Development of Predictive Model for Levels of Glycemic Control as a Measure of Diabetes Self-management

Zainab Amna

000402510

MS Healthcare Biotechnology

Supervisor

Prof. Dr. Attya Bhatti

HEALTHCARE BIOTECHNOLOGY

ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

AUGUST, 2024

Development of Predictive Model for Levels of Glycemic Control as a Measure of Diabetes Self-management

By

Zainab Amna

000402510

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Healthcare Biotechnology

Thesis Supervisor:

Prof. Dr. Attya Bhatti

HEALTHCARE BIOTECHNOLOGY

ATTA-UR-RAHMAN SCHOOL OF APPLIED BIOSCIENCES

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

AUGUST, 2023

FORM TH-4

National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by: (Student Name & Regn No.) **Zainab Amna, 00000402510**

Titled: "Development of predictive model for the level of glycemic control as a measure of diabetes self-management" be accepted in partial fulfillment of the requirements for the award of **MS Degree in Healthcare Biotechnology** with (A grade).

Examination Committee Members

1. Name: **Prof. Dr. Peter John** Signature:

2. Name: **Dr. Maria Shabbir** Signature:

3. Name: **Muhammad Saeed Mughal** Signature:

Supervisor's name: **Prof. Dr. Attya Bhatti** Signature:

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

Date: 22/08/2024

Dr. Hussain Mustatab Wahedi
Head of Department (HoD)
Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

Head of Department:

Date: 22/08/2024

COUNTERSINGED

A/Principal & Dean
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST, Islamabad

Dean/Principal

Date: 22/08/2024

Declaration

I Zainab Amna certify that this research work titled "*Development of predictive model for levels of glycemic control as a measure of diabetes self-management*" is my own work. The work has not been presented elsewhere for assessment. The material that has been used from other sources it has been properly acknowledged / referred.

Signature of Student

Zainab Amna

0000402510

THESIS ACCEPTANCE CERTIFICATE

Certified that content and forms of thesis entitled **"Development of predictive model for the level of glycemic control as a measure of diabetes self-management"** submitted by **Zainab Amna** has been fond excellent for the requirement of the degree.

Supervisor: _____

Professor Dr. Attya Bhatti
Atta-ur-Rahman School of Applied Biosciences (ASAB)
National University of Sciences and Technology (NUST)

Head of Department: _____

Atta-ur-Rahman School of Applied Biosciences (ASAB)
National University of Sciences and Technology (NUST)

Principal ASAB: _____

Atta-ur-Rahman School of Applied Biosciences (ASAB)
National University of Sciences and Technology (NUST)

Certificate for plagiarism

It is certified that M Phil/MS Thesis titled "*Development of predictive model for the level of glycemic control as a measure of diabetes self-management*" of Ms. Zainab Amna with Reg No. 000402510has been examined by us. We undertake the follows:

a) Thesis has significant new work as compared already published or are under consideration to be published elsewhere. No sentence, equation, diagram, table, paragraph or section has been copied verbatim from previous work unless it is placed under quotation marks and duly referenced.
b) The work presented is original and own work of the author (i.e. there is no plagiarism). No ideas, processes, results or words of others have been presented as Author own work.
c) There is no fabrication of data or results which have been compiled.
d) There is no falsification by manipulating research materials, equipment or processes, or changing or omitting data or results such that the research is not accurately represented in the research record.
e) The thesis has been checked using Turnitin (copy of originality report attached) and found within limits as per HEC plagiarism Policy and instructions issued from time to time.

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

(Supervisor)
Dr. Attya Bhatti
Associate Professor
ASAB, NUST

Copyright Statement

Acknowledgements

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought which Allah setup in my mind to improve it. Whosoever helped me throughout the course of my thesis, whether my parents or any other individual was Allah's will, so indeed none be worthy of praise but Allah.

I would also like to express special thanks to my supervisor Prof. Dr. Attya Bhatti for her help throughout my thesis and also for Molecular genetics course which she has taught me.

I would also like to thank my GEC member for being on my thesis guidance and evaluation committee and express my special thanks to Prof. Dr. Peter John for his help. I am also thankful to Afnan Waqar, Anum Mustafa, Hijab Fatima, Pakeezah Shahid and my lab fellows for their support and cooperation.

I am profusely thankful to my beloved parents who raised me when I was not capable of walking and continued to support me throughout in every department of my life. My deepest gratitude goes to my parents, whose steadfast emotional and financial support has been the bedrock of my research. Their unwavering belief in me and relentless encouragement have propelled me forward in every step of this journey.

I would also like to pay special thanks to elder brothers Zain-ul-Abdeen and Zubair Abdullah for their tremendous support and cooperation. Each time I got stuck in something, they came up with the solution. I appreciate his patience and guidance throughout the whole thesis.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my exceptional parents and adored siblings whose tremendous support and cooperation led me to this wonderful accomplishment.

Abstract

Diabetes mellitus is a chronic metabolic disorder that affects millions of people worldwide and has been related to serious complications such as cardiovascular disease, neuropathy, nephropathy, and retinopathy. Effective diabetes care is dependent on maintaining glycemic control, which is frequently measured using glycated hemoglobin (HbA1c) values. HbA1c is an important biomarker that represents average blood glucose levels over the previous two to three months and is used to determine long-term glycemic management and the risk of problems.

The goal of this research is to create a reliable and accurate HbA1c prediction model utilizing machine learning (ML) technologies. The objectives include assessing clinical and demographic data to find significant variables, creating and comparing multiple machine learning models, and determining the best effective model for HbA1c prediction. In addition, the study investigates the clinical consequences of precise HbA1c forecasts, as well as the obstacles and ethical problems involved with machine learning in health care.

The study indicated that Gradient Boosting Regression (GBR) outperformed other models, with the lowest Mean Squared Error (MSE) and greatest R^2 values. Random Forest Regression also fared well, although Neural Network Regression (NNR) and Support Vector Regression (SVR) were less effective due to their sensitivity to feature scaling and parameter adjustment. Accurate HbA1c forecasts can assist healthcare practitioners anticipate levels and enhance glycemic control, but issues like individual variability and data security must be addressed.

Key Words: *Gradient Boost regression model, HbA1c level, Machine learning, Predictive models, Self-management, Type 2 diabetes mellitus*.

Table of Contents

List of Figures

List of Tables

List of Acronyms

T2DM: Type 2 Diabetes mellitus

T1DM: Type 2 Diabetes mellitus

HbA1c: Hemoglobin A1c

ADA: American Diabetes Associstion

IDF: International Diabetes Federation

GCK: Glucokinase

AI: Artificial Intelligence

ML: Machine Learning

GBR: Gradient Boost Regression

SVR: Support Vector Regression

IGT: Impaired Glucose Tolerance

DR: Diabetic Retinopathy

CDSS: Clinical Decision Support System

MSE: Mean Squared Error

R^2: Coefficient of Determination

MLA: Machine Learning Algorithms

INTRODUCTION

1.1 Diabetes mellitus

Diabetes mellitus is chronic, multifactorial disorder that is characterized by physiologically abnormal hyperglycemic condition. Elevated blood sugar levels are mainly due to abnormality in insulin secretion or its action. The molecular mechanisms involved in insulin synthesis, release, and detection are strictly regulated since these processes are necessary for maintaining glucose homeostasis (Galicia-Garcia et al., 2020). Any one of the mechanisms underlying these processes could be flawed, resulting in a metabolic imbalance that fuels the disease's progression. Diabetes is a complex disorder that is characterized by fat and protein imbalances in association with carbohydrate imbalances and hyperglycemia. These metabolic disturbances influence the functioning of multiple organs mainly affecting vasculature of organs (Banday et al., 2020). There are 5 types of diabetes mellitus: Type 1 diabetes, type 2 diabetes, gestational diabetes, MODY (Maturity onset diabetes of young), and NDM (Neonatal diabetes mellitus). Type 2 diabetes is more prevalent, comprising about 90% of all cases (Banday et al., 2020).

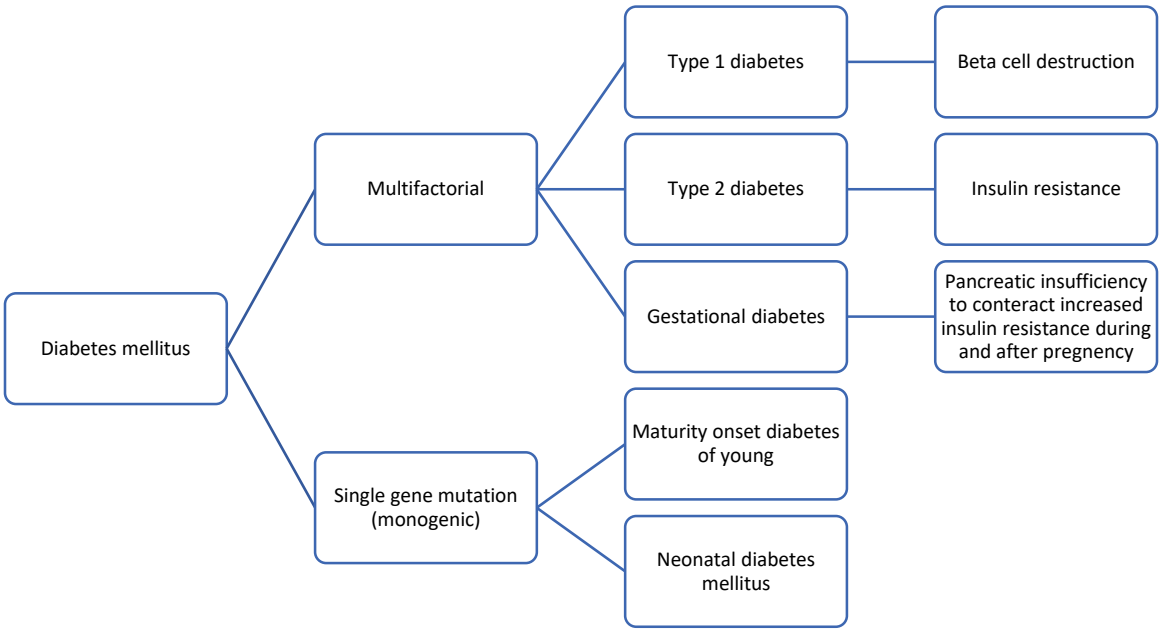


Figure 1.0.1: Classification of diabetes mellitus.

Type 2 diabetes mellitus is referred to as adult-onset diabetes or non-insulin dependent diabetes mellitus (NIDDM). T2DM advances slowly and without symptoms; even mild hyperglycemia takes years to develop, and as a result, the condition is mainly undetected until the advanced stages of the disease manifest the classic symptoms of severe hyperglycemia, which include weight loss, growth impairment, blurred vision, polyuria, and polydipsia (Banday et al., 2020).

1.1.1 Pathophysiology of type 2 diabetes:

Insulin resistance and β-cell dysfunction are the two main insulin related abnormalities. Insulin resistance arise from the disturbance of multiple cellular pathways, resulting in reduced sensitivity or reaction of peripheral tissue cells, specifically the liver, adipose tissue and muscle, to insulin (Banday et al., 2020). Elevated blood glucose levels are a result of dysfunctions in insulin secretion and action in Type 2 Diabetes Mellitus (T2DM). Firstly, decreased insulin sensitivity causes β-cell hyper function, which in turn causes an increase in insulin production as a coping mechanism to keep blood sugar levels normal and avoid hyperglycemia. But as time goes on, β-cell function deteriorates and insulin production is no longer enough to compensate for decreasing insulin sensitivity (Banday et al., 2020). Consequently, because normal glycaemia cannot be maintained, hyperglycemia sets in. Through many biological pathways, factors such as hormones, carbohydrates, and amino acids cause β-cells to produce insulin. Stressors such as hyperglycemia and hyperlipidemia put dysfunctional β-cells under pressure, which can result in oxidative stress, ER stress, and inflammation (Galicia-Garcia et al., 2020). These can ultimately affect insulin production and play a role in the pathophysiology of T2DM. Inadequate processes related to the synthesis or secretion of insulin worsen β-cell dysfunction.

1.1.2 Epidemiology of type 2 diabetes:

The prevalence of diabetes mellitus has more than doubled worldwide over the last three decades, making it one of the biggest public health concerns facing every country. Over the past three decades, the prevalence of type 2 diabetes mellitus (T2DM) and prediabetes has increased dramatically worldwide, especially in emerging nations. Apart from the early development of type 2 diabetes in young adults, there is a discernible upward trend in the prevalence of both T2DM and prediabetes in children and adolescents (Chen et al., 2011).

According to IDF (International Diabetes Federation) 536 million adults (1 in 10 adult) are living with diabetes. By 2029 and 2045, this number is expected to increase to 643 million and 783 million, respectively. 6.7 million Individuals (1 every 5 seconds) died due to diabetes in 2021. Impaired Glucose Tolerance (IGT) puts 540 million persons at high risk of developing type 2 diabetes.

In South East Asia, 1 in 11 adults (about 90 million) have diabetes, according to the IDF 2021 Diabetes Atlas. By 2029, there will be 113 million adults with diabetes worldwide, and by 2045, there will be 151 million. More than half of adult diabetics go undiagnosed. Diabetes will be a factor in 747,000 fatalities in 2021.

