◆ Research paper
DOI: 10.5281/zenodo.10791178

Novel Risk Factors for Diabetes: A Comprehensive Analysis for Enhanced Disease Diagnosis

Irfan Chaudhuri[1], Md Shah Alam[2], and Md. Kamrujjaman[3*]

[1]Department of Chemistry, Colby College, Maine, USA

[2]Department of Mathematics, University of Houston, Houston, Texas- 77204, USA

[3]Department of Mathematics, University of Dhaka, Dhaka-1000, Bangladesh

* Corresponding author Email: kamrujjaman@du.ac.bd

Abstract: Diabetes mellitus, Type II, is a prevalent chronic condition with significant health implications worldwide. For management and intervention to be successful, an accurate and prompt diagnosis is essential. The accessibility of several data resources offers the chance to investigate novel concepts and improve the accuracy of diabetes diagnosis. Using the extensive Diabetes Dataset from the Iraqi community, this research seeks to evaluate diabetes-related biomarkers. The dataset consists of medical records and laboratory findings from 1000 people who were either diagnosed with diabetes, did not have the disease, or were expected to develop it. Blood sugar level, age, gender, creatinine ratio (Cr), body mass index (BMI), urea, cholesterol (Chol), fasting lipid profile, and HbA1c are among the many factors that are taken into account. A proper diabetes diagnosis is crucial since it has an immediate influence on patient care and management tactics. With the use of the Diabetes Dataset, this project aims to create a more precise diagnostic framework that would aid healthcare professionals in making decisions that will enhance patient outcomes. This research intends to improve the present diagnostic techniques by revealing hidden patterns, correlations, and prediction signs within the dataset. The study offers insights into the association between diabetes and numerous variables through careful data analysis and data-driven decision-making. This work helps continuing efforts to improve patient care and treatment in this difficult area by addressing the urgent need for better diabetes diagnostic techniques.

Keywords: Diabetes, risk factor, data set, chronic disease.

Introduction

Diabetes mellitus type II is a common and complicated chronic condition that creates serious health issues for people all over the world. Accurate and prompt diagnosis is essential for efficient management and intervention due to its rising incidence and related consequences. The accessibility of vast data resources offers the chance to experiment with new ideas and improve the precision of diabetes diagnosis. In this paper, we will analyze biomarkers linked with diabetes by delving into the vast Diabetes Dataset and graphs, gathered from the Iraqi society in regards to type II diabetes.

A comprehensive collection of medical data and laboratory analyses from the Medical City Hospital and the Specialized Center for Endocrinology and Diabetes at Al-Kindy Teaching Hospital are included in the Ahlam Rashid-contributed Diabetes Dataset (Rashid, 2020). The dataset consists of data from 1000 patients who fall into one of three categories: diabetic, non-diabetic, or predicted diabetic. This dataset is a significant resource for examining the many facets of diabetes since it includes a wide variety of variables, such as blood sugar level, age, gender, creatinine ratio (Cr), body mass index (BMI), urea, cholesterol (Chol), fasting lipid profile, and HBA1C.

The significance of accurate diabetes diagnosis cannot be overstated, as it directly impacts patient care and management strategies. Misdiagnosis or a delay in diagnosis can result in ineffective treatment strategies, higher medical expenses, and significant problems. As a result, this study aims to use the Diabetes Dataset to create a more accurate diagnostic framework that will help healthcare practitioners make wise decisions and improve patient outcomes.

The paper aims to reveal hidden patterns, correlations, and prediction signs within the collection. To distinguish between the three classifications of people with diabetes—Diabetic, Non-Diabetic, and Predicted-Diabetic—accurately, we must first discover the distinguishing characteristics and build a strong model that can do so. By utilizing the power of thorough data analysis and data-driven decision-making, this research aims to improve the current diagnostic procedures.

The Diabetes Dataset has been thoroughly analyzed in this research, providing insightful information on the medical data related to diabetes. We want to address the urgent need for more accurate and precise diagnosis methods in the field of diabetes by examining this dataset. By sharing our research, we want to support ongoing efforts to improve diabetes diagnosis, which will subsequently result in improved patient care and management techniques in this difficult area. In this paper, we will study the relationship between Diabetes and several attributes in the dataset such as blood sugar level, creatinine ratio (Cr), body mass index (BMI), urea, cholesterol (Chol), fasting lipid profile, and HBA1C.

Dataset collection and representation

The Mendeley Diabetes types dataset where the dataset used by the algorithm is taken from. The laboratory of Medical City Hospital and (the Specialized Center for Endocrinology and Diabetes-Al-Kindy Teaching Hospital) obtained the data from Iraqi society. The database is filled with information that has been taken from the patient's file. Medical information and laboratory analysis make up the data. The database contains 844 diabetic individuals, 53 pre-diabetics, and 103 people without diabetes.

The dataset includes information on 1000 patients. It includes several diabetes-related characteristics, such as the number of patients, blood sugar level, age, gender, creatinine ratio (Cr), body mass index (BMI), urea, and cholesterol (Chol), among others. These characteristics should be carefully chosen since any irrelevant factor might cause the findings to be skewed. Data analysis must include the removal of outliers since they provide statistical findings that are unimportant and might contradict assumptions (Rajput & Khedgikar, 2022). A database is initially applied using a box plot to eliminate any outliers.

Body Mass Index (BMI)

Body Mass Index (BMI) is a numerical value calculated based on an individual's weight and height (Why Is BMI Important? n.d.). It is frequently employed as a widely acknowledged, straightforward approach to determine whether a person is a healthy weight for their height. A person's overall weight status—underweight, normal weight, overweight, or obese—can be determined using their BMI.

The following is the BMI calculation formula:

- Weight (in kilos) divided by height (in meters) equals BMI.

- If height is measured in inches and weight is measured in pounds, the equation becomes:

- Weight (in pounds) / Height (in Inches) x 2 x 703 is the formula for BMI.

The resulting BMI value is a numeric representation of a person's body mass relative to their height. Based on specified ranges, the BMI number is interpreted differently:

- Underweight if BMI is less than 18.5

- BMI of 18.5 to 24.9 indicates a normal weight

- Overweight if BMI is between 25 and 29.9

- Obese, BMI 30 or higher

Diabetes vs BMI

Figure 1 is a strip plot below that shows the BMI values of 1000 patients who are classified into three categories: N (non-diabetic), P (pre-diabetic), and Y (diabetic). The x-axis represents the diabetes category, and the y-axis represents the BMI value. Each dot on the plot corresponds to one patient's BMI value.

