in the regression model.

Omnibus Test of Model Coefficients

Omnibus Tests of Model Coefficients			
	Chi-square	df	Sig.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Step 1	Step	270.039	8	.000
	Block	270.039	8	.000
	Model	270.039	8	.000

Table 4: *Omnibus Test of Model Coefficients*

Table 4 revealed the omnibus test of model coefficients which checks whether your logistic regression model with predictors fits significantly better than a model without any predictors (i.e. the null model). The table shows Chi-Square of 270.039 which indicate how much better the model fits compared to the null model. 8 Degree of Freedom equal to the number of predictors in the model and the p-values 0.000 shows a highly significant ($p < 0.001$) indicating the model provides a significantly better fit than a model without predictors. All the three values Chi-Square values, Degree of Freedom and P-Values for the Step, Block and Model are the same indicating the 8 predictors are significantly improved model. Overall, the Omnibus test of model coefficients indicates that logistic regression model with pregnancies, glucose, blood pressure, skin thickness, insulin, BMI, Diabetes Pedigree Function and Age significantly predicts the likelihood of diabetes.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	723.445[a]	.296	.408

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Table 5. *Model Summary*

Table 5 reveals that -2LL is 723.445 which reveals the final value after estimation and the fact that the change stopped at iteration 5 means the model converged normally, Cox & Snell R^2 is 0.296 which indicate that the pseudo R^2 value (not directly comparable to R^2 in linear regression) it suggests that about 29.5% of the variation in the dependent variable can be explained by the predictors in your logistic regression model while Nagelkerke R^2 is 0.408 indicating that about 40.8% of the variance in the outcome variable is accounted for by the model, this is considered a moderate to strong effect size in medical sciences. Overall, the logistic regression model explains roughly 30-41% of the variability in diabetes status which indicates that predictors collectively have a substantial influence on the outcome. The model converged well and is likely providing stable parameter estimates.

Hosmer and Lemeshow Test of Diabetes Mellitus

Step	Chi-square	df	Sig.
1	8.323	8	.403

Table 6a. *Hosmer and Lemeshow Test*

		Outcome = No		Outcome = Yes		
		Observed	Expected	Observed	Expected	Total
Step 1	1	74	74.336	3	2.664	77
	2	73	70.819	4	6.181	77
	3	72	67.854	5	9.146	77
	4	61	63.898	16	13.102	77
	5	59	58.824	18	18.176	77
	6	53	53.046	24	23.954	77
	7	40	45.910	37	31.090	77
	8	40	34.479	37	42.521	77
	9	18	21.811	59	55.189	77
	10	10	9.021	65	65.979	75

Table 6b. *Contingency Table for Hosmer and Lemeshow Test*

Table 6 revealed Hosmer and Lemeshow Test goodness of fits test grouped participants into deciles based on predicted probabilities of diabetes. Observed and expected frequencies within each group were closely aligned, with no large deviations detected. The Chi-square statistic was 8.323 (df = 8, p = 0.403), indicating no significant lack of fit. This suggests that the logistic regression model adequately fits the data, with predicted probabilities closely matching the observed outcome.

Classification Table (Confusion Matrix) to assess the predictive accuracy of the logistic regression model in identifying diabetes outcome

			Predicted		
			Outcome		
			No	Yes	Percentage Correct
Step 1	Outcome	No	445	55	89.0
		Yes	112	156	58.2
	Overall Percentage				78.3

a. The cut value is .500

Table 7. *Classification Table to assess the predictive accuracy of the logistic regression model in identifying diabetes outcome*

Table 7 revealed the classification table was used to evaluate the predictive performance of the logistic regression model in identifying diabetes outcomes at the probability of 0.500. The model correctly classified 445 non-diabetes patients as No and misclassified 55 as diabetic resulting in an accuracy of 89.0% for the non-diabetic group. Among diabetic patients 156 were correctly predicted 601 out of 768 cases giving an overall classification accuracy of 78.3%. For instance, a 2024 study on type 2 diabetes using logistic regression reported an accuracy of 78.26%, closely matching your result. Another study (2025) on diabetes risk modelling found logistic regression achieved approximately 74% accuracy when compared with other methods. These results indicate that the model performs strongly in predicting no-diabetic cases but is less effective in detecting diabetic cases. This imbalance suggests that while the predictors are reliable for ruling out diabetes, they are less accurate for confirming its presence which may be due to class imbalance or weaker predictor associations with diabetes status. When compared with recent literature, these findings align with observed trends.