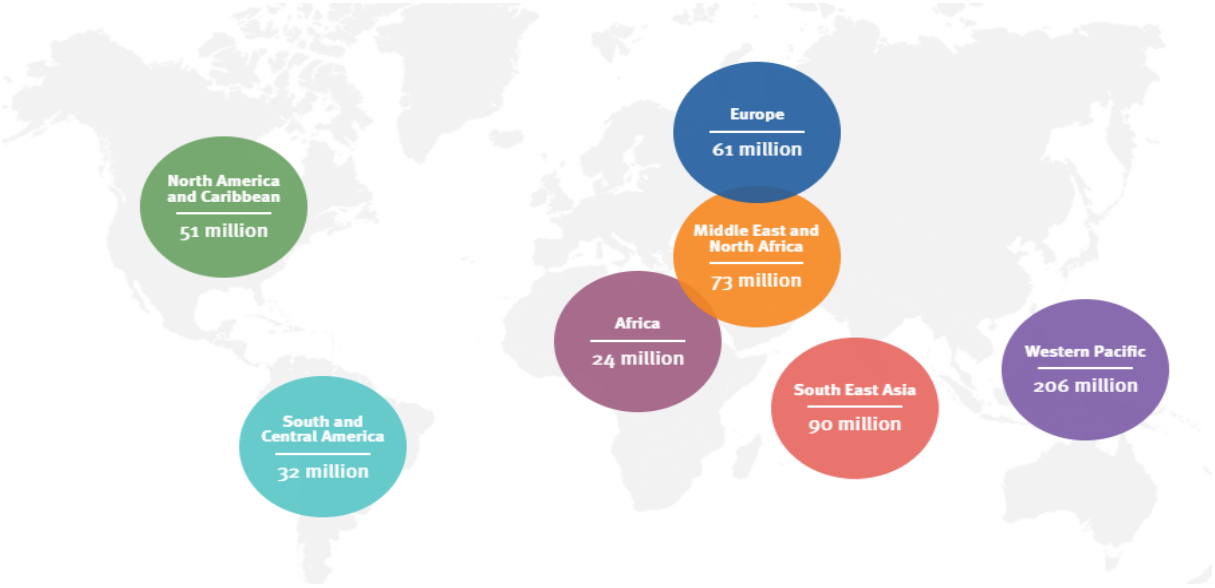


Figure 1.0.2: Diabetes around the world in 2021 (IDF Diabetes Atlas)

The global burden of diabetes is increasing at a faster rate. Pakistan ranks in the 7th position with the highest burden of diabetes (Ishaque et al., 2016). Thus, this is an alarming situation that should be addressed by healthcare and the clinical system. Diabetes care expenses surpass the average per capita healthcare spending by a factor of at least 3.2, escalating to 9.4 times in cases of complications (Khan et al., 2020). Early prevention, self-management, and intervention strategies must be employed to reduce mortality and morbidity rates.

1.1.3 Etiology of type 2 diabetes:

A range of behavioral and environmental risk factors, in addition to genetic and epigenetic predispositions, are responsible for the epidemic of type 2 diabetes (Chen et al., 2011). The chances of an individual developing diabetes are five times higher in individuals who have blood relations with diabetic patients (Liu, 2020). Several gene mutations, like mutations in the insulin gene and the gluco-kinase gene, are also responsible for type 2 diabetes (Liu, 2020). Several physical factors like age, unhealthy diet, BMI, reduced exercise, sedentary lifestyle, psychological stress, blood pressure, blood lipids, smoking, viral infections, taking steroids, and other medications contribute to the onset of diabetes and its increased worldwide prevalence (Khan et al., 2020; Liu, 2020). Several pathological conditions also contribute to the etiology of type 2 diabetes mellitus. The many microbial species that make up the gut microbiota are essential to human physiology, since they influence immune response, gut barrier integrity, and metabolism. Age, diet, and lifestyle factors can all lead to dysbiosis, which can upset the metabolic balance and aggravate conditions like type 2 diabetes (T2DM). Insulin resistance and low-grade inflammation can result from dysbiosis brought on by things like a high-fat diet. More investigation is needed to fully comprehend the clinical significance of gut dysbiosis in type 2 diabetes (Galicia-Garcia et al., 2020). The persistence of diabetic problems in the face of glycemic management is known as metabolic memory. It is seen in large-scale clinical studies and is associated with oxidative stress, non-enzymatic protein glycation, chronic inflammation, and epigenetic modifications. MicroRNAs (miRNAs) are essential for post-transcriptional gene regulation and modulate the expression of genes linked to insulin production and β-cell activity, which helps to maintain metabolic memory (Galicia-Garcia et al., 2020). Age-related insulin resistance, oxidative stress, and the development of type 2 diabetes are all increasingly linked to mitochondrial dysfunction. It involves damage brought on by oxidative stress, deficiencies in mitochondrial biogenesis, and compromised mitophagy. The risk of type 2 diabetes is also influenced by mitochondrial genetics, which includes mtDNA variations (Galicia-Garcia et al., 2020).

In understanding the multifaceted nature of this disease, various factors come into play, ranging from genetic predispositions to environmental influences. Within the context of South Asia, particularly Pakistan, a complex interplay of genetics, maternal health, gender disparities, and age

demographics, diet, obesity, physical inactivity, depression and sleep pattern contributes to the prevalence and onset of diabetes.

1.1.3.1 Genetics:

Research indicates that individuals may be genetically predisposed to type 2 diabetes. While lifestyle factors were not sufficiently taken into account, ethnic variances across the nation suggest genetic influences (Hakeem & Fawwad, 2010).

1.1.3.2 Maternal Hyperglycemia and Undernutrition:

Metabolic syndrome, macrosomia, and altered placental shape in offspring are all increased by gestational diabetes. Even in women who do not have diabetes, higher fasting plasma glucose levels are associated with a higher risk of macrosomia and preeclampsia. Pregnancy-related malnutrition results in Intrauterine Growth Retardation (IUGR) and impaired child metabolic capability. Pakistan has high rates of low birth weight and maternal death, which raises the risk of metabolic problems in young children (Hakeem & Fawwad, 2010).

1.1.3.3 Gender:

Gender influences type 2 diabetes risk through hormonal variations, body composition differences, and lifestyle variances, impacting insulin sensitivity, genetic predisposition, and healthcare access disparately between men and women (Kautzky-Willer et al., 2023). Poor glucose tolerance is more common in women, while diabetes is more common in men (Hakeem & Fawwad, 2010).

1.1.3.4 Dietary habits:

The relevance of a nutritionally balanced diet is emphasized in diabetes management guidelines; nevertheless, the effect of malnutrition on diabetes risk has received more attention recently. There is evidence that micronutrient deficits increase the risk of diabetes (Hakeem & Fawwad, 2010). Nonetheless, little is known about how undernutrition contributes to the development of type 2 diabetes. Increased blood glucose and lipid-rich particles are caused by the high-calorie Western diet, which also increases oxidative stress, inflammation, and reactive oxygen species (ROS). Insulin resistance (IR) and type 2 diabetes mellitus (T2DM) are facilitated by this. Increased ROS levels contribute to diabetic problems by inducing ER stress, mitochondrial malfunction, and activation of pathways such as NADPH oxidase (NOX). Elevated levels of free fatty acids (FFAs)

worsen the functioning of mitochondria (Galicia-Garcia et al., 2020). Different global relationships between food groups and T2D risk have been found by prospective epidemiological research. Eating food high in plant matter reduce danger, while those low in energy density provide protection (Kolb & Martin, 2017).

1.1.3.5 Physical inactivity:

Risk of most metabolic disorders are linked with low physical activity (Hakeem & Fawwad, 2010). High physical activity levels, including both recreational and work-related activities, reduce the incidence of diabetes by about 29%, according to epidemiological research (Kolb & Martin, 2017). Controlled research show that exercise improves glycemic management and insulin sensitivity (Kolb & Martin, 2017). Insulin sensitivity is improved by substituting moderate-to-intense physical activity for sedentary time. Obesity and type 2 diabetes mellitus (T2DM), which is linked to persistent low-grade inflammation, are caused by decreased physical activity and an increase in sedentary behavior (Galicia-Garcia et al., 2020). Physical activity and exercise increase anti-inflammatory cytokines and lower levels of inflammatory markers, such as CRP, IL-6, and IL-18, in the blood. Additionally, exercise increases the production of antioxidants, which lowers oxidative stress (Galicia-Garcia et al., 2020). The risk of diabetes is doubled by sedentary behavior, especially extended TV watching, regardless of physical activity level (Kolb & Martin, 2017).

1.1.3.6 Obesity:

One of the best indicators of type 2 diabetes is obesity, those who are overweight have increased incidence of the disease. Obesity rates are high in both women and children and there is evidence linking obesity to insulin resistance and metabolic risk (Hakeem & Fawwad, 2010). Because obesity increases insulin resistance, a defining feature of type 2 diabetes mellitus, it has a substantial impact on the regulation of systemic glucose levels (T2DM) (Banday et al., 2020). Even while not all people with type 2 diabetes are overweight or obese, there is a clear correlation between insulin resistance and hyperglycemia and excess body fat, particularly visceral or abdominal fat (Banday et al., 2020).

1.1.3.7 Socio-economic status:

Whether assessed by income, occupation, or education, type 2 diabetes (T2D) and socioeconomic level are consistently inversely correlated globally. When compared to higher status groups, a low socioeconomic status is associated with a 39–60% higher relative risk of type 2 diabetes (Kolb & Martin, 2017). Lower socioeconomic categories have a higher risk of diabetes, which is mediated by factors including smoking and inactivity and illiteracy about the disease. When combined with the right dietary and lifestyle modifications, income growth may lower the risk of type 2 diabetes. Socioeconomic factors like poverty, gender bias, a lack of educational opportunities, flaws in the healthcare system, and a reliance on non-evidence-based preventive and care strategies are acknowledged as major contributors to diabetes risk, even though research on the socio-cultural determinants of diabetes in Pakistan is lacking (Hakeem & Fawwad, 2010).

1.1.3.8 Depression:

There is a substantial correlation between newly diagnosed type 2 diabetes and depression (Hakeem & Fawwad, 2010). The findings of prospective and cross-sectional research on the relationship between stress and T2D are inconsistent. Long-term research, however, indicates a strong connection between the developments of diabetes. There is consistent evidence that those who have symptoms of anxiety or depression are more likely to develop type 2 diabetes (T2D), and this association may be reciprocal (Kolb & Martin, 2017).

1.1.3.9 Sleep pattern:

The risk of being diagnosed with type 2 diabetes (T2D) increases by approximately 20–39% for people exposed to higher noise levels or more fine particulate matter over 10 years or who live near busy roads, according to epidemiological studies (Kolb & Martin, 2017). Sleep duration and quality are contributing factors, with nighttime noise or light exposure causing disturbances. Shorter sleep duration is linked to increased diabetes risk, with each hour shorter associated with a 9% higher risk (Kolb & Martin, 2017).

1.2 Diabetic complications:

Diabetes is a chronic illness that lowers a person's quality of life and has a high death rate, which raises the possibility of complications from the disease. Cardiovascular disorders, diabetic foot ulcers, neuropathy, nephropathy, and retinopathy are among the major consequences of diabetes.

Hyperglycemia is the primary cause of these problems since it destroys blood vessels and neurons (Banday et al., 2020). Diabetes can cause complications in many organ systems, including retinopathy, which can lead to blindness, nephropathy, which can cause renal failure, hypertension, coronary heart disease, and peripheral and autonomic neuropathy, which can impair nerves (Banday et al., 2020). Moreover, diabetes raises the risk of coronary heart disease, peripheral artery disease, and cerebrovascular disease—a condition known as atherosclerotic cardiovascular disease—significantly, which adds to the high rates of morbidity and death linked to diabetes.

1.2.1 Diabetic nephropathy:

Thirty to forty percent of people with diabetes have diabetic kidney disease (DKD), which increases the risk of end-stage renal disease (ESRD) and death (Demir et al., 2021). Notwithstanding strict blood glucose management, lipotoxicity and oxidative stress play a role in its development, leading to structural alterations in the kidney. The molecular pathways encompass elements such as abnormal lipid signaling and TGFβ1 [12]. Novel therapeutic approaches focus on active lipids and compounds such as VEGF-B, in addition to SGLT2 inhibitors that mitigate renal fibrosis and inflammation (Chen et al., 2011). Results are improved by therapeutic advancements such SGLT2 inhibitors and RAAS inhibition (Selby & Taal, 2020). The course of DN varies, and albuminuria may return. Hyperfiltration and albuminuria are the first signs of DN, a primary cause of ESRD, which is followed by a loss in renal function. Patients with DKD have a significant mortality rate, primarily from cardiovascular illness. Controlling hemodynamic and metabolic abnormalities is essential to stopping the course of DKD (Sagoo & Gnudi, 2020). Diabetic nephropathy emphasizes falling eGFR or persistent albuminuria. It is essential to screen for microalbuminuria, and strict glucose and pressure management is advised. Nonproteinuric diabetic kidney disease presents difficulties in diagnosis and treatment, necessitating new strategies (Samsu, 2021).

1.2.2 Diabetic retinopathy:

Diabetic retinopathy (DR), a common consequence of both type 1 (T1DM) and type 2 (T2DM) diabetes, is characterized by damage to the neurons and blood vessels of the retina, which results in loss of vision. Globally, its prevalence is growing as the incidence of diabetes rises (Milluzzo et al., 2021). Obesity, hypertension, and hyperglycemia are modifiable risk factors; genetic susceptibility and the duration of diabetes are non-modifiable (Milluzzo et al., 2021). Early

treatment greatly lowers the chance of severe visual loss, and screening programs have contributed to a decrease in the prevalence of DR (Milluzzo et al., 2021). On the other hand, DR pathogenesis is facilitated by a complex gene-environment interaction that involves epigenetic processes (Milluzzo et al., 2021). To guarantee efficient management, efforts are being made to maximize screening frequency through the use of prediction models (van der Heijden et al., 2020). While stem cell therapy holds promise for vascular regeneration, therapeutic methods that target inflammatory cytokines and vascular endothelial growth factor are effective (Marques et al., 2021). Understanding the molecular and cellular pathology of DR is advancing rapidly, providing insights for targeted therapeutic interventions (Antonetti et al., 2021).

1.2.3 Diabetic cardiomyopathy:

Diabetes mellitus (DM) and diabetic cardiomyopathy (DC) are closely related conditions that ultimately result in heart failure (HF). Although cardiac fibrosis and hypertrophy were associated with the old concept, new research highlights the significant impact of reduced cardiomyocyte contraction caused by protein alterations, opening up new therapeutic options (Tan et al., 2020). According to Dillmann, ischemic cardiomyopathy and coronary artery disease are two major causes of HF, which is common in DM patients and further complicates the clinical picture (Dillmann, 2019). Diagnoses of DC can be difficult because the condition was first recognized in 1972 and is characterized by HF without the usual cardiac comorbidities. However, recent research has identified a unique phenotype associated with diastolic dysfunction and left ventricular hypertrophy that significantly affects the long-term prognosis and quality of life in diabetes patients (Lorenzo-Almorós et al., 2022).

Diabetes and heart failure have complicated interactions, with each amplification of the other. Although a conclusive diagnosis usually necessitates invasive procedures like endomyocardial biopsy, which are rarely used in clinical practice, diagnosis entails detecting indicators of HF while ruling out other heart disorders (Dillmann, 2019). Although there aren't any specific treatments for DC at this time, therapeutic approaches that focus on HF and diabetes show potential. Experimental data points to possible uses for several diabetic drugs in the treatment of DCM; however, additional studies involving DCM patients must be conducted in order to confirm these results and create efficient treatment plans (Paolillo et al., 2019).

1.3 Glycemic control:

Glycemic control refers to maintaining the blood glucose levels within the optimum range. According to the World Health Organization, just 36% of diabetic individuals have their blood sugar levels under control (Abd-Elraouf, 2020). The risk of complications from diabetes, including kidney disease, nerve damage, eye issues, and cardiovascular disease, is decreased when glycemic control is maintained. Additionally, it improves diabetic patients' mood and gives them more energy for daily tasks.