From the strip plot, we can observe some patterns and trends:

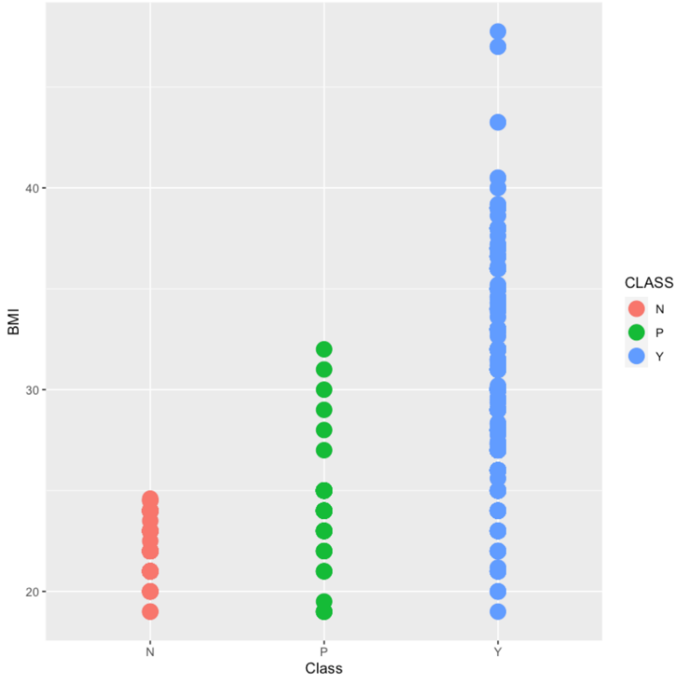

Figure 1 Diabetes vs BMI.

- The majority of the patients are diabetic (Y), followed by non-diabetic (N) and pre-diabetic (P).

- The BMI values range from 0 to 50, with most of them falling between 20 and 40.

For each category, the BMI means, medians, modes, first quartiles, and third quartiles are as Table 1.

Table 1 Category and value

Category	Mean	Median	Mode	1st Quartile	3rd Quartile
N, non-diabetic	22.4	22	24	21	24
P, pre-diabetic	23.9	24	24	23	25
Y, diabetic	30.8	30	30	28	33

- The lowest BMI values are seen in the non-diabetic individuals (N), the majority of whom fall within the normal weight range (18.5 to 24.9).

- Compared to non-diabetic patients, pre-diabetic patients (P) had somewhat higher BMI values, with some of them falling within the overweight range (25 to 29.9).

- The group of diabetes patients (Y) with the highest BMI values—the majority of whom are obese (30 or higher)—is composed of diabetic patients.

- The overlap between the groups, particularly between P and Y, suggests that there isn't a definite line separating them based just on BMI.

- There are some outliers in each group, such as a patient who is not diabetic and has a BMI of 47.8 and a patient who is diabetic and has a BMI of 0.

These findings suggest that diabetes and BMI are positively correlated, which means that having a higher BMI tends to increase your chance of developing diabetes. This does not necessarily suggest causality, however, as there may be other factors that have an impact on both variables. Furthermore, because it does not take into consideration muscle mass, bone density, or body form, BMI is not a perfect indicator of body fatness or health. A causal link between diabetes and BMI must thus be established via additional investigation and analysis.

Relationship between BMI and Diabetes

The relationship between Body Mass Index (BMI) and diabetes has been extensively studied, highlighting the significant impact of weight on the development and management of the disease. BMI, which is a numerical representation of body mass about height, is a key marker of a person's weight status and has become an important tool in comprehending the link between obesity and diabetes (Nguyen et al., 2011).

The risk of acquiring type 2 diabetes and BMI are strongly positively correlated, according to research (Gupta & Bansal, 2020). In comparison to those with normal BMI values, individuals with higher BMI values, especially those in the overweight and obese categories (BMI 25), have shown an increased vulnerability to developing diabetes. This connection can be explained by some reasons, including insulin resistance, persistent lowgrade inflammation, and poor glucose metabolism, all of which are more common in people who are overweight or obese.

Additionally, BMI is a useful indicator of the development of diabetes and related problems. Insulin resistance, dyslipidemia, hypertension, and cardiovascular illnesses are all associated with higher BMI values and all of them play a role in the emergence of problems connected to diabetes. To achieve glycemic control and reduce the risk of long-term problems, people with obesity-related diabetes may need more intense management measures.

Although BMI is a helpful screening tool and predictor of risk factors associated with being overweight, it has limitations in terms of capturing individual differences in body composition, muscle mass, and adipose tissue distribution. Therefore, using other measurements like waist circumference, body fat percentage, and waist-to-hip ratio can offer a more thorough evaluation of metabolic health and the risk of diabetes.

In conclusion, it is generally known that there is a connection between BMI and diabetes and that those with higher BMI numbers have a higher chance of getting and maintaining the disease. BMI is a useful measure for identifying people who are more at risk, directing therapeutic approaches, and assessing how well treatment plans are working. To encourage early identification, prevention, and focused treatments to lessen the burden of the illness and its effects on people and healthcare systems, it is crucial to comprehend the complex link between BMI and diabetes.

Cholesterol

The liver naturally produces cholesterol, a waxy, fatty molecule that is also acquired from dietary sources. It is a crucial part of cell membranes and is required for many physiological functions, such as the creation of hormones, vitamin D, and bile acids. While the body needs cholesterol for optimal operation, high amounts can lead to health issues, notably cardiovascular illnesses. There are two primary types of cholesterol:

- low-density lipoprotein (LDL) cholesterol, often referred to as "bad" cholesterol.

- high-density lipoprotein (HDL) cholesterol, known as "good" cholesterol.

When evaluating cholesterol levels, it is important to consider the ratio of LDL cholesterol to HDL cholesterol, as well as other lipid parameters such as triglycerides. High levels of triglycerides, a type of fat found in the blood, can also contribute to cardiovascular disease risk. The ratio of LDL cholesterol to HDL cholesterol, as well as other lipid factors like triglycerides, should be taken into account when assessing cholesterol levels. The risk of cardiovascular disease can also be increased by having high blood triglyceride levels.

A blood test called a lipid profile is commonly used to measure these cholesterol levels. The findings reveal the amounts of triglycerides, LDL cholesterol, HDL cholesterol, and total cholesterol. Based on these findings, medical practitioners can determine a person's cardiovascular disease risk and suggest the best dietary changes, drugs, or other treatments to control cholesterol levels and lower the risk of problems. Also, an ideal cholesterol level test should include:

- Total cholesterol: Often called 'serum cholesterol' or 'TC' refers to one’s overall level of cholesterol

- Non-HDL cholesterol: the total cholesterol minus HDL cholesterol. High non-HDL cholesterol, including LDL cholesterol, leads to a buildup of cholesterol in your arteries. Ideally, it should be as low as possible.

- HDL cholesterol: HDL cholesterol ('good' cholesterol) helps clear the cholesterol out of arteries, while LDL cholesterol ('bad' cholesterol) can clog them up. HDL cholesterol should ideally be high, around 1.4mmol/L, but HDL levels higher than this may not give any extra benefit.