Variables in the Equation (Regression Coefficients)

Variables in the Equation

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1[a]	Pregnancies	.123	.032	14.747	1	.000	1.131	1.062	1.204
	Glucose	.035	.004	89.897	1	.000	1.036	1.028	1.043
	Blood Pressure	-.013	.005	6.454	1	.011	.987	.977	.997
	Skin Thickness	.001	.007	.008	1	.929	1.001	.987	1.014
	Insulin	-.001	.001	1.749	1	.186	.999	.997	1.001
	BMI	.090	.015	35.347	1	.000	1.094	1.062	1.127
	Diabetes Predigree Function	.945	.299	9.983	1	.002	2.573	1.432	4.625
	Age	.015	.009	2.537	1	.111	1.015	.997	1.034
	Constant	-8.405	.717	137.546	1	.000	.000		

a. Variable(s) entered on step 1: Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, DiabetesPredigreeFunction, Age.

Table 8. *Variables in the Equation*

Table 8 revealed the logistic regression analysis identified several significant predictors of diabetes mellitus. Glucose concentration emerged as the strongest predictor with each 1 mg/dl increase associated with 3.6% higher likelihood of diabetes ($p < 0.001$). Body mass index (BMI)

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

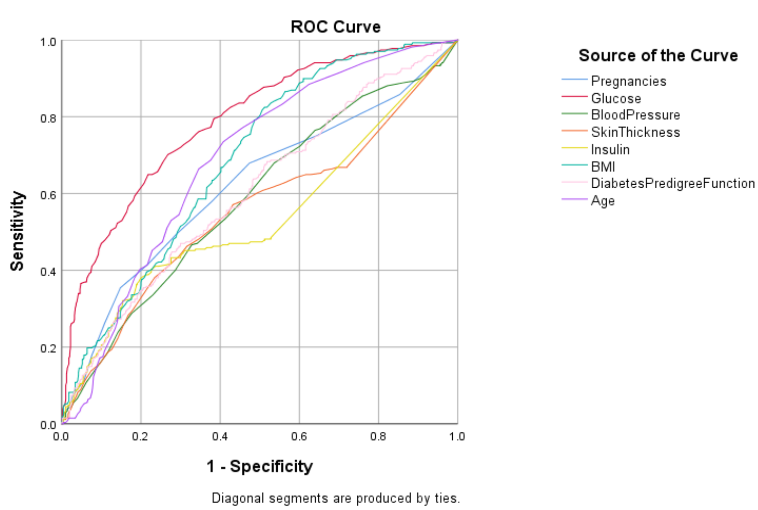
also showed a strong effect, where each unit increase corresponded to a 9.4% rise in diabetes odds ($p < 0.001$). Similarly, a higher diabetes pedigree function reflecting genetic predisposition more than doubled the odds of developing diabetes (Exp(B) = 2.573, $p = 0.002$) while the number pregnancies significantly increased risk by 13.1% per additional pregnancy ($p < 0.001$). Conversely, skin thickness, insulin levels, and age did not exhibit significant associations after adjustment for other factors. Overall, the findings suggest that pregnancies are the most influential risk factors for diabetes, whereas blood pressure plays a modest protective role and other clinical variables were not significant predictors.

ROC Curve Analysis of Diabetes Predictors

Outcome	Valid N (Listwise)	Percentage
Positive (Diabetes = Yes)	268	34.9%
Negative (Diabetes = No)	500	65.1%
Total	**768**	**100%**

Table 9. *ROC Curve of Diabetes Predictors*

The analysis included a total of 768 valid cases after listwise deletion of missing data, out of these, 268 cases (34.9%) were classified as positive for the outcome (i.e. diagnosed with diabetes mellitus) while 500 cases (65.1%) were classified as negative (no diabetes diagnosis). This note indicates that larger values of the test variables correspond to stronger evidence for a positive actual state meaning higher predictor scores are associated with the likelihood of having diabetes.