1.3.1 Diabetes self-management and glycemic control:

The key to managing diabetes effectively is keeping blood glucose levels within the recommended range, which can be achieved mainly via self-management and medication compliance. Self-management involves activities like exercise, diet, medication adherence, glucose monitoring, and foot care, as well as emotional, role, and medical components. Studies regularly demonstrate that interventions enhance these behaviors in at least one way (Almutairi et al., 2020).

Due to the necessity of juggling several activities, including insulin or medication administration, exercise, food, and routine glucose testing, self-management can be difficult (Al-Khawaldeh et al., 2012; Martínez et al., 2016). Inadequate information and motivation are two obstacles to effective self-management, underscoring the significance of diabetes education initiatives in improving self-management practices and quality of life (Li et al., 2020). In order to enhance glycemic control and lower complications, programs such as DESMOND (Diabetes Education and Self-Management for Ongoing and Newly Diagnosed) offer vital information on nutrition, exercise, and lifestyle modifications (Ahola & Groop, 2013).

To sum up, self-management is an essential part of diabetes therapy and calls for thorough instruction and assistance to handle the different obstacles that patients have when trying to manage their illness.

1.3.2 Factors influencing glycemic control:

Comprehending the complex interplay of variables impacting glucose control is imperative for proficient management of diabetes. Every factor, from socioeconomic status to lifestyle decisions,

has a substantial impact on glycemic outcomes, emphasizing the need for integrated approaches to diabetes care.

1.3.2.1 Physical activity:

The study in 2020 found strong correlations with a number of variables, such as the length of diabetes, regular exercise, anthropometric measurements, and abnormalities in the lipid profile (Abd-Elraouf, 2020). Significantly, worse glycemic control was associated with longer duration of diabetes, perhaps as a result of increasing impairment of insulin secretion. In line with other research connecting physical activity to better insulin sensitivity and lower fasting blood sugar levels, irregular exercise was also linked to poor management.

1.3.2.2 Physiological factors:

Higher HbA1c levels were also significantly predicted by anthropometric measures like waist circumference and BMI, as well as by anomalies in the lipid profile, specifically elevated levels of total and LDL cholesterol (Abd-Elraouf, 2020). These results highlight how crucial it is to address these variables in the management of diabetes in order to enhance glycemic control and general health outcomes.

1.3.2.3 Environmental factors and their influence on sleep pattern and depression:

There is a complex interplay among sleep disorders, environmental variables, and glycemic management in people with type 2 diabetes mellitus (T2DM). Various factors influences glycemic management, including light exposure, noise pollution, sleep disruptions, and community inequities (Afroz-Hossain et al., 2019). For example, irregular sleep schedules and artificial light exposure at night can throw off circadian rhythms, which can result in poor glycemic control and insulin resistance. Further complicating diabetes management are neighborhood inequities, which include things like food insecurity and crime rates that affect patients' ability to stick to behavioral adjustments (Afroz-Hossain et al., 2019).

1.3.2.4 Age, ethinicity and socio-economic factors:

Glycemic control is influenced by a number of factors, including age, race, socioeconomic status, clinical features, and patient-centered variables (Cheng et al., 2019). The association between glycemic control and socioeconomic status is mediated by variables such as health habits, service

quality, and personal traits (Houle et al., 2016). Due to financial hardship and restricted access to high-quality care, patients with lower socioeconomic level may find it difficult to control their diabetes, which can have a negative impact on glycemic outcomes (Houle et al., 2016). Improving the management of diabetes and its consequences requires addressing these socioeconomic inequalities. As a result, specific diabetes management techniques are required.

1.3.2.5 Dietary habits:

A number of variables affecting diabetic patients' eating patterns. It implies that by addressing obstacles associated with dietary therapy, nutrition education and counseling can have a good effect on blood glucose control (Jun et al., 2016). Patients who received nutrition instruction, for example, demonstrated improved adherence to dietary guidelines as seen by higher blood glucose influence behavior scores.

1.3.3 Markers of glycemic control:

The importance of glycemic control is emphasized in the management of diabetes, with particular attention to monitoring techniques such hemoglobin A1c (HbA1c) laboratory tests and self-monitoring of blood glucose (SMBG) (Krhač & Lovrenčić, 2019). While HbA1c is a common indicator of average glycemia over months, its limitations force researchers to investigate other biomarkers for a more thorough evaluation of glycemic control, such as fructosamine, glycated albumin, and 1,5-anhydroglucitol, in addition to data from continuous glucose monitoring systems (CGMS) (Krhač & Lovrenčić, 2019). Self-monitoring of blood glucose (SMBG) is still crucial for daily management, although it doesn't offer thorough details on transient variations. Precise assessment of glucose fluctuation is necessary, ideally by regular glucose testing. CGM is a potential answer, but there are still issues with data assessment and standardization. A mix of glycemic indicators with shorter and longer durations increases the risk prediction of diabetes (Kohnert et al., 2015).

1.3.4 Glycemic control assessment using the HbA1c level as the gold standard:

Glycated hemoglobin, or HbA1c, is a critical biomarker for evaluating diabetes patients' long-term blood sugar control (Krhač & Lovrenčić, 2019). It is frequently used to track glycemic control since it shows average blood sugar levels over a period of one to three months (Krhač & Lovrenčić, 2019). Despite its importance, methodological variability has made accurate HbA1c measurement

difficult (Krhač & Lovrenčić, 2019). In order to improve clinical accuracy in the management of diabetes, standardization initiatives are still ongoing (Krhač & Lovrenčić, 2019). Numerous studies have demonstrated the correlation between a diabetic's HbA1c level and quality of life, with every 1% decrease in HbA1c level corresponding to a 5% improvement in life quality (Abd-Elraouf, 2020). Glycosylated hemoglobin (HbA1c) measurement is now recognized as the most trustworthy marker for long-term glycemic management since it properly represents a patient's blood glucose levels over the preceding two to three months as well as the standard of diabetes care that the general public receives (Abd-Elraouf, 2020).

Higher risks of diabetes-related morbidity and mortality are associated with elevated HbA1c levels (Almutairi et al., 2020). According to seven studies comparing the intervention and control groups, the intervention group's HbA1c was significantly lower (Almutairi et al., 2020). Furthermore, every intervention showed an improvement in HbA1c in the intervention group, while the control group showed various degrees of improvement or minor increases among studies (Almutairi et al., 2020).

The studies with the highest reductions were those in which the mean baseline HbA1c was greater than 10% (Almutairi et al., 2020). Interestingly, just one study showed sustained but statistically insignificant reductions in HbA1c after six months, and that trial continued to a 24-month follow-up (Almutairi et al., 2020).

Diabetes-related hazards, such as mortality and microvascular complications, may be reduced by 21% with every 1% drop in mean HbA1c (Almutairi et al., 2020). Benhalima et al. (2011) have highlighted that recent trials such as ACCORD, ADVANCE, and VADT have cast doubt on the advantages of strict control, especially for individuals with advanced Type 2 diabetes and heightened cardiovascular susceptibility [33].

In conclusion, new research indicates the need for tailored treatment targets, taking into account variables like life expectancy, cardiovascular risk, and comorbidities, even though HbA1c is still a key tool for evaluating glycemic control and predicting diabetes-related hazards (Benhalima et al., 2011). Furthermore, continued research into substitute markers for improved glucose control emphasizes how diabetes care is changing (Kohnert et al., 2015).

1.4 Genetic basis of HbA1c level:

Hemoglobin A1c (HbA1c) levels indicates average blood glucose levels over the previous two to three months that are impacted by a mixture of environmental, genetic and epigenetic factors. Lifestyle variables including nutrition, exercise, stress, and medication adherence are also important in affecting HbA1c levels. Environmental factors have the ability to either magnify or lessen the impact of genetic predispositions, even though hereditary factors still offer a predisposition. Histone alterations and DNA methylation are epigenetic modifications that can impact HbA1c levels by influencing gene expression. Lifestyle factors like stress, exercise, and diet can affect these changes.

The genetic influences on hemoglobin A1c (HbA1c) levels, a critical marker for the diagnosis and treatment of type 2 diabetic mellitus (T2D), have been highlighted by recent studies. Numerous genetic loci, including the FTO and GCK genes, have been identified as being associated with HbA1c by genome-wide association studies (GWAS) (Leong & Wheeler, 2018). Unaffected by BMI, the FTO gene, which is linked to obesity, affects HbA1c levels via glycemic pathways (Leong & Wheeler, 2018). Conversely, glucokinase, an essential enzyme in glucose metabolism, is encoded by the GCK gene and influences HbA1c levels directly through glycemic pathways (Leong & Wheeler, 2018).

Ten genetic loci were shown to be strongly correlated with glycated hemoglobin A1c (HbA1c) in a recent genome-wide association study (GWAS). They included four known loci (HK1, MTNR1B, GCK, and G6PC2/ABCB11) and six new loci (FN3K, HFE, TMPRSS6, ANK1, SPTA1, and ATP11A/TUBGCP3) (Soranzo, 2011). Three loci were identified as being connected to hyperglycemia: GCK, G6PC2, and MTNR1B (Soranzo, 2011). It appears from the study that hyperglycemia has a role in influencing these genetic relationships with HbA1c levels (Soranzo, 2011). A common mutation at any of the seven non-glycemic loci may impact HbA1c levels through erythrocyte biology, as these loci are associated with iron storage problems and hereditary anemias.

Furthermore, rs1039215, a low-frequency mutation unique to African ancestry, was found to have a significant impact on HbA1c levels in a study that analyzed genetic data across multiple ancestries (Sarnowski et al., 2019). Furthermore, a considerable amount of the variance in HbA1c

in African-Americans and Hispanics can be explained by a mutation in the G6PD gene (rs1050828), which significantly lowers HbA1c levels (Sarnowski et al., 2019). SNP variant rs76723693, another uncommon G6PD coding variant, is linked to lower HbA1c levels and has consistent effects on various cohorts of African-American people (Sarnowski et al., 2019).

These results highlight how hereditary factors—specifically, variations in genes like FTO and GCK—have a significant impact on HbA1c levels, which in turn affects T2D diagnosis and treatment. In order to provide more precise diabetes diagnosis and efficient treatment plans catered to each patient's unique genetic profile, it is imperative to comprehend these genetic impacts.

Knowing the genetic foundation of type 2 diabetes HbA1c values can assist identify patients who are more likely to develop problems and create individualized treatment plans that maximize blood glucose control. But it's important to understand that genetics is only one aspect of the picture; medication and lifestyle changes are also vital for properly treating diabetes.

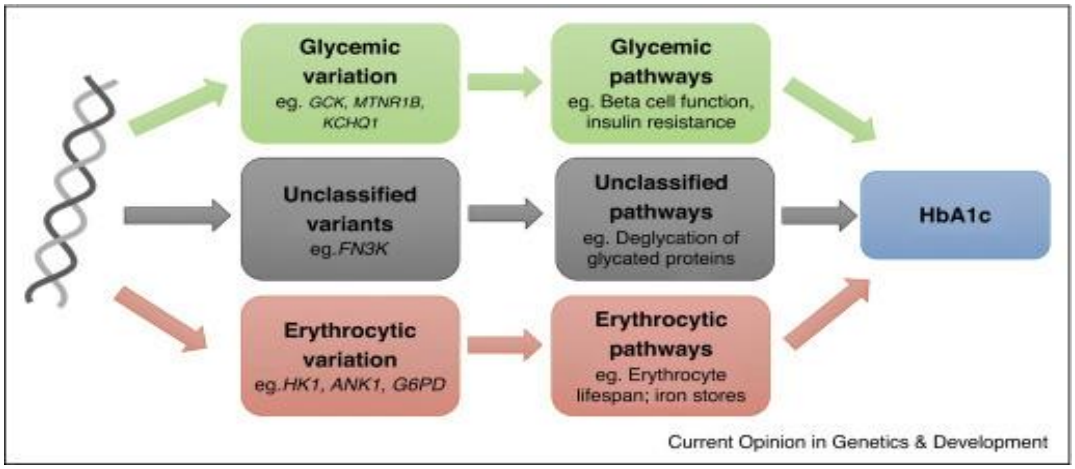


Figure 1.0.3: Pathways influencing HbA1c level. (Leong & Wheeler, 2018)

1.5 Machine learning in disease management:

Machine learning and algorithms assists medical professionals in the diagnosis and prediction of diseases, machine learning techniques can be widely used to the solving of problems in the medical arena. However, because medical data and information are jumbled, disorganized, and high dimensional, it is difficult to extract knowledge and information from them (Singh et al., 2019). Additionally, there are outliers and noise in the data that was gathered for this report. The primary

useful approach will be examining several machine learning strategies. By confirming and validating the correctness of machine learning approaches, the effectiveness of these methods is assessed. The usefulness and applicability of various machine learning techniques, such as the decision tree algorithm, support vector machine method, random forest method, evolutionary algorithm-based models, and swarm intelligence-based techniques, in the diagnosis and treatment of diseases (Singh et al., 2019).

1.5.1 Applications of machine learning in diabetes self-management:

Particularly when it comes to diabetes, artificial intelligence (AI) has become a ground-breaking instrument in the management of disease. AI transforms healthcare by utilizing clinical and genetic data to tailor medical interventions through the integration of machine learning (ML) algorithms and prediction models (Li et al., 2020). Diagnose, treat, and manage diabetes with the help of trained models and AI-driven clinical decision support systems (CDSS) (Amann et al., 2020). By analyzing patient data, machine learning algorithms provide personalized and predictive skills for managing diabetes (Chaki et al., 2022).

Artificial intelligence (AI) is important for managing diabetes because it can improve clinical decision support, enable continuous patient monitoring, expedite healthcare administration, and offer individualized interventions (Reddy et al., 2019). Compared to traditional methods, these AI-driven approaches have many benefits, such as easy deployment, low cost, high efficiency, and broad coverage (Li et al., 2020). Nevertheless, issues like patient involvement and implementation in clinical settings still need to be resolved (Su et al., 2019).

Through the use of technology, AI in diabetes self-management enables more informed decision-making and improved disease management (Su et al., 2019). Research shows that AI-driven therapies, such as telemedicine and machine learning algorithms, are superior to traditional techniques for enhancing glycemic control (Taylor, 2013). Lower HbA1c levels are correlated with patient activation and use of remote monitoring devices, highlighting the significance of focused interventions in the management of diabetes (Su et al., 2019).