- TC: HDL ratio: The ratio of HDL to total cholesterol is known as the TC: HDL ratio. This ought to be as low as is practical. High is defined as 6 or higher. Even if the TC: HDL ratio is low, it is still vital to include all the other "numbers" in addition to the TC: HDL ratio. For example, it is crucial to consider both HDL cholesterol and non-HDL cholesterol

Cholesterol levels in diabetes patients might differ based on some variables, including personal traits, diabetes care, and other comorbidities. For those with diabetes, there are some broad recommendations for ideal cholesterol levels (Heart UK, n.d.):

Table 2 Healthy Cholesterol Level

	Units: mmol/L	Units: mg/dL
Total (serum) cholesterol	below 5.0	below 193
Non-HDL cholesterol	below 4.0	below 155
LDL cholesterol	below 3.0	below 116
HDL cholesterol	above 1.0 for a man above 1.2 for a woman (ideally around 1.4)	above 39 for a man above 46 for a woman
TC: HDL ratio	The above 6 is considered high risk	The above 6 is considered high risk
Fasting triglyceride	below 1.7	below 150
Non-fasting triglyceride	below 2.3	below 204

However, it's critical to stress that these target levels may change depending on a person's unique circumstances, including age, the existence of other medical disorders, and overall cardiovascular risk. Diabetes patients should establish their ideal cholesterol objectives and create a thorough plan for controlling cholesterol levels in close consultation with their healthcare professionals.

Diabetes vs Cholesterol

The Figure 2 strip plot is a type of graphical display that shows the individual data points along a single axis. It can be useful to visualize the distribution and variation of a variable, as well as to compare different groups or categories of data. In this case, the cholesterol levels are displayed on the y-axis in mmol/L. There are 103 non-diabetic patients (N), 53 pre-diabetic patients (P), and 844 diabetic patients (Y) in the data set. The strip plot will be used to examine the association between cholesterol levels and diabetes in 1000 participants.

The strip plot of the data displays some intriguing patterns and trends. We can observe, for instance, that:

- The range of cholesterol values in the three groups is 0 to 10.6 mmol/L.

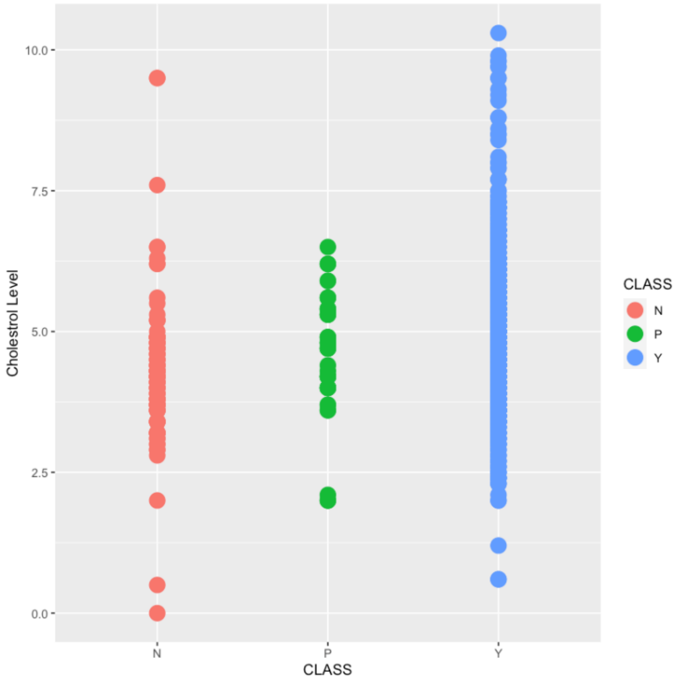

Figure 2 Diabetes vs Cholesterol.

- Each class's average cholesterol level is greater than the healthy person's recommended level of 5 mmol/L.

- Except for the non-diabetic class (N), which has a median of 4.2 mmol/L, the median cholesterol level for each class is likewise greater than the optimum range.

- There are more outliers with high cholesterol levels than low cholesterol levels, as shown by the fact that the mode cholesterol level for each class is lower than the mean and median.

- The distribution of the middle 50% of the data for each class is displayed in the first and third quartiles. The diabetic class (Y) has a wider distribution than the pre-diabetic class (P), the non-diabetic class (N), and the non-diabetic class (N). This shows that diabetes individuals' cholesterol levels vary more than those of non-diabetic or pre-diabetic patients.

- The strip plot also reveals some potential extreme values or outliers in the data. For instance, some diabetic individuals (Y) have very low cholesterol levels below 1 mmol/L, but some non-diabetic patients (N) have extremely high cholesterol levels exceeding 9 mmol/L. These numbers might be a sign of measurement or data collecting mistakes or abnormalities.

These findings suggest that there is a positive correlation between cholesterol levels and diabetes, i.e., that greater cholesterol levels are likely to be associated with higher levels of diabetes. This does not, however, suggest a causative link between the two variables, as other variables, such as age, gender, food, activity, genetics, etc., may affect both cholesterol levels and diabetes. We would need to do further experiments or analyses that account for these confounding variables to demonstrate a causal association.

An effective tool for comparing and visualizing the distribution of a single variable over many groups or categories is the strip plot, in conclusion. In this illustration, we looked at the correlation between cholesterol levels and diabetes in 1000 individuals using a strip plot. We discovered a strong correlation between the two variables; however, it is not necessarily a causative one. To better comprehend the data, we also found some descriptive statistics and possible outliers.

Moreover, diabetes can contribute to alterations in lipid metabolism, including increased levels of total cholesterol, LDL cholesterol, and triglycerides, so we will need to dive deeper and study the respective data of HDL, LDL, and triglycerides.

Diabetes vs High-Density Lipoprotein

Figure 3 is a strip plot graph that depicts the relationship between diabetes classes (Non-Diabetic, Pre-Diabetic, and Diabetic) and their corresponding levels of high-density lipoprotein (HDL) cholesterol. N, P, and Y stand for non-diabetic patients, pre-diabetic patients, and diabetic patients, respectively, while the X-axis symbolizes the various groups of diabetes patients. The HDL cholesterol levels are displayed on the Y-axis and range from 0 to 10.6 mmol/L.

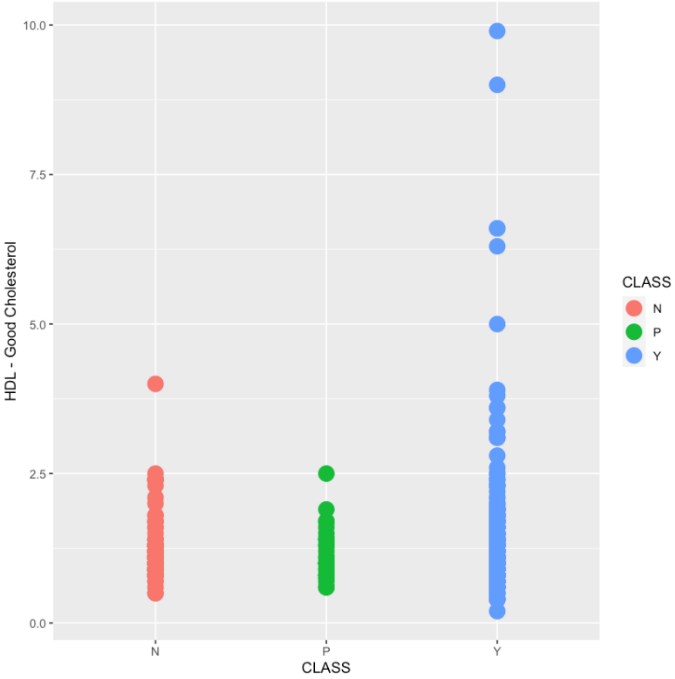

Figure 3 Diabetes vs HDL.