International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

Figure 1: ROC Curve to evaluates how well each predictor carriable distinguishes between patients with diabetes (Positive = Yes) and those without diabetes (Negative =No).

Figure 1 revealed the ROC (Receiver Operating Characteristic) curve to evaluates how well each predictor variable distinguishes between patients with diabetes (Positive = Yes) and those without diabetes (Negative = No). Sensitivity (y-axis) = True Positive Rate while Specificity (x-axis) = False Positive Rate. The Glucose Curve is farthest from the diagonal line meaning it is the most reliable single variable in predicting diabetes. Age and Pregnancies also show some predictive value but weaker. Blood Pressure, Skin Thickness and Insulin have curves close to the diagonal, meaning they are not strong individual predictors in this dataset. The diagonal line represents random guessing (AUC = 0.05). This indicate that glucose should be considered a primary predictor variable of diabetes mellitus. Age and number of pregnancies may still be useful in combination with other predictors but on their own they are weaker. The low AUC fo Insulin and blood pressure suggests they may not add much predictive value unless they interact with other variables.

Arae Under the Curve

Test Result Variable(s)	Area
Pregnancies	.620
Glucose	.788
Blood Pressure	.586
Skin Thickness	.554
Insulin	.538
BMI	.688
Diabetes Pedigree Function	.606
Age	.687

The test result variable(s): Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Table 10. *Area Under the Curve*

Table 10 revealed Area Under the Curve from the ROC analysis measures how well each variables cam discriminate between patients with diabetes (Positive actual state) and those without (negative actual state). The table reveals that Glucose 0.788 has good predictor among all variables, indicating strong ability to distinguish between a diabetic and no-diabetic patients. BMI

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com

0.688 shows fair predictor indicating moderately useful. Age 0.687 shows fair predictor similar usefulness to BMI. Pregnancies 0.620 shows fair predictors indicating a limited but noticeable discrimination ability. Diabetes Pedigree Function 0.606 shows fair predictor indicating borderline usefulness. Blood pressure 0.586 shows poor predictor. Skin Thickness 0.554 shows poor predictor indicating low individual predictive strength while Insulin 0.538 shows poor predictor indicating weakest among all variables.

Discussion of Findings

This study demonstrated that glucose concentration, BMI, diabetes pedigree function and number of pregnancies were the most significant predictors of diabetes mellitus. Elevated glucose emerged as the strongest predictor while obesity and hereditary risk also substantially increased the likelihood of diabetes. The model achieved good fit and moderate explanatory power, correctly classifying 78.3% of cases, though it performed better at identifying non-diabetes than diabetic patients. ROC analysis confirmed glucose as the most discriminative variable followed by BMI and age whereas insulin, skin thickness and blood pressure contributed minimally. These results reinforce the multifactorial etiology of diabetes, highlighting the interplay of clinical, genetic and demographic factors in disease prediction.

CONCLUSION

The study established that diabetes mellitus can be effectively predicted using clinical an demographic variables, with glucose concentration, BMI, diabetes pedigree function and number of pregnancies identified as key risk factors. Although this logistic regression model showed good accuracy in classifying non-diabetes cases, its lower sensitivity in detecting true diabetic patients suggests the need for more complex modelling approaches. Future research could explore the integration of additional lifestyle, socioeconomic and biochemical variables as well as the predictive performance and capture the nonlinear relationships underlying diabetes risk.

REFERENCE

American Diabetes Association. (2021). *Standards of medical care in diabetes—2021*. Diabetes Care, 44(Suppl. 1), S1–S232.

Bello, O. M., Adesina, K. T., & Akinyemi, O. O. (2022). Comparative analysis of machine learning algorithms for diabetes prediction using clinical datasets in Nigeria. *Nigerian Journal of Clinical Practice, 25*(6), 812–820.

Bergenstal, R. M., et al. (2018). Effectiveness of continuous glucose monitoring in patients with type 1 diabetes. *Diabetes Care.*

Bergenstal, R. M., et al. (2018). Recommendations for the use of continuous glucose monitoring in adults with diabetes.

Bode, B. W., et al. (2016). The impact of continuous glucose monitoring on glycemic control. *Journal of Diabetes Science and Technology*.