By providing individualized support, advice on meal planning and exercise, and resolution of issues regarding blood glucose levels and complications, artificial intelligence natural language models, such as ChatGPT, help people with diabetes manage their condition (Farajollahi et al.,

2021). Furthermore, glycemic control can only be achieved through self-management behavior and medication adherence, which are highlighted by longitudinal research (Zheng et al., 2018).

Artificial intelligence (AI) algorithms have the potential to significantly improve glycemic control and self-management techniques in remote areas with limited access to healthcare (Ansari et al., 2023). AI revolutionizes disease management by utilizing ML and predictive models to provide patients with diabetes with individualized, effective, and easily accessible care.

1.6 Objective:

1. To select features for model by analyzing the effect of different diabetes management variables on blood glycemic level in Pakistani patient data.
2. To build and validate machine learning model to predict level of glycemic control which serves as a measure of diabetes self-management.
3. Validate the prediction of HbA1c levels using a machine learning model based on gene expression analysis.

LITERATURE REVIEW

2.1 Type 2 diabetes and current approaches of glycemic control:

Diabetes mellitus, especially Type 2, is a major global health issue that affects healthcare resources, morbidity, and death. Maintaining optimal glucose regulation is essential for controlling diabetes and reducing the chance of complications. But maintaining and attaining glycemic control is still a complex issue driven by a wide range of variables.

A study by on veterans with poorly controlled diabetes revealed a gap in disease self-management even in the presence of competent medical care. Their results highlighted how important self-efficacy programs are for raising glycemic control and quality of life in diabetes patients. They emphasized how crucial it is to comprehend elements like self-efficacy and openness to change in order to improve diabetic self-management techniques (Nelson et al., 2007).

A thorough analysis of glycemic control in Shanghai among patients with type 2 diabetes was conducted, providing insight into a number of variables impacting control results. Their research demonstrated the effects of waist-to-hip ratio, medical nutrition therapy, and disease duration on the state of glycemic control. They underlined how important it is to take into account a variety of factors in order to attain the best possible glycemic control in diabetic population (Yang, n.d.).

A cross-sectional study in Sri Lanka to investigates the variables affecting adults with diabetes' ability to control their blood sugar levels. Their data revealed participants' inadequate glucose control despite their claimed adherence to medication, food restriction, and exercise. They underlined the necessity for specialized interventions to overcome obstacles to efficient diabetes control and underlined the complexity of self-management practices (Amarasekara et al., 2015).

Abd-Elraouf studied glycemic control in Anhui, China, and found that patients with type 2 diabetes mellitus had a low prevalence of excellent control. Their investigation revealed that glycemic control outcomes were significantly influenced by variables like gender, education level, and length of illness. In order to enhance control outcomes, they emphasized how crucial it is to address socioeconomic issues in diabetes care regimens (Abd-Elraouf, 2020).

The connection between glycemic control and a healthy lifestyle score was investigated in Chinese patients with type 2 diabetes mellitus. Their results made clear how important it is to integrate several healthy lifestyle choices in order to achieve and sustain appropriate glycemic control. To improve the results of diabetes care, they underlined the need of encouraging healthy behaviors (Xing et al., 2022).

This investigation exploring the connection between a healthy lifestyle score and glycemic control among Chinese patients with type 2 diabetes mellitus was expanded by Che et al. Their research highlighted the significance of health-related activities in glycemic control outcomes, including self-management and coping mechanisms. They emphasized the necessity of all-encompassing methods that take lifestyle factors into account while developing diabetes care plans (Che et al., 2023).

Houle et al looked at how socioeconomic status affected glycemic control in persons with type 2 diabetes mellitus in Canada. Their research clarified the complex interactions that exist between glycemic control results, health behaviors, and socioeconomic determinants. They emphasized how crucial it is for diabetes management programs to take socioeconomic variables into account (Houle et al., 2016).

Pamungkas et al concentrated on how poor sleep quality affects type 2 diabetes mellitus patients in India's glycemic control. Their research showed how important it is for diabetes management plans to take a holistic approach to treating sleep-related problems. They emphasized how important it is for diabetic treatment regimens to take sleep quality into account (Pamungkas et al., 2017).

A hospital-based study in India and found that among patients with problems, poor glycemic control was significantly correlated with a number of clinical and demographic variables. Their results highlighted the significance of individualized therapies based on individual characteristics and offered insights into the factors impacting glycemic control outcomes in this population (Kodakandla et al., n.d.).

Haghighatpanah et al looked into other variables associated with inadequate glycemic control in individuals with problems from type 2 diabetes mellitus. Their research demonstrated the significance of several clinical and demographic variables, including age, gender, the length of diabetes, and the kind of medicine taken, in affecting the results of glycemic control. They emphasized how important it is to take these aspects into account when developing diabetes care plans in order to enhance control and lessen the burden of complications (Haghighatpanah et al., 2018).

Together, these studies highlight the complex and multifaceted relationship between glycemic control and type 2 diabetes mellitus, highlighting the role of lifestyle choices, comorbidities, socioeconomic factors, and self-management practices. In order to improve diabetes outcomes and lessen the burden of complications on afflicted populations, it is imperative to comprehend and treat these various aspects.

2.2 Role of self-management in glycemic control of type 2 diabetes:

The only way to survive with diabetes is to control and manage blood glucose levels within the ideal range. Diabetes is an incurable disease. The management of diabetes is the cornerstone of its control. Medication adherence and self-management are the two primary facets of diabetes care. Since medication adherence is insufficient for managing diabetes on its own, it is actually a component of self-management. The degree to which individuals take their medications at the prescribed dose and time is known as medication adherence. Poor glycemic control is primarily caused by persons who adhere to medication regimens but neglect other self-management practices, as has been seen (Ansari et al., 2023).

2.2.1 Components of self-management for type 2 diabetes:

For those with long-term illnesses like diabetes, self-management includes emotional, role, and medical control. Improved glycemic management is associated with behaviors like exercise, diet, medication adherence, blood glucose monitoring, and foot care. Even though different self-management behaviors are evaluated in separate research, treatments always result in at least one behavior improvement (Almutairi et al., 2020).

Maintaining glucose levels within the ideal range and lowering the risk of complications from diabetes are the two main objectives of self-management. Diabetes is challenging to self-manage since it necessitates juggling a number of activities, including consistent glucose testing, regular exercise, a nutritious diet, and the right amount of medication or insulin (Al-Khawaldeh et al., 2012). Because type 2 diabetes involves integration and commitment to various behaviors related to nutrition, exercise, medication, and insulin, management of the condition is complicated and challenging (Martínez et al., 2016).

2.2.2 Challenges and barriers to self-management in type 2 diabetes:

Research indicates that interventions significantly improve blood glucose monitoring, physical activity, nutrition, and foot care. The goal of "patient activation," a technique that allows patients to participate in their own care, is to provide people the knowledge, abilities, and self-assurance they need. Patients who are more engaged in their care are more likely to adopt healthy habits and have better results. The goal of patient activation interventions is to raise activation levels as determined by the Patient Activation Measure (PAM) (Almutairi et al., 2020).

The quality of self-management depends on the patient's knowledge and the patient's motivation to improve his behavior, but patients normally do not adhere to self-management plans (Ansari et al., 2023). Thus, it is important to promote self-management behavior and recommend personalized activities to maintain blood glucose levels. Conventionally, diabetes education programs were launched for better diabetes management through medication adherence and lifestyle changes. One of the obstacles to diabetes self-management is inadequate knowledge. It was shown that diabetes education plays an important role in enhancing the quality of life by improving self-management behavior in diabetics (Li et al., 2020). DESMOND (Diabetes Education and Self-Management for Ongoing and Newly Diagnosed) is an education program for people with type 2 diabetes that provides knowledge about diet habits, exercise, and other lifestyle changes for improvement of glycemic control and about how diabetics can reduce the risk of diabetic complication (Ahola & Groop, 2013).

2.3 HbA1c level as a measure of diabetes self-management quality:

Diabetes mellitus (DM), which affects millions of people globally, poses a serious threat to global health. A multimodal strategy is necessary for the effective management of diabetes, with glucose control serving as a fundamental component of care. Hemoglobin A1c (HbA1c) values are frequently used to inform treatment choices and are an essential indicator for evaluating long-term glycemic management. In light of research examining the effects of HbA1c awareness, diabetes education, self-management interventions, and physical activity on glycemic control and quality of life in people with type 2 diabetes mellitus (T2DM), we examine the significance of knowing one's HbA1c levels in diabetes self-management in this section.

It is essential to comprehend one's HbA1c values in order to effectively treat diabetes. In a research 686 persons with type 2 diabetes were included (Heisler et al., 2005). Of them, only 25% correctly reported their most recent HbA1c result, and 66% were not aware of it. HbA1c awareness was found to be correlated with education level and good provider communication. People who knew their HbA1c values tended to have greater understanding of diabetes management and to evaluate their diabetes control more accurately. Nevertheless, neither self-efficacy nor self-management actions increased as a result of this understanding. The study highlights that additional techniques to improve motivation and support are required for optimal diabetes care; simply presenting information on HbA1c levels may not be adequate (Heisler et al., 2005).

Globally, a number of interventions, including diabetes self-management education/support (DSME/S) programs, have been put into place to help people with type 2 diabetes mellitus (T2DM) improve their glycemic control. A meta-analysis demonstrated that DSME/S treatments significantly lowered HbA1c levels, particularly in individuals whose baseline HbA1c was greater than 7.5% (Bekele et al., 2021). These therapies also decreased participants' emergency visits and increased their sense of self-efficacy. It has been discovered that group-based, lay-led self-management interventions can significantly lower healthcare utilization and enhance glycemic control in diabetics (Bekele et al., 2021; Tay et al., 2021).

Nevertheless, different populations may respond differently to DSME/S. For example, no statistically significant difference in HbA1c levels was observed between the usual care and

intervention groups in a trial that focused on African-American adults with T2DM. In spite of this, participants' quality of life significantly improved after receiving DSME (Cunningham et al., 2018). Additionally, it was discovered that patient activity and self-management practices were related to patient recollection of HbA1c readings. Recalling HbA1c values may improve health-promoting behaviors, but obstacles like low educational attainment may reduce the motivating power of this information (Cunningham et al., 2018; Willaing et al., 2013).

Glycemic regulation is significantly aided by physical activity in addition to knowledge and self-management techniques. A meta-analysis revealed a correlation between moderate improvements in physical activity and lower levels of HbA1c and fasting glucose (Boniol et al., 2017). Furthermore, in certain populations, such Hispanic/Latino people with or at risk of T2DM, the time and intensity of physical activity may affect HbA1c levels and body mass index (Kerr et al., 2024; Utinane & Tīcmane, n.d.).

All things considered, controlling and comprehending HbA1c readings are crucial elements of successful diabetes self-management. Interventions emphasizing physical exercise, education, and self-management support can greatly enhance the quality of life and glycemic control of people with diabetes. To address current issues in diabetes care and customize interventions for particular populations, more research is necessary.

2.4 Introduction to Artificial intelligence and Machine learning in Healthcare:

Healthcare is one of the many industries that artificial intelligence (AI) and machine learning (ML) have changed. Artificial Intelligence (AI) is the imitation of human intelligence in machines, allowing them to do tasks like speech recognition, visual perception, and decision making. As a branch of artificial intelligence, machine learning focuses on teaching algorithms to learn from data and make predictions or judgments based on it. Because these technologies can analyze enormous volumes of data and spot trends that human clinicians would miss, they have a particularly significant impact on the healthcare industry (Ellahham, 2020).

AI and ML have become extremely useful tools in the treatment of diabetes, providing predictive models for managing complications and determining diabetes risk. By utilizing large datasets, these technologies improve diagnostic precision, advance predictive modeling, and customize

patient treatment, thereby revolutionizing the field of healthcare (Fan et al., 2021). In order to forecast and treat problems including retinopathy, nephropathy, and cardiovascular illnesses, machine learning algorithms examine large datasets from individuals with diabetes (Ellahham, 2020; Fan et al., 2021; Kasula, 2023).

2.4.1 Machine learning models in diabetes care and glycemic control:

In order to improve diabetes treatment and control, machine learning models are being applied more and more. These algorithms help with early diagnosis, individualized treatment strategies, and improved glycemic control by analyzing massive datasets to find trends and forecast outcomes. Machine learning models offer insights that help healthcare professionals make wise decisions and improve patient care (Dankwa-Mullan et al., 2019).

2.4.1.1 Diabetes risk prediction:

To identify people who are at high risk of acquiring diabetes, machine learning algorithms assess genetic, metabolic, and lifestyle data (Ellahham, 2020). This allows for targeted therapies and preventive measures.

2.4.1.2 Automated screening of complications:

Numerous machine learning algorithms, including Random Forest, Support Vector Machines, Gradient Boosting, and Neural Networks, show promise in anticipating issues like retinopathy and dividing up the risk of cardiovascular illnesses into different categories (Kasula, 2023). By evaluating retinal scans and giving precise diagnoses, AI systems like IDx-DR identify diabetic retinopathy early on, relieving ophthalmologists of their workload and guaranteeing prompt treatment (Ellahham, 2020). Artificial intelligence (AI) significantly aids in diabetes risk assessment by utilizing noninvasive, reliable predictive models and machine learning algorithms that analyze vast datasets and numerous risk factors effectively. These models have demonstrated high accuracy in identifying individuals at high risk of diabetes, allowing for early intervention and improved management (Liu, 2020). AI also enhances diabetes diagnosis through efficient, noninvasive methods, outperforming traditional diagnostic approaches (Liu, 2020).

2.4.1.3 Patient Self-Management:

Patients can better control their diabetes by monitoring their blood glucose levels, getting individualized guidance, and using wearable technology and smartphone applications with AI capabilities (Ellahham, 2020). AI systems help medical professionals by quickly analyzing patient data and providing treatment and follow-up care suggestions. This improves decision-making effectiveness and makes the best use of available resources (Ellahham, 2020).

In T2D management, machine learning algorithms are used to evaluate patient data and forecast results. Demographics, test findings, illness features, drugs, and economics are examples of input variables. Nephropathy, angiopathy, eye disease, and neuropathy are examples of consequences, and poor glycemic management is a major cause of these. Predictive models are constructed using a variety of machine learning algorithms, including as neural networks and Bayesian networks (Fan et al., 2021).