The 1000 patient dataset contains 844 people with diabetes (Y), 53 people with pre-diabetes (P), and 103 people without diabetes (N). Following are the HDL cholesterol values for each category. An explanation of the strip plot graph between diabetes and HDL is given in the passages that follow:

- The non-diabetic patients (N): The HDL cholesterol levels for the non-diabetic class (N) vary from 0.5 to 4 mmol/L. This class's mean HDL cholesterol value is 1.2 mmol/L, median HDL cholesterol is 1.1 mmol/L, and mode HDL cholesterol is 0.9 mmol/L.

- The pre-diabetic patients (P): The HDL cholesterol levels for people in the pre-diabetic class (P) vary from 0.6 to 2.5 mmol/L. For this class, the mean HDL cholesterol value is 1.1 mmol/L, the median is 1 mmol/L, and the mode is 1 mmol/L.

- The diabetic patients (Y): The range of HDL cholesterol in the diabetes class (Y) is 0.2 to 9.9 mmol/L. The median and mode values for HDL cholesterol in this class are 1.1 mmol/L and 1.2 mmol/L, respectively.

The graph highlights some commonalities and distinctions between the classes. With a mean of 1.2 mmol/L, a median of 1.1 mmol/L, and a mode of 0.9 mmol/L, class N, for instance, has the widest range in HDL levels, ranging from 0.5 to 4 mmol/L. The first and third quartiles are, respectively, 0.9 and 1.3 mmol/L. This shows that the majority of non-diabetic people have HDL levels that are normal or high, which are thought to be protective against cardiovascular illnesses.

With a mean of 1.1 mmol/L, a median of 1 mmol/L, and a mode of 1 mmol/L, Class P has a narrower range of HDL values, ranging from 0.6 to 2.5 mmol/L. The first and third quartiles are, respectively, 0.8 and 1.4 mmol/L. This indicates that pre-diabetic individuals' HDL levels are marginally lower than those of non-diabetics, which may raise their risk of developing diabetes or cardiovascular problems.

With a mean of 1.2 mmol/L, a median of 1.1 mmol/L, and a mode of 1.1 mmol/L, Class Y has the greatest range in HDL values, ranging from 0.2 to 9.9 mmol/L. The first and third quartiles are, respectively, 0.9 and 1.3 mmol/L. This demonstrates that diabetes individuals' HDL levels are highly variable, with some having either low or extremely high numbers. High HDL levels may be a sign of various health issues or genetic predispositions, whereas low HDL levels are linked to an increased risk of cardiovascular illnesses.

The strip plot graph can help us understand the relationship between HDL and diabetes by comparing the distribution and statistics of the three classes. It can also help us identify outliers or anomalies in the data that may require further investigation or explanation.

The strip plot graph also highlights the range and major tendency for each class of diabetes patients as it graphically represents the distribution of HDL cholesterol levels across them. However, it does not imply that maintaining HDL cholesterol levels and managing diabetes risk depend greatly on regulating HDL cholesterol levels. Further analysis and consideration of other factors are necessary to draw definitive conclusions about the significance of diabetes on HDL cholesterol levels.

Diabetes vs Low-Density Lipoprotein

Figure 4 is a strip plot graph that illustrates the relationship between diabetic patients' classes (Non-Diabetic, Pre-Diabetic, and Diabetic) and their corresponding levels of low-density lipoprotein (LDL) cholesterol.

Three kinds of diabetes patients are represented on the strip plot's x-axis: N (non-diabetic), P (pre-diabetic), and Y (diabetic). The LDL cholesterol ranges from 0 to 10.6 mmol/L on the y-axis. The LDL level of each dot on the figure corresponds to a specific patient in each class.

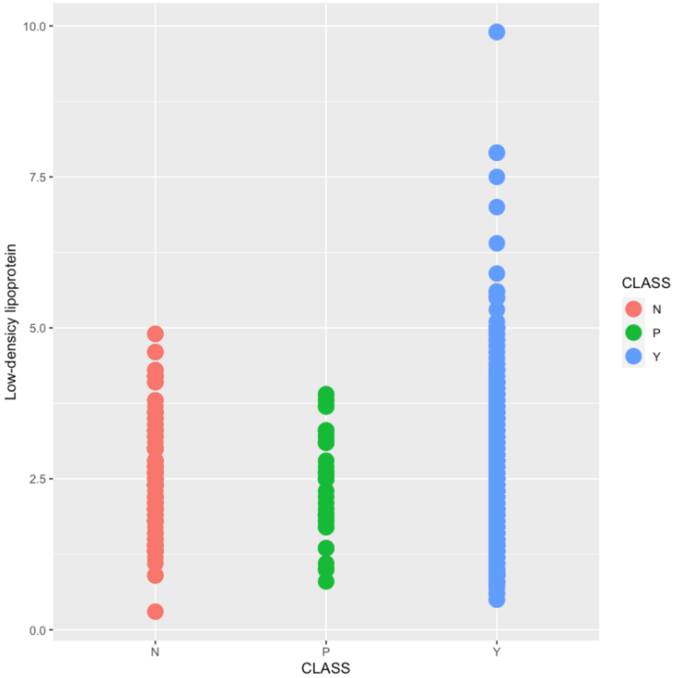

Figure 4 Diabetes vs LDL.

The strip plot displays some intriguing similarities and contrasts across the three patient groups. For instance:

- The non-diabetic patients (N): With a mean of 2.6 mmol/L, a median of 3 mmol/L, and a mode of 2.6 mmol/L, the non-diabetic patients (N) had the lowest range of LDL values, ranging from 0.3 to 4.9 mmol/L. The first and third quartiles are, respectively, 1.9 and 3.25 mmol/L. This shows that there are no outliers or extreme results among the non-diabetic individuals, and the majority of them have healthy LDL levels below 3 mmol/L.