Buse, J. B., et al. (2020). Insulin treatment in type 2 diabetes: An updated review.

Chatterjee, A., Khaw, K. T., & Wareham, N. J. (2016). Genetic risk, lifestyle factors, and the risk of type 2 diabetes: A Mendelian randomization study. *Diabetes Care*, *39*(3), 487–495.

Colberg, S. R., et al. (2016). Physical activity/exercise and diabetes: A position statement of the American Diabetes Association. *Diabetes Care.*

Davies, M. J., et al. (2018). Efficacy of GLP-1 receptor agonists in type 2 diabetes: A systematic review and meta-analysis.

Fisher, L., et al. (2016). Behavioral strategies for diabetes management.

Gonzalez, A. L., et al. (2019). Gene therapy for diabetes: Recent advances and future prospects.

Gonzalez, J. S., et al. (2016). Stress, diabetes, and chronic illness. *Psychosomatic Medicine, 78*(7), 885–895.

Gonzalez, J. S., et al. (2018). Social support and diabetes management. *Diabetes Spectrum.*

Gonzalez, J. S., et al. (2020). Healthcare system challenges in diabetes management. *Diabetes Care*, *43*(3), 654–661.

International Journal of Interdisciplinary Research
ISSN(Online): 3090-2959
Vol 2 no 2 (2026): July 2026
https://journal.as-salafiyah.id/index.php/ijir/index
Email: ijireditor7@gmail.com
Goyal, S., et al. (2020). Mobile health interventions for diabetes management: A systematic review. *Diabetes Technology & Therapeutics.*

Gunter, M., et al. (2019). Challenges in the adoption of health technologies. *Journal of Health Communication.*

Hu, F. B. (2003). Resolved: There is a significant relationship between obesity and diabetes. *Diabetes Care, 26*(3), 858–863.

International Diabetes Federation. (2019). *IDF diabetes atlas* (9th ed.). IDF.

International Diabetes Federation. (2021). *IDF diabetes atlas* (10th ed.). IDF.

Inzucchi, S. E., et al. (2015). Management of hyperglycemia in type 2 diabetes, 2015: A patient-centered approach. *Diabetes Care.*

Kovacs, S., et al. (2021). The artificial pancreas: Current status of research and clinical use.

Kovatchev, B. P., et al. (2017). Closed-loop control of blood glucose. *Diabetes Care.*

Lyssenko, V., et al. (2009). Genetic variation in GIPR influences the risk of type 2 diabetes. *Diabetes, 58*(7), 1368–1374.

Mahajan, A., et al. (2018). Identification and impact of type 2 diabetes loci in 433,540 individuals. *Nature Genetics, 50*(11), 1590–1601.

Marmot, M. G., et al. (2008). *Closing the gap in a generation: Health equity through action on the social determinants of health.* World Health Organization.

Ng, M., et al. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet, 384*(9945), 766–781.

Owolabi, E. O., Onuaha, H., & Adewale, A. (2020). Challenges in diabetes prediction and prevention in sub-Saharan Africa: A systematic review. *African Journal of Diabetes Medicine, 28*(2), 28–34.

Patterson, C. C., et al. (2019). Global estimates of incidence of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice, 157*, 107840.

Rudolph, A., et al. (2018). Environmental pollutants and diabetes: A review of the literature. *Environmental Health Perspectives, 126*(3), 037001.

Saeedi, P., et al. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045. *Diabetes Research and Clinical Practice, 157*, 107843.

Schwingshackl, L., & Hoffmann, G. (2014). Adherence to a Mediterranean diet and risk of type 2 diabetes.

Sinha, R., et al. (2021). Telemedicine and diabetes management: A meta-analysis. *Diabetes Research and Clinical Practice.*

Szyf, M. (2015). The role of epigenetics in the developmental origins of health and disease. *Nature Reviews Genetics*, *16*(12), 849–861.

Tay, J., et al. (2018). Low-carbohydrate diets and cardiovascular disease: A systematic review.

Thabit, H., et al. (2018). Continuous glucose monitoring: A review. *Diabetes Technology & Therapeutics.*

Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global epidemiology of type 2 diabetes and its cardiovascular implications. *Nature Reviews Cardiology, 15*(2), 85–98.