2.4.2 Current Predictive Models for Diabetes Care and Glycemic Control:

Predictive models that analyze data from electronic health records (EHRs) and other sources use machine learning algorithms to control diabetes and associated consequences. For example, employed machine learning (ML) algorithms to forecast unfavorable outcomes in patients with non-adherent Type 2 Diabetes (T2D), with a particular emphasis on poor glycemic control and comorbidities including nephropathy and eye disease (Fan et al., 2021). Machine learning (ML) models are useful for risk stratification and crucial predictor identification in diabetes management. They also provide hypothesis-free analysis and improve our understanding of predictive factors in diabetes treatment. Recent studies demonstrate the application of machine learning (ML) models in the management of diabetes, finding important predictors and facilitating the early identification of diabetic nephropathy (Kasula, 2023). Validation studies have confirmed the accuracy of predictive models that use patient data and biomarkers to accurately determine the probability of acquiring type 2 diabetes (Abbasi et al., 2012). Furthermore, baseline glycemic status is a significant predictor of glycemic control in type 2 diabetes patients receiving combination medication (Del Parigi et al., 2019). The Bayesian network had the best prediction performance for HbA1c values. The amount of hypoglycemic medications and types of insulin taken were important indicators of poor glycemic control. The study indicates that improving

glucose monitoring and medication adherence can postpone T2D consequences, despite its small-sample, single-center design (Fan et al., 2021). Another study found predictors of T2D patients' achievement and maintenance of a HbA1c target of ≤7% by analyzing clinical trial data with machine learning. The results of the investigation showed that fasting plasma glucose (FPG) and baseline HbA1c were the best indicators of glycemic management. In particular, by using HbA1c as a crucial metric, recent studies highlight the vital role that AI and predictive models play in diabetes self-management and glycemic control (Wan, 2021). With an emphasis on early identification and individualized therapies, these models improve diabetes prediction and management by including a variety of patient data. The KMAP-O model, for instance, emphasizes the need for dependable tools and long-term observation to comprehend the dose-response relationship between interventions and outcomes. It blends knowledge, motivation, attitude, and preventative actions to optimize diabetes care (Wan, 2021). Through automated retinal screening and clinical decision support systems, for example, AI applications like machine learning and neural networks further increase predictive accuracy for diabetes complications (Dankwa-Mullan et al., 2019; Ellahham, 2020). Research also shows that mobile health interventions can improve HbA1c levels and self-management practices. This shows the promise of current communication technologies and ongoing assistance (Riangkam et al., 2021). Overall, by improving patient participation and streamlining treatment plans, these developments in AI and predictive analytics are revolutionizing the way that diabetes care is provided (Cichosz et al., 2015).

2.4.2.1 Important variables found in these models:

- Duration of Diabetes: Patient's diabetes duration is a major predictor of complications and inadequate glycemic control (Fan et al., 2021).
- Types of Insulin Used: Glycemic control and the likelihood of unfavorable outcomes are influenced by the types and dosages of insulin used (Fan et al., 2021).
- Number of Hypoglycemic Drugs: A patient's use of hypoglycemic medications plays a significant role in controlling their blood sugar levels and averting consequences (Fan et al., 2021).

2.4.3 Assessment of Current Model Accuracy and Performance:

A number of important indicators are commonly used to assess the effectiveness of machine learning models in the treatment of diabetes:

- AUC-ROC: The model's capacity to distinguish between positive and negative cases is measured by the Area Under the Receiver Operating Characteristic Curve (AUC-ROC).
- Accuracy: The percentage of true findings (true positives and true negatives) in relation to all cases analyzed.
- Sensitivity (Recall): The model's accuracy in recognizing positive cases.
- Specificity: The model's capacity to recognize negative situations with accuracy (Jahani & Mahdavi, 2016).

2.4.4 Model Comparisons:

Comparative research demonstrates that sophisticated machine learning strategies frequently outperform conventional statistical approaches:

Gradient Boosting Machine (GBM) and Logistic Regression (LR): Lai et al. showed that both models had good predictive accuracy; GBM's AUC was 84.7%, while logistic regression's was 84.0% (Lai et al., 2019).

Neural Networks augmented with Memetic Algorithms: Neural network models augmented with memetic algorithms outperformed conventional logistic regression models, achieving an accuracy of 93.2% (Jahani & Mahdavi, 2016).

Promising improvements in glycemic control, tailored treatment, and early diagnosis can be achieved by integrating AI and ML into diabetes care. In order to improve patient outcomes and save healthcare costs, it is imperative that this field sees ongoing research and development to improve the accuracy and reliability of predictive models. The management of diabetes and other chronic diseases will probably depend more and more on these technologies as they develop.

2.5 Challenges and considerations in implementation of predictive models for diabetes care:

AI-driven diabetes care implementation is fraught with difficulties. Because dose errors have serious repercussions for diabetics, it is imperative to ensure the safety and dependability of AI algorithms. This calls for stringent validation and regulatory procedures (Diaz C et al., 2023; Juneja et al., 2022). AI systems become more complex as a result of individual diversity in response brought on by genetics, lifestyle, and coexisting medical conditions. Individual needs must be met while simultaneously managing data security and quality issues (Diaz C et al., 2023; Juneja et al., 2022). Interoperability with the current healthcare infrastructure is crucial, and AI systems must take into consideration outside variables like stress and disease that affect blood glucose levels (Domingo-Lopez et al., 2022). Successful integration also requires resolving ethical issues including data protection and algorithmic bias, guaranteeing affordability, building trust with patients and healthcare professionals via openness, and offering the required education and training (Deepa & Sivasamy, 2023; Iparraguirre-Villanueva et al., 2023). For these systems to be deployed effectively and to be improved over time, academics, regulators, and technology developers must work together continuously (Deepa & Sivasamy, 2023; Iparraguirre-Villanueva et al., 2023).

2.5.1 Ethical considerations in implementing predictive models for diabetes care:

Informed permission, maintaining transparency in AI decision-making processes, and protecting patient privacy and data security are ethical considerations in AI-driven diabetic treatment (Akinci D'Antonoli, 2020; Murdoch, 2021). It's also critical to address any biases in training data, make sure AI upholds rather than compromises patient autonomy, and have a "human-in-the-loop" mindset (Naik et al., 2022). In order to prevent healthcare inequities and adhere to all pertinent rules, developers must guarantee accessibility for a varied range of individuals (Naik et al., 2022). Sustaining ethical norms and cultivating confidence in AI applications for diabetes care requires ongoing observation, instruction, and cooperation (Akinci D'Antonoli, 2020; Murdoch, 2021; Naik et al., 2022).

2.6 Genetic Variations in Hemoglobin A1c Levels: Consequences for the Management of Diabetes:

Hemoglobin A1c (HbA1c) levels are impacted by genetic, environmental, and epigenetic variables and represent the average blood glucose during the previous two to three months. Diet, exercise, stress, and medication adherence are examples of lifestyle factors that have a major impact on HbA1c levels. Many genetic loci, including as FTO and GCK, have been linked to HbA1c in recent research, regardless of BMI (Leong & Wheeler, 2018). HbA1c is influenced by both known and unknown loci, some of which are influenced by glycemic pathways and others by erythrocyte biology, according to genome-wide association studies (GWAS) (Soranzo, 2011). The significance of genetic variants in the diagnosis and management of diabetes is highlighted by the impact that specific mutations, especially in African and Hispanic populations, have on HbA1c levels (Sarnowski et al., 2019). Comprehending these hereditary factors is essential to creating individualized treatment regimens for ideal blood glucose regulation. Effective diabetes control still depends on maintaining a healthy lifestyle and taking medications as prescribed.

2.6.1 Glucokinase (GCK):

One of the most important enzymes in the metabolism of glucose is glucose kinase (GCK), which primarily controls the amount of glucose in the liver and pancreatic beta cells (Matsutani et al., 1992). The 7p13 chromosome contains the GCK gene, which is highly variable and produces a wide range of splice variants. These variations have been linked to a number of phenotypes, such as maturity-onset diabetes of the young type II (MODY2) and noninsulin-dependent diabetic mellitus (NIDDM) (Stoffel et al., 1992). The clinical significance of GCK gene mutations has been highlighted by the identification and correlation of certain mutations, such as a C-to-T transition in exon 5, with late-onset non-insulin-dependent diabetic diabetes mellitus (Katagiri et al., 1992).

2.6.1.1 Genetic Variation of the GCK Gene in T2DM Risk:

Studies highlight the significance of GCK in glucose metabolism and diabetes risk in this population. The impact of a common mutation in the glucokinase (GCK) gene on metabolic characteristics and the risk of type 2 diabetes was investigated in Pima Indians (Muller et al., 2014). A new 3' untranslated region (3'UTR) SNP's A allele was linked to higher risk of diabetes and lower rates of energy expenditure and glucose oxidation (Muller et al., 2014). A number of other

single-nucleotide polymorphisms (SNPs) in GCK were also linked to metabolic characteristics. An investigation was conducted on the relationship between the GCK gene and the risk of type 2 diabetes mellitus (T2DM), with a specific focus on the variant GCK rs1799884 linked to fasting plasma glucose (FPG) levels (Wang et al., 2013). The minor A-allele of GCK rs1799884 was found to be significantly associated with an elevated risk of type 2 diabetes, especially in Caucasians, according to a meta-analysis (Wang et al., 2013). A study looked at the relationship between Type 2 Diabetes Mellitus (T2DM) and changes in the glucokinase (GCK) gene, specifically the -30G/A polymorphism. GCK -30G/A genotypes were found to have significant correlations with a number of metabolic markers in both T2DM cases and healthy controls (Fatima et al., 2020). Furthermore, GCK mRNA expression was found to be downregulated in T2DM patients with particular genotypes, indicating a potential role in the advancement of the disease (Fatima et al., 2020).

2.6.1.2 Glucokinase's Function in Insulin Release and Glucose Sensing:

In pancreatic β-cells, glucokinase is essential for the release of insulin in response to glucose and is involved in glycolysis and glucose metabolism. Different from the liver, pancreatic β-cells have a special regulation for it that is essential for accurate glucose sensing; defects in this area may result in diabetes (Matschinsky, 1990). The impact of the SNP GCK rs1799884 on glucose metabolism was demonstrated by a study conducted on a Chinese population, which indicated that it significantly affected insulin secretion phases and fasting glucose levels (Hu et al., 2010). Comprehending these genetic variances can facilitate the creation of focused diabetes remedies. The therapeutic potential of glucokinase modulation requires further investigation.

2.6.1.3 Type 2 Diabetes, HbA1c level and Hepatic GCK Expression Correlates:

Diabetes type 2 has been linked to decreased hepatic glucokinase (GCK) expression (T2D). According to a study examining GCK expression in liver biopsies from both diabetic and non-diabetic participants, GCK expression was considerably lower in those with a HbA1c >7.0% (Haeusler et al., 2015) . Targeting glucokinase activators in T2D treatment is a possibility because of the decrease of hepatic GCK expression, which is thought to be a significant factor in T2D dysregulation.

The glucokinase enzyme, which is encoded by the GCK gene, is essential for controlling glucose metabolism. It affects the synthesis of glucose in the liver as well as the release of insulin from pancreatic beta cells. Glucokinase in the liver promotes glycogen synthesis and inhibits gluconeogenesis by helping to convert glucose to glucose-6-phosphate (Iynedjian, 2009). Reduced glycogen storage and increased hepatic glucose output are the outcomes of impaired GCK function, which raises HbA1c levels and causes hyperglycemia during fasting (Matschinsky, 1990). Glucokinase serves as a glucose sensor in pancreatic beta cells and is necessary for the appropriate release of insulin. GCK gene mutations result in insufficient insulin production in response to glucose, which raises HbA1c and causes postprandial hyperglycemia (Steele et al., 2013). Furthermore, a decrease in GCK activity impairs the effective insulin-regulated glucose absorption in peripheral tissues, hence aggravating hyperglycemia (Matschinsky et al., 2011; Saltiel & Kahn, 2001). Because of this, GCK gene expression and activity are essential for preserving glucose homeostasis, and dysregulation of these processes is intimately linked to the etiology of type 2 diabetes and elevated HbA1c levels.

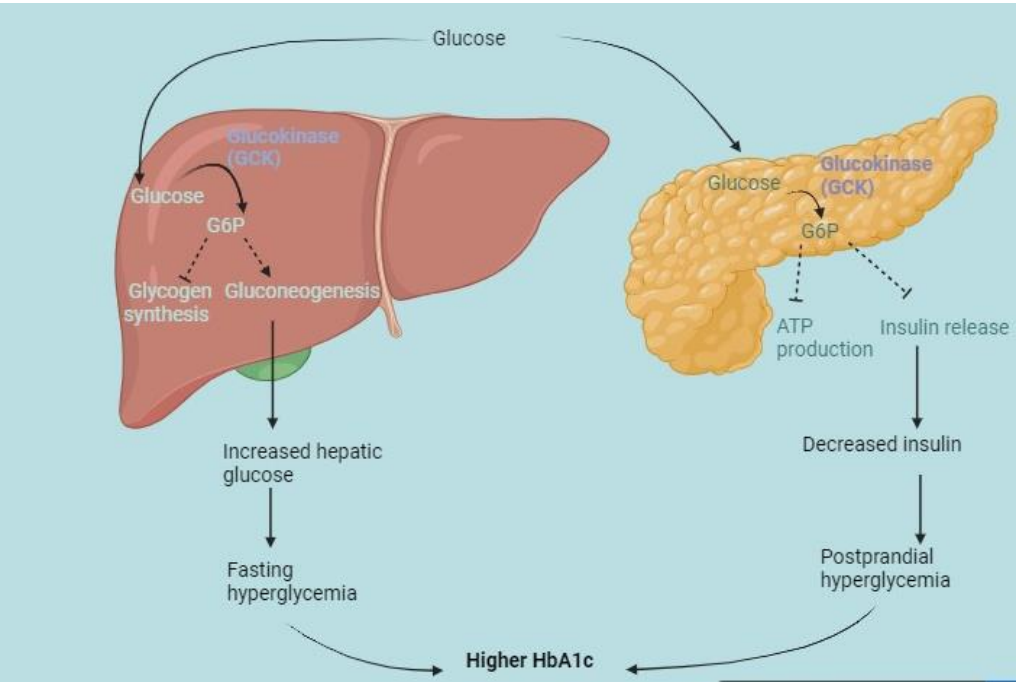


Figure 2.0.1: Pathways showing aberrant GCK expression leading to higher HbA1c level.

2.6.1.4 HbA1c as a Diagnostic Marker for Hyperglycemia Associated with GCK Mutations:

The usefulness of HbA1c in identifying patients with hyperglycemia caused by a glucokinase (GCK) mutation was assessed by observational case-control studies. The ability to differentiate between GCK mutation carriers and patients with type 1 or type 2 diabetes, as well as between individuals with GCK mutations and controls, was demonstrated using HbA1c (Steele et al., 2013). Age-related HbA1c reference ranges may help identify hyperglycemia brought on by GCK mutations, avoiding incorrect diagnosis and treatment (Steele et al., 2013).