- The pre-diabetic patients (P): With a mean of 2.5 mmol/L, a median of 2.5 mmol/L, and a mode of 1.9 mmol/L, the pre-diabetic patients (P) had somewhat greater ranges of LDL values, from 0.8 to 3.9 mmol/L. The first and third quartiles are, respectively, 1.9 and 3.2 mmol/L. This indicates that there are a few outliers or extreme numbers among the pre-diabetic individuals, some of whom have borderline or high LDL levels, exceeding 3 mmol/L.
- The diabetic patients (Y): With a mean of 2.6 mmol/L, a median of 2.5 mmol/L, and a mode of 2.5 mmol/L, diabetic individuals (Y) had the widest range of LDL values, ranging from 0.5 to 9.9 mmol/L. The first and third quartiles are, respectively, 1.8 and 3.3 mmol/L. This suggests that there are multiple outliers or extreme results in this group of diabetic individuals, many of whom had high or extremely high LDL levels, exceeding 3 mmol/L.

The strip plot also reveals considerable overlap among the three patient groups, particularly between P and Y, indicating that not all patients with diabetes have high LDL levels and not all patients with pre-diabetes have low LDL levels. However, there may be a definite pattern that the LDL levels tend to rise and fluctuate more widely when the diabetes status worsens from N to P to Y.

According to the research, diabetes and LDL cholesterol levels are positively correlated, which means that diabetic individuals are more likely than non-diabetic ones to have higher LDL levels (American Diabetes Association, n.d.). Numerous things, including insulin resistance, obesity, inflammation, or hereditary predisposition, might be the cause of this. Heart attacks and strokes, which are frequent consequences of diabetes, are made more likely by high LDL levels.

Therefore, each data point on the strip plot has an important implication for the health and well-being of the individual patient it represents. It shows their LDL cholesterol level right now and their probable risk of future cardiovascular issues. Furthermore, it gives a foundation for comparison with other patients in the same class or across classes. The patients' and their physicians' decisions on the most suitable measures or therapies to reduce LDL and avoid or control problems associated with diabetes are based on the information provided.

Diabetes vs Triglycerides

The Figure 5 strip plot is a type of graphical display that shows each observation as a dot along a single axis. It might help analyze a variable's distribution and spot anomalies or data gaps. The strip plot in this instance depicts the association between triglyceride (TG) levels and diabetes in 1000 individuals.

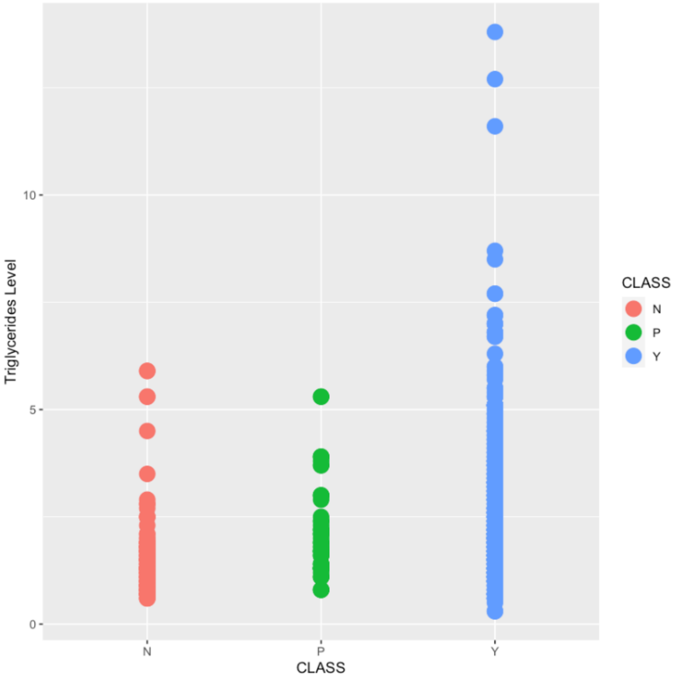

Figure 5 Diabetes vs TG.

The three categories N, P, and Y on the strip plot's x-axis stand for non-diabetic, pre-diabetic, and diabetic patients, respectively. Each patient's TG level is displayed on the y-axis in mmol/L. Below 1.7 mmol/L is the optimal TG level for a healthy individual. We may compare the TG levels across several types of diabetes and examine how they differ within each type using the strip plot.

The strip plot reveals the following:

- The non-diabetic patients (N): Of the three groups, the non-diabetic patients (N) have the lowest range and mean TG values. The majority of them have TG levels that are below the borderline or normal range of 2 mmol/L. The symmetric distribution of this category is indicated by the median and mode, both of which are 1.3 mmol/L.

- The pre-diabetic patients (P): The range and mean of TG levels are a little greater in pre-diabetic patients (P) compared to non-diabetic individuals. Some of them have TG levels that are high or extremely high, over 2 mmol/L. This category has a right-skewed distribution because the mean is 1.3 mmol/L and the median is 1.8 mmol/L.
- The diabetic patients (Y): Of the three groups, diabetes patients (Y) had the widest variation and mean of TG levels. TG levels in many of them are more than 2 mmol/L, and in some cases, they are even higher than 10 mmol/L, which is exceedingly high. This category's median and mean values, both 2.1 mmol/L, point to a symmetric distribution with some outliers.

According to the strip plot, diabetes and TG levels are positively correlated, implying that people with greater TG levels are more likely to have the disease than those with lower TG levels. This is in line with earlier research that showed elevated TG levels to be a risk factor for diabetes. The strip plot does not, however, prove a causal relationship between TG level and diabetes as there may be other variables that affect both variables.

Each data point in the strip plot reflects the TG level for a specific patient in that category. This can assist us in locating particular instances that may require more attention or intervention. As an illustration, we may see that a small percentage of non-diabetic people have extremely high TG levels, which may be a sign that they are at risk for developing diabetes or other health issues in the future. Similar to this, we may see that certain diabetic individuals have very low TG levels, which may signify a good course of therapy or management of their illness.

However, correlation does not imply causation, meaning that we cannot conclude that diabetes causes high TG levels or vice versa based on this data alone. We would need more evidence and comparison with other biomarkers to establish a causal relationship between these two variables.

According to some sources, normal TG levels for healthy people are below 1.7 mmol/L or 150 mg/dL (Understanding Your Cholesterol Test Results, n.d.). Based on this criterion, we can see that most patients in this data set have elevated TG levels, regardless of their diabetes status. This may indicate that they are at risk for cardiovascular diseases or other health problems related to high cholesterol.

Therefore, patients need to monitor their blood sugar and cholesterol levels regularly and follow their doctor's advice on how to manage their diabetes and lower their TG levels through diet, exercise, medication, or other interventions.

Diabetes vs Very-low-density Lipoprotein (VLDL) - carries Triglycerides

Figure 6 is a strip plot graph that displays the relationship between diabetic patient classes (Non-Diabetic, Pre-Diabetic, and Diabetic) and their corresponding very low-density lipoprotein (VLDL) cholesterol levels. The classifications of diabetic patients N, P, and Y are represented on the X-axis; N stands for non-diabetic patients, P for pre-diabetic patients, and Y for diabetic patients. The VLDL cholesterol readings, which range from 0 to 37.5, are represented on the Y-axis.

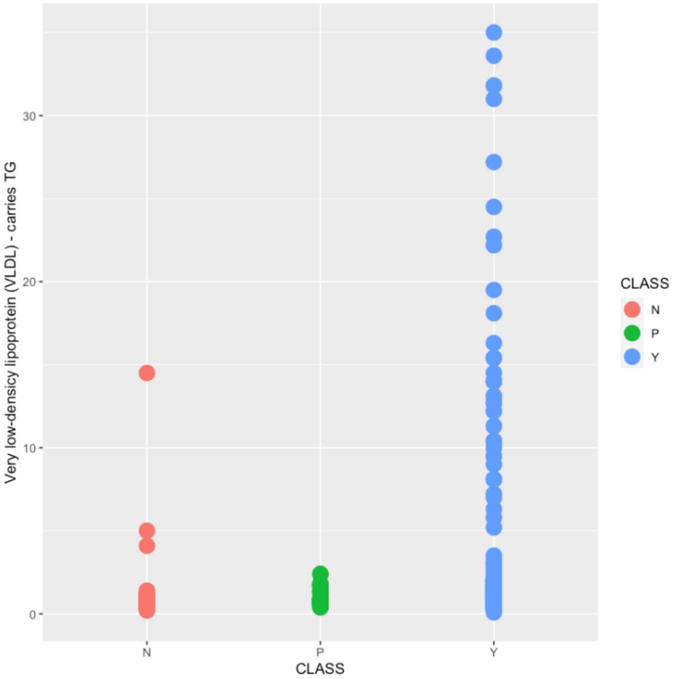

Figure 6 Diabetes vs VLDL.

There are 844 people with diabetes (Y), 53 people with pre-diabetes (P), and 103 people without diabetes (N) in the sample of 1000 patients. The VLDL cholesterol level of each patient from each class is represented by a data point on the graph.

Now let's discuss the characteristics of VLDL levels for each class:

- Non-Diabetic (N) Class: N class for non-diabetic: With a mean value of 0.9, the VLDL values in this class vary from 0.2 to 14.5. The mean value is 0.8, while the median is 0.7. Between the first and third quartiles, there is a difference of 0.5. These results suggest that the majority of non-diabetic VLDL levels are generally modest and fall within a normal range.
- Pre-Diabetic (P) Class: VLDL values for the Pre-Diabetic (P) Class vary from 0.4 to 2.4, with a mean value of 1.0. 0.8 is the median value, whereas 0.7 is the mode. The first and third quartiles are 0.6 and 1.3 respectively. According to these results, those who are pre-diabetic may have somewhat greater VLDL levels than those who are not diabetic, but these levels are still within a healthy range.
- Diabetic (Y) Class: With a mean value of 2.0, the VLDL levels in this class vary from 0.1 to 35. The mean value is 0.9, with 1.0 serving as the median. 1.5 and 3.0 are in the first and third quartiles, respectively. The diabetic group's VLDL levels had the broadest range of values, with some people having much higher amounts. This suggests that higher VLDL levels and diabetes are related.

The strip plot graph helps us to explore potential correlations between VLDL levels and diabetes and visually assess the distribution of VLDL levels within each class.

The VLDL levels of specific patients within their respective groups are represented by the implications of the data points in the graph. Examining the graph allows us to see any trends or outliers by observing the distribution and concentration of VLDL levels within each class.

We may infer the following conclusions about how VLDL levels and diabetes interact from the data:

- Non-Diabetic (N) Class: The majority of non-diabetics have VLDL levels that are within a normal range and are reasonably low.
- Pre-Diabetic (P) Class: Although still within an acceptable range, people in the pre-diabetic stage may have somewhat greater VLDL levels than people who are not diabetic.
- Diabetic (Y) Class: People with diabetes often have higher levels of VLDL, with some people having dramatically raised amounts. For diabetic individuals, managing and regulating VLDL levels is crucial to lowering their risk of cardiovascular problems.

Overall, the strip plot graph provides insights into the distribution and variation of VLDL levels across different classes of diabetic patients. It suggests that VLDL levels may increase as individuals progress from non-diabetes to pre-diabetes and diabetes. Monitoring and managing VLDL levels are crucial for diabetes management and reducing the risk of associated cardiovascular diseases.

Diabetes vs Urea

Figure 7 is a strip plot graph that illustrates the relationship between diabetic patient classes (Non-Diabetic, Pre-Diabetic, and Diabetic) and their corresponding urea levels. The classifications of diabetic patients N, P, and Y are represented on the X-axis; N stands for non-diabetic patients, P for pre-diabetic patients, and Y for diabetic patients. The levels of urea, which range from 0 to 37.5 mmol/L, are represented on the Y-axis.

There are 844 people with diabetes (Y), 53 people with pre-diabetes (P), and 103 people without diabetes (N) in the sample of 1000 patients. The urea level of each patient in each class is represented by a data point on the graph.

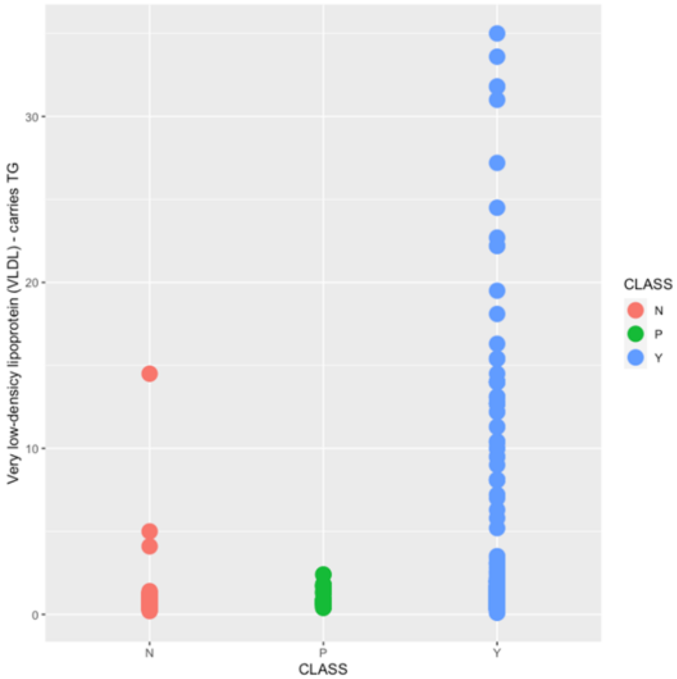

Figure 7 Diabetes vs Urea.