2.6.1.5 Clinical Implications:

According to Haeusler et al. (2014), one important contributing element to the dysregulated glucose metabolism in type 2 diabetes is the inhibition of hepatic GCK expression (Haeusler et al., 2015). It may be possible to improve T2D management and glycemic control by focusing on GCK expression. In summary, these results highlight the critical function of GCK in glucose metabolism and its consequences for the etiology and diagnosis of diabetes. Additional investigation into the complex processes behind GCK gene regulation and its effect on glucose homeostasis may provide important new therapeutic approaches for the treatment of diabetes.

METHODOLOGY

3.1 Study subjects:

This investigation was carried out at Immunogenetics Lab, ASAB, NUST. Patients with type 2 diabetes are the study's subjects, and the Federal Government Polyclinic Hospital in Islamabad's ethics council has given its approval. The patients with prior gestational diabetes, type 1 diabetes (differentiated from T2DM by age at onset and ketoacidosis), and other kinds of diabetes were excluded. Before data and blood samples from the federal government polyclinic hospital in Islamabad's department of endocrinology were collected, all patients gave their informed consent.

Over the course of six months, from December 2023 to May 2024, cross-sectional data on these individuals was gathered and recorded from the endocrinology department of the federal government polyclinic hospital in Islamabad without altering the study environment and patient's treatment plan. Using HbA1C as a screening parameter, a thorough face-to-face interview was carried out with patients during this time.

In the study, 80 human participants made up the case control group. Purple-capped EDTA tubes were used to collect 1 cc of blood samples. Fifteen samples were taken from a healthy control group of the same age, and sixty-five samples were taken from patients with type 2 diabetes.

3.1.1 Inclusion and exclusion criteria:

3.1.1.1 Inclusion criteria:

Male and female patients who are 18 years of age or older will participate in the study. Participants must have had type 2 diabetes for at least six months with a verified diagnosis. This criterion guarantees that the subjects are in a stable enough state to support the long-term glycemic control aim of the study. Every participant must also give their informed consent.

Participants should have healthcare records dating back six months, which contain vital information such blood glucose and HbA1c levels as well as other pertinent health data, to guarantee the collection of accurate and pertinent data. Participants also need to speak the language

used in the in-person interview well. By doing this, you can be confident that their responses will be honest and trustworthy, which is crucial for the validity of the study.

3.1.1.2 Exclusion criteria:

The study excludes people with cognitive impairments that would impede their capacity to comprehend and give informed consent. Individuals who have undergone significant surgery within the last three months will also be excluded, as the healing process may impact research factors, as will participants with serious medical illnesses like active cancer that could potentially interfere with the study. Also excluded from the study will be those whose life expectancy is shorter than a year due to a terminal illness. Due to the varied management and metabolic considerations during pregnancy, pregnant individuals will not be admitted. To avoid confounding effects, participants who are presently engaged in another clinical experiment that might affect the outcomes of this study will be excluded. Individuals without adequate clinical records for the previous six months, which are required for thorough data collection and analysis, will also be disqualified.

3.2 Data collection:

Patients' health can be fully understood by academics and healthcare practitioners through the use of direct clinical measurements and patient questionnaires. Numerous pieces of information are included in these data sources, such as demographics, medical history, medication compliance, lifestyle choices, test findings, and coexisting diabetes complications.

3.2.1 Structured questionnaire:

The questionnaire is a structured form that was designed to assess patients' self-management. A thorough interview was conducted with each patient to fill out the questionnaire to gather information on their medical history, demographics, and clinical details, covering age, sex, disease duration, family medical history, overall health status, any existing micro- and macro-vascular complications and lifestyle factors such as diet and physical activity, and psychosocial aspects. This method capture patient-reported outcomes and experiences, complementing clinical data and identifying issues not apparent in clinical exams. This method of data collection provides valuable insights into the patient's perspective, ensuring a more holistic approach to healthcare.

Questionnaire contain following sections:

Table 3-1: Questionnaire sections.

Sr.no.	Section
1.	Patient information
2.	Diabetes related information
3.	Symptoms related information
4.	Physical health and diabetic complication information
5.	Diabetes self-management information
6.	Biochemical profile of patient

3.2.2 Direct clinical measurement:

The method of data collecting includes direct clinical measurement in addition to the collection of patient-reported information via structured questionnaires. Direct clinical measurement includes information gleaned straight from clinical evaluations and observation processes. These include physical measurements like height and weight, body mass index (BMI), laboratory results like HbA1c levels and blood glucose readings, and blood pressure. Direct clinical measures have the advantage of providing objective and reliable health data, which is essential for correct diagnosis and continuous treatment efficacy monitoring. This thorough approach guarantees that the basis of our study is a solid combination of objective clinical data and subjective patient feedback.

3.2.2.1 Reported clinical measurement:

Clinical measurements, such as uric acid levels, liver function tests, renal function tests, and lipid profile tests, were used to evaluate each patient's biochemical profile. Important information on liver health, kidney function, cardiovascular health, and other metabolic diseases can be gained from these tests.

3.3 Data analysis:

3.3.1 Data organization:

An Excel spreadsheet was used to systematically enter and arrange the patient data. To ensure that all pertinent information was carefully recorded, each patient's data was entered into separate rows. The spreadsheet was organized with distinct data categories represented by columns that had clear labels.

3.3.2 Data encoding:

Most of the gathered data is categorical such as gender, frequency of physical activity, frequency pf blood glucose monitory, suffering from any hypertension and anxiety etc. For these category data to be utilized in an algorithm or machine learning model, they must be transformed into numerical data. Two types of data encoding methods were utilized: one-hot encoding and label encoding.

For non-ordinal categorical variables-where there is no intrinsic order among the categories like gender, suffering from hypertension or not, stay hydrated or not, feeling symptoms of anxiety or not, One-Hot Encoding was employed. By converting each category into a binary vector, this method produces a representation where each category is denoted with a 1 in the appropriate location and a 0 in other places.

Table 3-2: Few examples of One-Hot encodings.

Sr.no.	Variables	Data	Encoding
1.	Gender	Female	0
		Male	1
2.	Marital status	Single	0
		Married	1
3.	Family history of diabetes, Smoking, Suffering from hypertension, Is patient checking blood glucose regularly, Do patient spare time for physical activity, Do patient follow specific dietary recommendations, Do	No	0
		Yes	1

	patient drink water, and Is patient taking medication for depression?		

For ordinal categorical data, such as frequency of physical activity, quality of sleep where categories have a meaningful order, label encoding was used. Using this strategy, each category is given a distinct integer according to its order.

Table 3-3: Few examples of Label encodings

Sr.no.	Variables	Label	Encoding
1.	Rating of self-management, How comfortable in using technology	Poor	0
		Average	1
		Fair	2
		Good	3
		Excellent	4
2.	Type of physical activity	Chores	0
		Walk	1
		Exercise	2
		Walk+ Exercise	3
3.	Frequency of physical activity	No	0
		Weekly	1
		Twice a week	2
		4 days in a week	3
		Daily	4
		Twice a day	5
4.	Quality of sleep	Poor	0
		Average	1
		Good	2

3.3.3 Data preprocessing and cleaning:

- Removing duplicate records: Duplicate records were removed to ensure the quality of the data.
- Removing variable with missing values: The variables with large number of missing values were removed. Removal could be a better option to prevent bias from large-scale imputation. For example, the entries of "suffering from any physical condition that effect patient's physical activities" and "dietary habit" were extremely low that it has to be excluded from final analysis.
- Imputing missing values: For variable that are most important or have few missing values, missing values were imputed with mean or median value of the variable. In such cases, imputing the missing value is more suitable rather than excluding the whole variable from final analysis. For example there were 7 missing values in "duration of diagnosis", these missing values were replaced with the mean value.
- Handling outliers: For continuous variable like HbA1c value, age, and random blood glucose levels, it is necessary to identify and remove significant outliers' values to ensure accurate building of machine learning model. Z-score was calculated by using formula Z-score=(X-Mean)/SD. Values with Z-score greater than 3 (|Z|>3) were considered significant outliers and were removed. Outlier values were removed from final dataset but were retained in the initial data analysis steps.

3.3.4 Statistical analysis for feature selection:

- Exploratory data analysis: Data of all variables are analyzed separately and in relation to target variable that is HbA1c level by calculating statistical values like mean, median, standard deviation (SD), and frequency distribution on Excel. Also, these statistical values were analyzed graphically by drawing histogram or bar graphs.
- To analyze the relation of each variable with HbA1c value following data analysis was performed using "Data Analysis" tool on Excel, P-values were calculated to identify statistically significant feature:
    - ANOVA
    - T-test

- A correlation matrix was created for the variable of choice. At this point, the majority of the statistically non-significant variables had already been eliminated.

3.4 Model development:

Google collaborator (Google Colab) (https://colab.research.google.com/) is a free cloud-based platform that enables users to build and run Python programs through a web interface. The data science and machine learning groups particularly value it for its user-friendliness and robust computing capabilities, which include free access to GPUs and TPUs for quicker model training. It offers real-time collaboration between several users, comes pre-installed with many popular Python modules, and interacts smoothly with Google Drive for simple sharing and saving.

- In order for the task to be completed on Google Colab, existing programs were changed. In order for subsequent scripts to be executed on the data, we first run the code to upload the data file.
- To make sure the data was correctly cleaned and preprocessed, another piece of code was run.
- Regression models are suitable for the prediction of HbA1c level as it is a continuous variable. Linear regression model, Neural Network regression model, Random forest regression model, Gradient Boosting regression model and Support Vector regression (SVR) model were build and were trained with 80% of the data while being tested with the remaining 20% of the data. The splitting of the dataset into training and testing dataset was random. The graphs were plotted for the predicted vs true value and predicted vs residual value.
- Another code was run to classify the predicted HbA1c value into “Excellent glycemic control”, “Good glycemic control”, “Fair glycemic control”, and “Poor glycemic control”. The reference values for classification were reported in the literature.

3.5 Model validation:

For model validation, new unlabeled dataset was utilized to predict their HbA1c level using built models. Our model will predict the HbA1c value and level of glycemic control in patients based on the features that were used in model building.

3.6 Model evaluation:

Performance of the models were evaluated by calculating Mean Squared error (MSE) and R^2 value for each model. Performance and accuracy of different models was compared by plotting graph for their MSE and R^2 value.

3.7 Quantitative Real Time PCR analysis:

Using real-time PCR, the expression levels of the GCK gene in patient blood samples in relation to their level of glycemic control were compared to control group. Specific primers for genes were utilized to determine how genes are expressed. GCK (glucokinase) gene expression study has the potential to greatly improve and validate machine learning models that predict HbA1c levels, a crucial indicator of glycemic management in individuals with type 2 diabetes. Because GCK is essential for insulin control and glucose metabolism, blood glucose and HbA1c levels are directly correlated with GCK expression levels. Researchers may verify the model's predictions and make sure they accurately reflect underlying biological pathways by comparing GCK expression with HbA1c data. For example, lower HbA1c levels should be correlated with high GCK expression, indicating efficient glucose metabolism.

3.7.1 Primer designing:

Primer for β-Actin was already reported in literature. These were checked for non-specific binding using UCSC insilico PCR. Primers for GCK gene was designed using Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi), GCK mRNA sequence was taken from NCBI (https://www.ncbi.nlm.nih.gov/gene/2645). Designed primers were checked on UCSC insilico PCR (https://genome.ucsc.edu/cgi-bin/hgPcr).

3.7.2 Sample collection:

Samples were taken into purple-capped EDTA tubes using a 3ml syringe. Before any blood was drawn, informed consent was obtained. Blood samples were accurately labeled with the sample number and then transported to the Immunogenetics Lab (IGL) at the Atta-ur-Rahman School of Applied Biosciences (ASAB), National University of Sciences and Technology (NUST),

Islamabad. They were then kept in an ice storage box to prevent hemolysis, clotting, and degradation. To ensure a high RNA yield, RNA was then extracted right away.

3.7.3 RNA extraction protocol:

Table 3-4: Protocol for RNA Extraction

Steps	Procedure	Quantity
1.	Add 750µl of Trizol reagent in an autoclaved and labelled Eppendorf tubes. Add 200µl of blood sample in trizol containing Eppendorf tube, vortex to homogenize.	750µl of Trizol reagent +200µl of blood sample
2.	Centrifuge it at 12000rpm for 10min in 4^0C refrigerated centrifuge.	
3.	Transfer the red/pink intermediate layer to new labelled Eppendorf tube. Incubate at room temperature for 5 min, then add 20µl of 5N glacial acetic acid, mix it vigorously and leave it at room temperature for 5 min. Then add 200µl of chloroform and mix vigorously and leave the mixture at room temperature for 10 min.	red/pink intermediate layer +20µl of 5N glacial acetic +200µl of chloroform
4.	Centrifuge it at 12000rpm for 10min in 4^0C refrigerated centrifuge.	
5.	Transfer the upper layer to the new centrifuge tube, then add 500µl of chilled isopropanol. Vortex the mixture to precipitate RNA, incubate it at room temperature.	Upper layer+500µl of chilled isopropanol

6.	Centrifuge it at 12000rpm for 10min in 4^0C refrigerated centrifuge.	
7.	Discard supernatant and wash the pellet with 1ml of 75% ethanol in DEPC treated water. Votex to dissolve pellet.	Pellet+1ml of 75% ethanol in DEPC treated water.
8.	Centrifuge it at 14000rpm for 10min in 4^0C refrigerated centrifuge.	
9.	Discard supernatant and air dried the pellet and dissolve the pellet in a 25µl buffer that is DEPC treated water.	Pellet+25µl DEPC treated water.
10.	Incubate it at 55-60^0C for 10 min before temporary storage at -80^0C.	

3.7.3.1 Quantitative analysis of extracted RNA by Nano drop:

Thermo Scientific Nano Drop ND-2000 spectrophotometer was used to quantify RNA. The purity of the RNA sample was ascertained using the absorbance ratio A260/A280. For the measurement, 1µl of RNA in DEPC treated water was employed, while 1µl of the treated water served as the blank.