Examining each class's urea levels' features:

- Non-Diabetic (N) Class: The range of urea concentrations in this class is 2 to 22, with a mean value of 4.7. The modal value is 4.7, with 4.4 serving as the median. The first and third quartiles are 3.3 and 5.5 respectively. According to these numbers, the majority of urea levels in the non-diabetic class fall within a normal range.
- Pre-Diabetic (P) Class: With a mean value of 4.5, the urea levels for this class vary from 2.1 to 17.1. The modal value is 4.8, with 4.4 serving as the median. The first and third quartiles are 3.4 and 5.0 respectively. These results indicate that urea levels in people in the pre-diabetic stage are comparable to those in the non-diabetic group.
- Diabetic (Y) Class: The range of urea concentrations in this class is 0.5 to 38.9, with a mean value of 5.2. 4.6 is the median number, while 4.3 is the mode. Between the first and third quartiles are 3.7 and 5.8, respectively. The diabetic group's urea levels vary more widely, with some people having increased amounts. The majority of urea levels in this class still fall within the normal range, it is crucial to remember.

By viewing the strip plot graph, we may determine potential associations between urea levels and diabetes by observing the distribution and concentration of urea values within each class.

The graph's data points represent the urea levels of specific patients within each of their respective classifications. It enables us to quickly spot any trends or outliers by graphically identifying the distribution and concentration of urea readings.

We may infer the following conclusions about the connection between urea levels and diabetes from the data: Blood Urea Nitrogen (BUN), n.d.

- Non-Diabetic (N) Class: The majority of non-diabetics have urea levels that fall within the normal range, which shows that their kidneys are functioning properly.
- Pre-Diabetic (P) Class: Individuals in the pre-diabetic stage show urea levels similar to those of non-diabetic individuals, suggesting no significant deviation in kidney function at this stage.
- Diabetic (Y) Class: People with diabetes may have slightly higher urea levels than people without diabetes, which might indicate renal damage. Although most of the diabetic class's urea levels are still within the normal range.

The strip plot graph offers important insights into how urea levels vary and is distributed across various types of diabetes patients. It implies that urea levels might not be a reliable predictor of diabetes on their own. Elevated urea levels in diabetics, however, may signal the necessity for addressing problems linked to diabetes and monitoring renal function. To reduce the risk of kidney-related issues, people with diabetes must maintain a healthy lifestyle and often check on the condition of their kidneys.

Diabetes vs Average Blood Sugar Level HbA1c

In Figure 8, the strip plot in this instance displays the HbA1c average blood sugar levels of patients with N, P, and Y diabetes statuses. The HbA1c test measures the amount of glucose bonded to hemoglobin in red blood cells. It displays the typical blood sugar value for the last two to three months. An increased risk of diabetic complications is indicated by a greater HbA1c.

First, the comic strip plot demonstrates several intriguing trends and connections between diabetes and HbA1c. The HbA1c levels are first seen to be grouped around various ranges for each category. The lowest HbA1c values are seen in the non-diabetic patients (N), with values ranging from 0.9% to 5.6%, a mean of 4.6%, a median of 4.9%, and a mode of 4%. The HbA1c levels in pre-diabetic patients (P) range from 5.7% to 6.4%, with a mean, median, and mode of 6%. With a mean of 8.9%, a median of 8.8%, and a mode of 8%, diabetic patients (Y) have the highest and most variable HbA1c readings. These values range from 2% to 16%. These ranges are consistent with the diagnostic criteria for diabetes, which define normal HbA1c as below 5.7%, pre-diabetes as between 5.7% and 6.4%, and diabetes as above 6.5% (Kaur et al., 2020).

Second, the chance of getting diabetes increases when HbA1c readings rise, as can be seen from the fact that there is a positive association between the two. The fact that the data points are more closely clustered on the right side of the plot than on the left makes this obvious. The fact that there is some overlap between the categories is another indication that some patients' HbA1c results may not correspond to their diabetes status. For instance, some non-diabetic individuals have HbA1c levels above 5.7%, which raises the possibility that they may eventually acquire pre-diabetes or diabetes. Similarly, some pre-diabetic patients have HbA1c values above 6.5%, which indicates that they may already have diabetes or are at high risk of developing it soon.

On the other hand, some diabetic patients have HbA1c values below 6.5%, which indicates that they may have good blood sugar control or are in remission from diabetes.

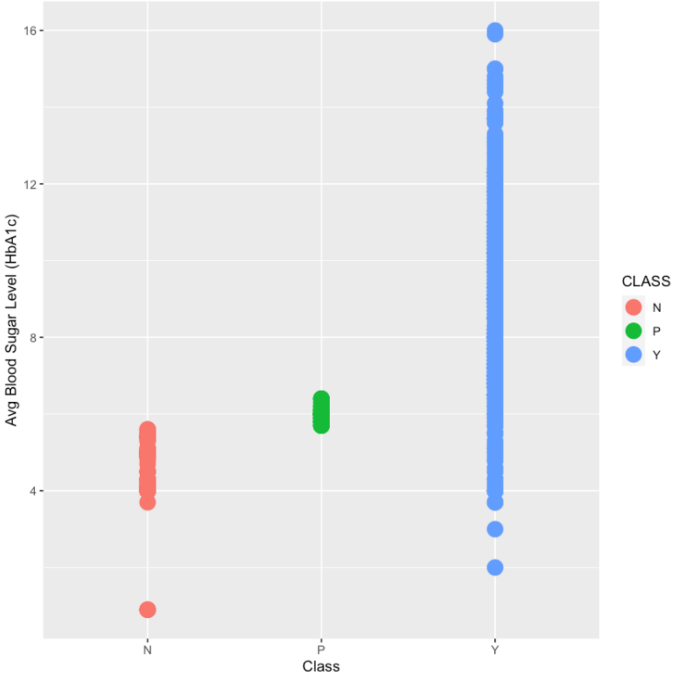

Figure 8 Diabetes vs Sugar Level.

Third, it is clear that each data point in the strip plot reflects the HbA1c and diabetes status of a specific patient. This implies that we may use this plot to contrast and compare the blood sugar levels of various people. The patients with an HbA1c value of 16% in the Y category or the patients with an HbA1c value of 2% in the Y category, for instance, are examples of outliers with extremely high or low HbA1c values for their category. We can also identify clusters of patients who have similar HbA1c values for their category, such as the group of patients with an HbA1c value of 6% in the P category or the group of patients with an HbA1c value of 8% in the Y category.

Finally, the strip plot graph between diabetes and average blood sugar level (HbA1c) demonstrates a distinct association between these two variables and offers helpful details regarding the distribution and variance of HbA1c values across various categories of diabetes status.

Diabetes and Age

Using the given dataset, we further investigated the association between age and diabetes in addition to looking at the biomarkers related to diabetes. Significant trends emerged from the examination of age distribution across the three groups of people (those with diabetes, those at risk for developing it, and those without diabetes), underscoring the importance of age for diabetes and biomarkers.

For individuals without diabetes (class N): With a mean age of 44.2, the age range was 25 to 77. The modal age was 50 and the median age was 44, showing a very evenly distributed population. According to the first quartile value of 38.5, 25% of non-diabetic people were under this age, and the third quartile value of 50, 75% of non-diabetic people were under this age. In the absence of diabetes, these figures offer a reference range for age distribution.

In the pre-diabetic group (class P): With a mean age of 43.3, the age range was 30 to 55. The distribution appears to be quite symmetrical given that the median and mean ages were 48 and 50, respectively. The first quartile value was 38, meaning that 25% of pre-diabetic people were under this age. The third quartile value was 50, meaning that 75% of pre-diabetic people were under this age. According to these results, those who are pre-diabetic are often a little bit younger than people who do not have the disease.

Among diabetic individuals (class Y): 20 to 79 years old was the age range, with a mean of 55.3. The greatest concentration of diabetic people occurred around this age, as indicated by the 55 median and mean ages. While the third quartile value of 60 indicated that 75% of diabetics were under this age, the first quartile value of 53 indicated that 25% of diabetics were below this age. According to this data, people with diabetes are often older than both pre-diabetic and non-diabetic people.

Age distribution patterns among the three groups of people have been discovered, and these patterns shed light on the association between age and diabetes. Our results indicate that age is a key factor in the onset and progression of diabetes, with diabetes prevalence rising with increasing age.

Conclusion

In this study, we investigated the effect of diabetes on various biomarkers, including BMI, cholesterol, HDL, LDL, TG, VLDL, urea, HbA1c, and age using strip plot graphs to analyze the data. Our results offer important light on the association between these biomarkers and diabetes, highlighting their importance in comprehending the course of the illness and assisting in its diagnosis and treatment. As the phases of diabetes advanced from non-diabetic to pre-diabetic and diabetic, the strip plot graphs showed unique patterns among the biomarkers. Pre-diabetic and diabetic people's BMIs were somewhat higher than those of non-diabetic people, suggesting a possible link between greater BMI and the onset of diabetes. Since cholesterol levels did not significantly differ between pre-diabetic and diabetic stages, it is possible that the development of diabetes does not directly affect cholesterol levels.

Between the various phases of diabetes, there were no appreciable variations in HDL levels, which is sometimes referred to as the "good" cholesterol. LDL levels, also known as "bad" cholesterol, did not show any significant differences. These findings imply that the development of diabetes may not have a significant impact on HDL and LDL levels and that these levels may instead be impacted by other variables.

However, TG levels showed a significant rise in diabetic people compared to non-diabetic and pre-diabetic people, suggesting a possible link between high TG levels and diabetes. Similar to how VLDL cholesterol levels increased gradually as diabetes worsened, this indicates that VLDL cholesterol may contribute to the onset and development of diabetes. As diabetes advanced, urea levels, a sign of kidney health, gradually increased, illustrating the effect of diabetes on renal health. According to the research, elevated urea levels can be a sign of diabetic nephropathy or diminished renal function. Notably, the long-term blood sugar management indicator HbA1c showed a definite rise with the development of diabetes. Because higher HbA1c levels are linked to worse glycemic control, this biomarker is important for controlling and monitoring diabetes.

Overall, our work offers insightful information on how different biomarkers and diabetes interact. The results imply that BMI, TG, VLDL, urea, and HbA1c can act as significant markers of diabetes development and may help with the identification and treatment of the condition. The underlying processes and possible therapeutic use of these biomarkers in diabetes require further study. Our work adds to the expanding body of information on the pathophysiology of diabetes by illuminating the connections between these biomarkers and diabetes and emphasizes the significance of thorough biomarker analysis in comprehending and treating the condition. It is also important to keep in mind that the influence of age may vary based on the particular biomarker when thinking about the association between age and biomarkers. Biomarkers including, cholesterol, HDL, and LDL, for instance, did not show much fluctuation with age. However, there were significant relationships between BMI, TG, VLDL, urea, and HbA1c levels and diabetes that were altered by age.

These results highlight the significance of age as a key aspect to take into account when discussing diabetes and biomarkers. Age is a risk factor as well as a potential confounding factor when interpreting the results of biomarker studies. Risk assessment, diagnosis, and management techniques specific to various age groups might benefit from an understanding of the age-related dynamics of biomarkers. The work emphasizes the importance of age in diabetes and biomarkers. It underlines the link between age and the prevalence of diabetes by highlighting the age distribution among diabetic, pre-diabetic, and non-diabetic persons. These results highlight the significance of age as a critical variable for evaluating the results of biomarkers like BMI, TG, VLDL, urea, and HbA1c levels and creating targeted diabetes treatment approaches. To enhance clinical outcomes for people of all ages and expand our understanding of the illness, future research should look more closely at the complex interactions between age, biomarkers, and diabetes.

Acknowledgments:

The research by M. Kamrujjaman was partially supported by the University Grants Commission (UGC), and the University of Dhaka, Bangladesh

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

The data is publicly available.

Reference

Cleveland Clinic (2023). Blood Urea Nitrogen (BUN): Testing, Levels & Indication. (n.d.). Cleveland Clinic. Retrieved May 29, 2023, https://my.clevelandclinic.org/health/diagnostics/17684-blood-urea-nitrogen-bun-test.

Gupta, S., & Bansal, S. (2020). Does a rise in BMI cause an increased risk of diabetes? Evidence from India. PLOS ONE, 15(4), e0229716.

Kaur, G., Lakshmi, P. V. M., Rastogi, A., Bhansali, A., Jain, S., Teerawattananon, Y., Bano, H., & Prinja, S. (2020). Diagnostic accuracy of tests for type 2 diabetes and prediabetes: A systematic review and meta-analysis. PloS One, 15(11), e0242415.

Lipids and Lipoproteins in Patients with Type 2 Diabetes | Diabetes Care | American Diabetes Association. (n.d.). Retrieved May 29, 2023, https://diabetesjournals.org/care/article/27/6/1496/22705/Lipids-and-Lipoproteins-in-Patients-With-Type-2.

Minakshi R. R. & Sushant S. K. (2022). Diabetes prediction and analysis using medical attributes: A Machine learning approach. Xi'an University of Architecture & Technology. https://doi.org/10.37896/JXAT14.01/314405.

Nguyen, N. T., Nguyen, X.-M. T., Lane, J., & Wang, P. (2011). Relationship between obesity and diabetes in a US adult population: Findings from the National Health and Nutrition Examination Survey, 1999-2006. Obesity Surgery, 21(3), 351–355.

Rashid, Ahlam (2020), "Diabetes Dataset", Mendeley Data, V1, doi: 10.17632/wj9rwkp9c2.1

Heart UK. (n.d.). Understanding your cholesterol test results. Retrieved May 29, 2023, https://www.heartuk.org.uk/cholesterol/understanding-your-cholesterol-test-results

Why is BMI Important? (n.d.). Retrieved May 29, 2023, https://www.diabetes.co.uk/bmi/why-is-bmi-important.html.

This paper DOI: 10.5281/zenodo.10791178
Journal New Website web1: https://ijgsw.eu.org/
web2: https://ijgsw.net/
Journal Old Website: http://ijgsw.comze.com/ is no longer used
You can submit your paper to email: Jichao@email.com
Or IJGSW@mail.com