3.7.4 Complementary DNA synthesis:

Complimentary DNA was synthesized using Thermo Scientific cDNA kit by following steps:

Table 3-0-5: Protocol for cDNA synthesis

Steps	Procedure	Quantity
1.	Take PCR tubes and add 2µl of RNA sample which will act as template for cDNA synthesis.	2µl RNA

	Add 1µl oligo dT primers and make the volume 12.5µl by adding DEPC treated water.	+ 1µl oligo dT primers+ 9.5µl DEPC treated water= 12.5µl
2.	Gently mix and centrifuge briefly	
3.	The mixture is heat shocked for 5 min at 65^0C, then quickly cooled on ice.	Heat shock
4.	Centrifuge the mixture.	
5.	Place the tubes on ice and add the remaining agents including 4µl of reaction buffer, 0.5µl of RNase inhibitor (20 Units), 2µl of 10mM dNTP mix, 1µl of Reverse Transcriptase enzyme (200 Units). Add 12.5µl of DEPC treated water to make a total volume of 20µl.	4µl reaction buffer+ 0.5µl of RNase inhibitor+ 2µl 10mM dNTP mix+ 1µl of Reverse Transcriptase enzyme + 12.5µl DEPC treated water= 20µl.
6.	Incubate in PCR at temperatures given in figure	
7.	Store the cDNA at -20^0C.	

3.7.5 Semi-quantitative PCR:

Complimentary DNA synthesis was confirmed using semi-quantitative PCR followed by gel electrophoresis.

1. Take PCR tubes and add following reagents:

Sr. no.	Reagent	Quantity

1.	cDNA	0.25 µl
2.	$MgCl_2$	2 µl
3.	Taq buffer	2 µl
4.	2mM dNTP mix	2 µl
5.	Forward primer of β-Actin	2 µl
6.	Reverse primer of β-Actin	2 µl
7.	Thermo Scientific Taq DNA polymerase	0.25 µl
8.	Nuclease free water	9.5 µl
	Total volume	=20 µl

2. The reaction mixture was spin loaded to thermal cycler at following conditions:

3.7.6 Gel electrophoresis:

The amplification of cDNA was confirmed by running PCR products on 2% agarose gel.

1. For 2% agarose gel, 1.2 g of agarose was dissolved in 60ml distilled water, 3 µl of ethidium bromide was used as a dye. The gel was allowed to solidify in gel casting tray.
2. 1X TAE buffer was used that was diluted from 50X stock solution.
3. 5 µl of PCR product was loaded with 3 µl of loading dye.
4. Gel was analyzed on Dolphin Doc. Product length was compared with 50bp DNA ladder.

3.7.7 Real time PCR:

Using flouresence-based SYBER green technology and real-time PCR using the Applied Bios system 7300 RT PCR equipment, the amount of cDNA was determined. GCK quantitative expression determined via real-time PCR. B-actin served as an internal control to ensure that the data was normalized.

1. Take PCR tubes and add following reagents:

Sr.no	Reagent	Quantity
1.	Forward primer	0.8 µl
2.	Reverse primer	0.8 µl
3.	SYBER green qPCR master mix	10 µl
4.	cDNA	25ng
5.	Nuclease free water	
	Total volume	=20µl

2. The reaction mixture was loaded in real time PCR at following conditions:

Note: All reactions should be run in triplicates. Each PCR reaction was run with non-template control.

3.7.8 Statistical analysis:

Data was organized on Excel spreadsheet and was analyzed statistically. Statistical significance was defined as a P-value of less than 0.05 and a 95% Confidence Interval (CI) was applied to the data.

Sr.no.	Statistical analysis	Purpose
1.	Student's t test	Compare mean delta CT values from RT PCR
2.	Two way ANOVA	Analysis of variance among study participants
3.	Bonferroni's Multiple comparison test	Compare the expression of GCK gene in two groups

RESULTS

4.1 Demographic characteristics of study participants:

During the study, data of 650 patient was collected who were visiting hospital for routine checkup with confirmed diagnosis. Of total patients, 71% were female and 29% were male; majority of the individual lie in 55-56 years age group; 99% are married; 50% live in joint family and 50% have nuclear family system. 51.3% of the patients are not suffering from any diabetic complication, while 16% are suffering from cardiovascular (CVD) complication and 11.9% have ocular complication, 5.1% have renal complications while the rest have combination of complications. 63.7% of the patients have family history of diabetes while 36.3% do not have any family history. 8% are smokers, 4% have quitted smoking and 88% are non-smokers. 44% are suffering from hypertension, while 56% are non-hypertensive. The general and clinical characteristics are summarized in the table.

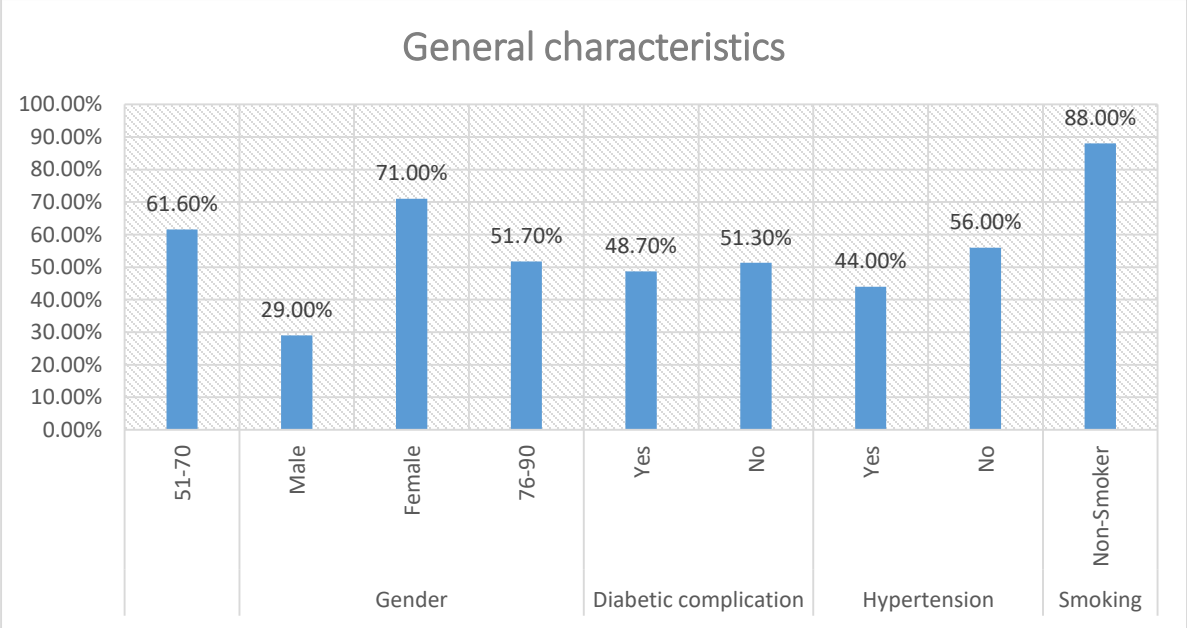


Figure 4.1: General characteristics of patients

Table 4-1: Summary of general and clinical characteristics

Sr.no	General characteristics	Average ±SD

1.	Age (Years)		55.19 ± 10.28
2.	BMI		30.75 ± 6.66
3.	Duration of diabetes (Years)		8.0 ± 6.0
	Biochemical profile		
1.	HbA1c value (%)		9.91 ± 2.19
2.	Random Blood glucose level (BSR) (mg/dl)		292.09 ± 93.66
3.	Fating blood glucose level (BSF) (mg/dl)		186.36 ± 70.48
4.	Lipid profile:	Cholesterol (mg/dl)	185.83 ± 53.96
		LDL (mg/dl)	120.34 ± 37.06
5.	Renal Function Test (RFT)	Urea (mg/dl)	30.41 ± 11.91
		Uric acid (mg/dl)	4.75 ± 1.44
		Creatinine (mg/dl)	0.81 ± 0.24
6.	Liver Function Test (LFT)	ALP (IU/L)	119.00 ± 53.36
		ALT (IU/L)	40.01 ± 26.81
7.	Serum ions level	Na (mmol/l)	136.52 ± 3.24
		K (mmol/l)	4.32 ± 0.68
		Cl (mmol/l)	99.26 ± 7.68

According to American Diabetes Association (ADA), 6.5% is the HbA1c level threshold for good glycemic control in patients without diabetic complications while 8.0% is the HbA1c level threshold for good glycemic control in patients with diabetic complications (https://diabetes.org/living-with-diabetes/type-2). Patients with type 2 diabetes were classified according to their degree of glycemic control, based on their HbA1c level. It is clear from the data that the majority of patients (46%) have poor glycemic control, whereas very few patients (9%+12%) have excellent and good glycemic control.

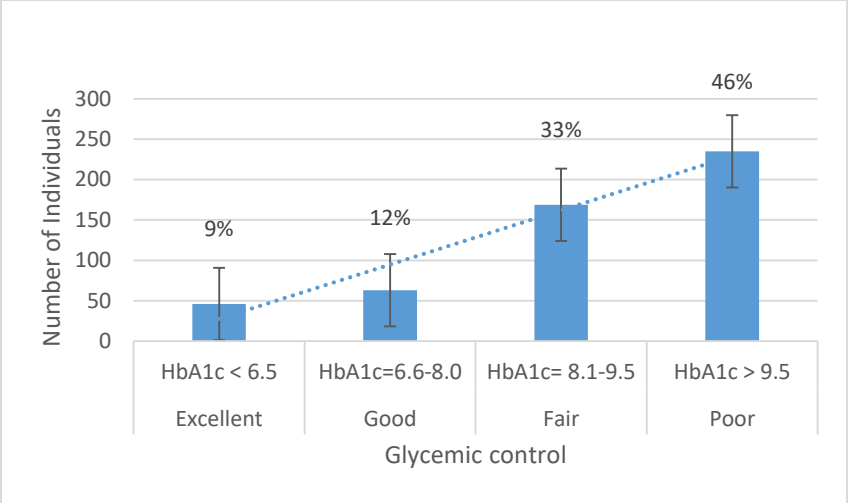


Figure 4.2: Number of patients with different levels of glycemic control

The individuals have varying level of comfort that they can use technology for their diabetes self-management as depicted by graph.

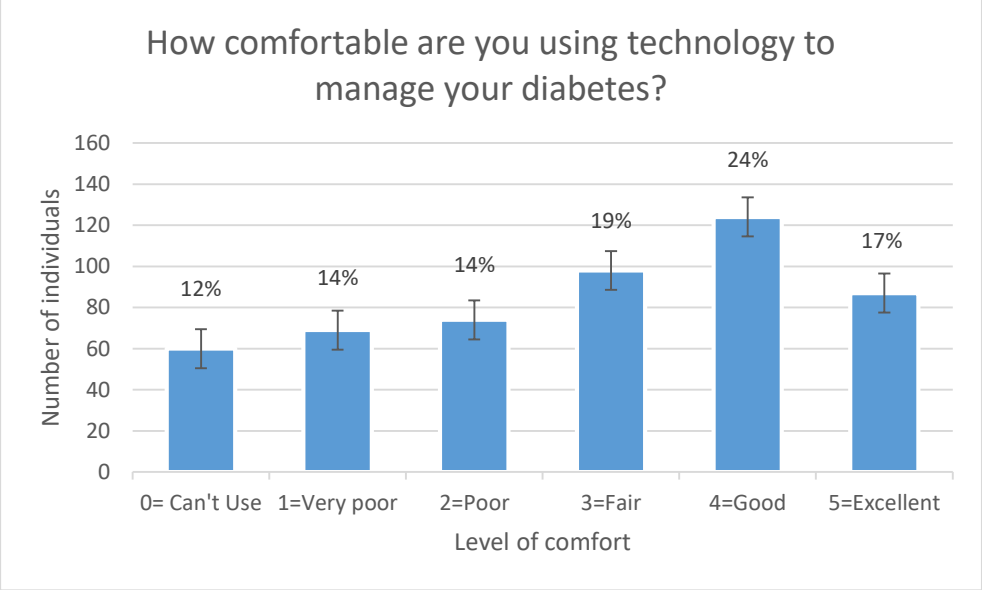


Figure 4.3: Number of patients with different level of ease to use technology

4.2 Feature selection:

The variable that have so many missing data like the variables of “Biochemical profile”, “Suffering from any physical condition that effect patient’s physical activities” and “Dietary habit”.

The relation of the variable with HbA1c value was analyzed and statistically significance P-value was calculated for HbA1c values in different groups of the variable. The statistical significance of variable was analyzed by T-Test if there are two groups within the variable while analyzed by ANOVA if there are two or more groups within the variable. P-value ≤ 0.05 was considered threshold and statistically significant features are selected. P-values are given in the table. Graphs were also plotted for the average HbA1c level in different groups of patients.

Table 4-2: Summary of feature selection

Sr.no.	Variable	P-value	Feature Selected
1.	Age	0.0134	✓
2.	BMI	P <0.001	✓
3.	Gender	0.47	×
4.	Family history of type 2 diabetes	0.046	✓
5.	Duration of diagnosis	P <0.001	✓
6.	Diabetic complication	0.75	×
7.	Suffering from hypertension	0.001	✓
8.	Smoking	0.9	×
9.	Random blood glucose level (BSR)	P <0.001	✓
10.	Rate of diabetes self-management	P <0.001	✓
11.	Frequency of physical activity	P <0.001	✓
12.	Frequency of blood glucose monitoring	0.000267	✓
13.	Frequency of sugar consumption	P <0.001	✓
14.	Stay hydrated or not	0.7	×
15.	Fast food consumption	0.05	✓
16.	Sleep quality	0.12	×
17.	Depression	0.000627	✓

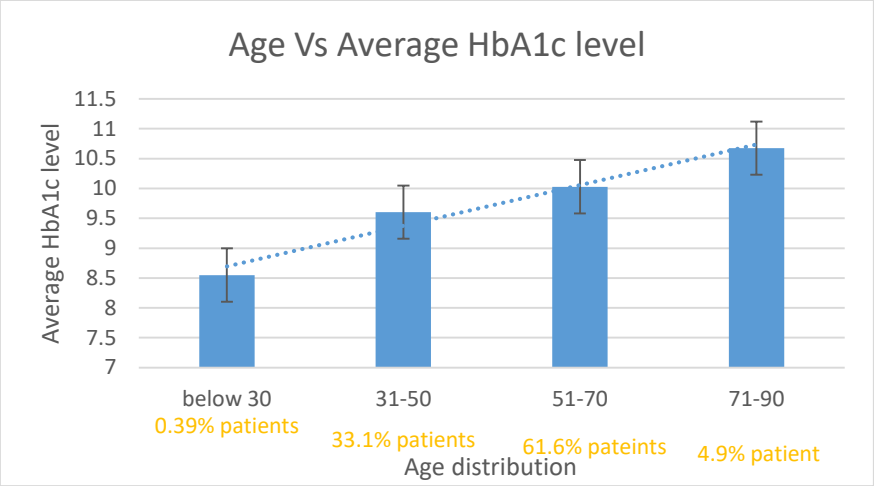


Figure 4.4: Graph showing relation between age and HbA1c level

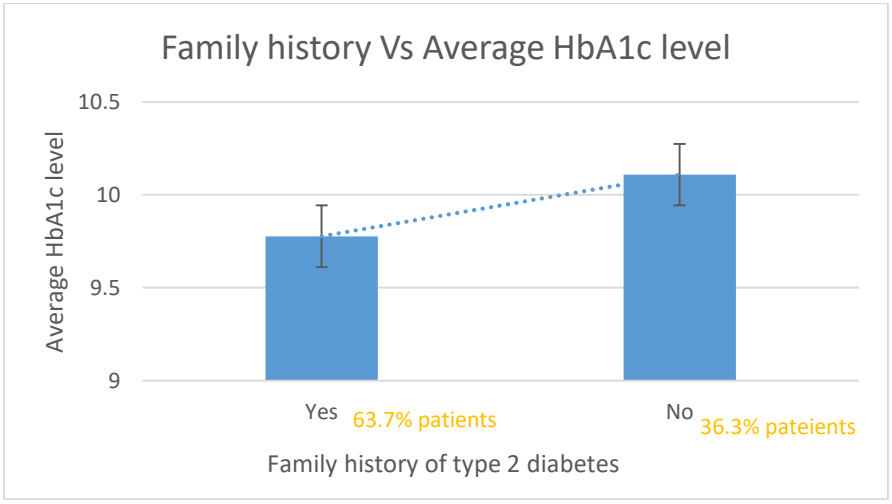


Figure 4.5: Graph showing relation between family history and HbA1c level

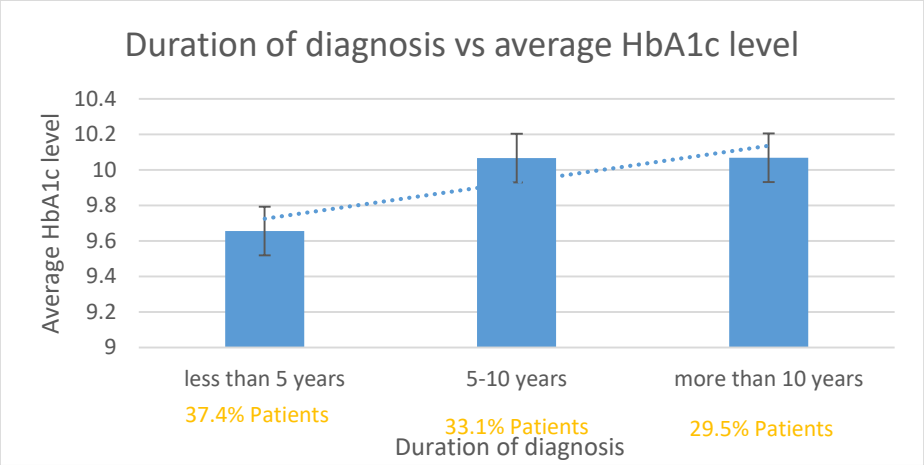


Figure 4.6: Graph showing relation between duration of diagnosis and HbA1c level

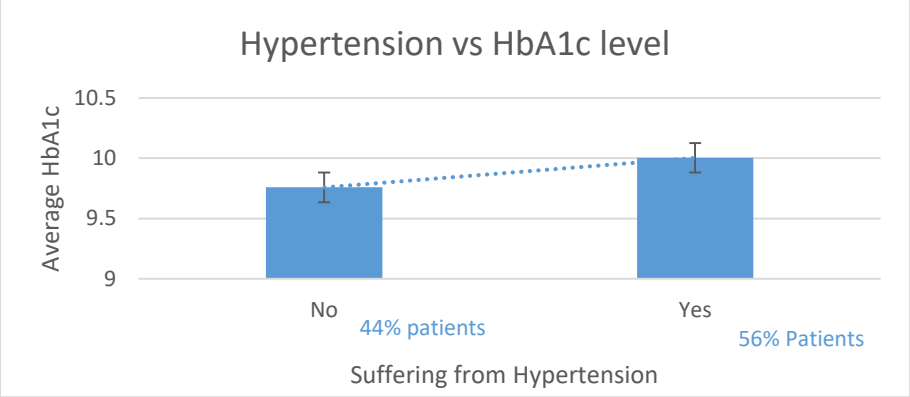


Figure 4.7: Graph showing relation between hypertension and HbA1c level

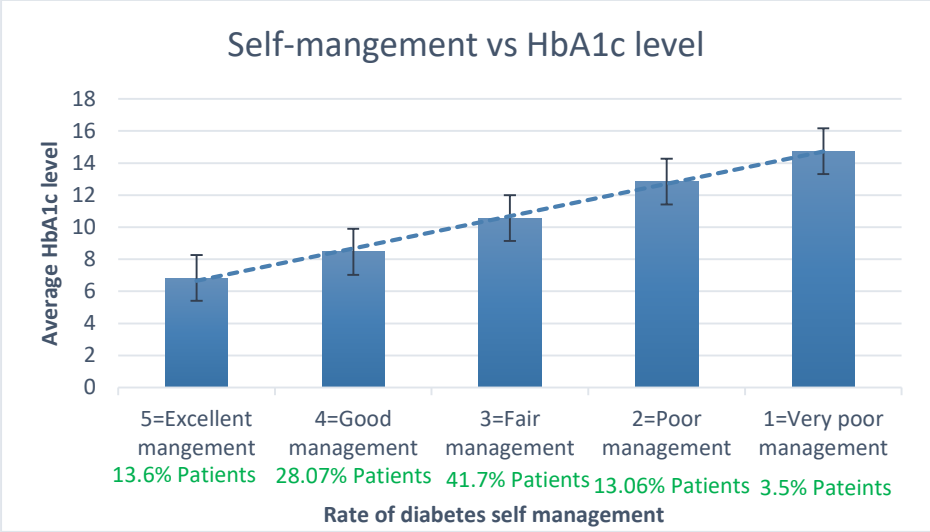


Figure 4.8: Graph showing relation between self-management and HbA1c level

A total of top 15 features are selected after the statistical analysis and a correlation matrix is constructed between these 15 variables to identify the features that have strong positive or negative correlation with the target variable that will be predicted by the regression model. Correlation coefficient value closest to 1 shows strong positive correlation and correlation coefficient values closest to -1 indicated strong negative correlation. As shown in the figure strong positive correlation is indicated by increasing intensity of red color while strong negative correlation is indicated by increasing intensity of green color.

	HbA1c level (%	Age	BMI	Duration of diagnosis	Family history of diabetes	Hypertension	Random Blood glucose levels	Rating of level of diabetes management	Frequency of glucose monitoring	Physical activity	Frequency ofphysical activity	Frequency of sweets consumption	How often you eat fast food	Quality of your sleep	Depression or anxiety	Comfortable of using technology
HbA1c level (%)	1															
Age	0.108527	1														
BMI	-0.06056	-0.08959	1													
Durartion of diagnosis	0.072229	0.265163	-0.06121	1												
Family history of diabetes	-0.07618	-0.15952	0.014077	0.0218788	1											
Hypertension	0.058224	0.105473	0.094116	0.0887632	0.045381	1										
Random Blood glucose levels	0.384279	-0.01128	-0.03567	-0.016027	-0.000969	0.0306531	1									
Rating of level of diabetes management	-0.90531	-0.09518	0.08668	-0.07884	0.091341	-0.070112	-0.37822	1								
Frequency of glucose monitoring	-0.21482	-0.01149	-0.00611	-0.057984	0.072822	-0.055589	-0.12325	0.23102	1							
Physical activity	-0.49358	-0.09201	-0.01447	-0.033829	0.085494	-0.109164	-0.22039	0.527293	0.15619298	1						
Frequency ofphysical activity	-0.72066	-0.09409	0.02971	-0.069096	0.076766	-0.13684	-0.30782	0.776091	0.23788384	0.748509	1					
Frequency of sweets consumption	0.059694	-0.00196	0.044675	0.0126157	0.020756	-0.027545	-0.01361	-0.11076	-0.1168397	-0.04761	-0.100902	1				
How often you eat fast food	-0.1078	-0.17005	0.060801	-0.146451	0.072147	-0.067992	-0.0056	0.053077	0.03585639	0.088669	0.0723123	0.1849663	1			
Quality of your sleep	-0.07807	-0.03386	-0.03209	-0.038332	-0.068725	-0.117056	-0.06387	0.071621	0.09466756	0.095071	0.1511933	-0.0152623	0.114814	1		
Depression or anxiety	-0.13327	-0.05747	0.001215	-0.013227	0.019208	0.0773747	-0.01128	0.110036	-0.0417684	0.056352	0.0854284	0.03525088	0.044607	-0.25924	1	
Comfortable of using technology	-0.23159	-0.21002	-0.08305	-0.14615	0.120134	-0.034383	-0.05311	0.229637	0.21759004	0.159025	0.194774	0.05932925	0.209787	0.131476	0.019953	1

Figure 4.9: Correlation matrix: strong positive correlation is indicated by increasing intensity of red color while strong negative correlation is indicated by increasing intensity of green color.

4.3 Model training:

Necessary libraries from python are imported for model development. The data of the selected features is loaded for training of regression model in later steps. As shown in the figure, the data file is successfully loaded; the word “float” means that the data of that variable have decimal values, while “int” means the data of that variable have integer values. “64” refers to 64 bits that is the size of computer processor. It is confirmed that our data file contain preprocessed and cleaned data as indicted by “0” in the figure.

```
HbA1c level (%)                                                  float64
Age                                                                int64
BMI                                                              float64
Durartion of diagnosis                                             int64
Any other family member suffering from diabetes                    int64
Suffering from hypertension                                        int64
Random Blood glucose levels                                        int64
How would you rate your current level of diabetes management       int64
How frequently do you check                                        int64
Do you spare some time for physical activity to manage your diabetes  int64
How frequently you practice physical activity                      int64
Do you consume sweets or sugary drink and how often                int64
Hw often you eat fast food                                         int64
What is the quality of your sleep                                  int64
Are you feeling symptoms of depression or anxiety                  int64
How comfortable are you using technology to manage your diabetes   int64
dtype: object
```

Figure 4.10: Floating of dataset for training of models

```
print(data.isnull().sum())
```

```
HbA1c level (%)                                                  0
Age                                                              0
BMI                                                              0
Durartion of diagnosis                                           0
Any other family member suffering from diabetes                  0
Smoking                                                          0
Suffering from hypertension                                      0
Random Blood glucose levels                                      0
How would you rate your current level of diabetes management     0
How frequently do you check                                      0
Do you spare some time for physical activity to manage your diabetes  0
How frequently you practice physical activity                    0
Do you consume sweets or sugary drink and how often              0
Do you drink enough water to stay hydrated                       0
Hw often you eat fast food                                       0
What is the quality of your sleep                                0
Are you feeling symptoms of depression or anxiety                0
How comfortable are you using technology to manage your diabetes  0
dtype: int64
```

Figure 4.11: Ensuring that cleaned data is loaded for training

Regression models are suitable for predicting HbA1c level as it is a continuous variable. Linear regression model, Neural Network regression model, Random Forest regression model, Gradient Boosting regression model, and Support Vector Regression model are trained and tested on the dataset.

A linear regression model predicts the value of a dependent variable (response variable) using one or more independent variables (predictors). The dependent and independent variables in the model are assumed to have a linear relationship.

A neural network regression model is intended to forecast continuous output values. It is made up of an input layer, one or more hidden layers, and an output layer, which are all layers of interconnected nodes, or neurons. When the objective is to forecast real-valued outcomes, such as the HbA1c value, this model is applied.

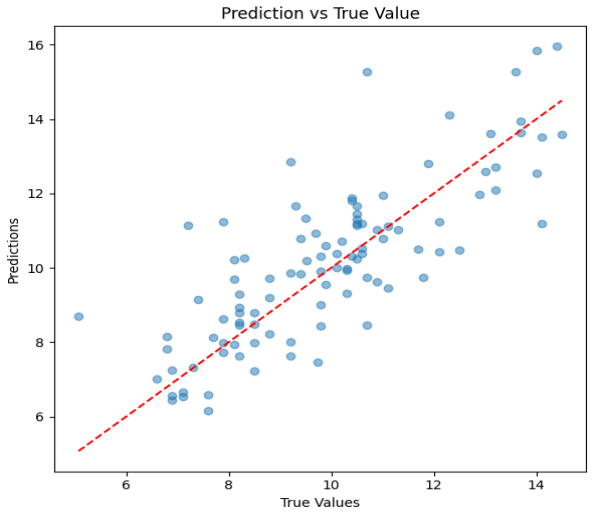
In Random Forest Regression mode there are 100 trees in the model at startup (n estimators=100). Depending on the size of the dataset and the difficulty of the task, this value can be changed. The significance of each characteristic in predicting the target variable can be automatically determined by Random Forests.

The gradient boosting regression is equipped with 100 boosting stages (n estimators=100). Each tree's contribution is influenced by the learning rate; lower rates may require more trees to model difficult datasets but may result in better performance. It indicates which features have the greatest influence on forecasts.

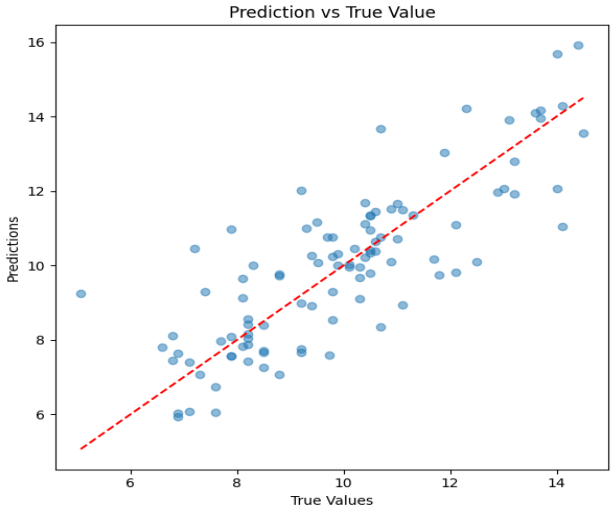
Support Vector Regression (SVR) model is used to predict continuous values in regression tasks. SVR seeks to ensure that the majority of predictions fall within a given margin of tolerance in addition to identifying the hyperplane in a high-dimensional space that best fits the data. This strategy emphasizes striking a balance between prediction accuracy and model complexity, which helps to decrease prediction errors and maintain model generalization.

The dataset is divided into training and testing dataset randomly. 80% data is used to train the model and the remaining 20% data is used to test the model. Graph is plotted between actual HbA1c values versus the HbA1c values predicted by the trained model during testing. A prediction

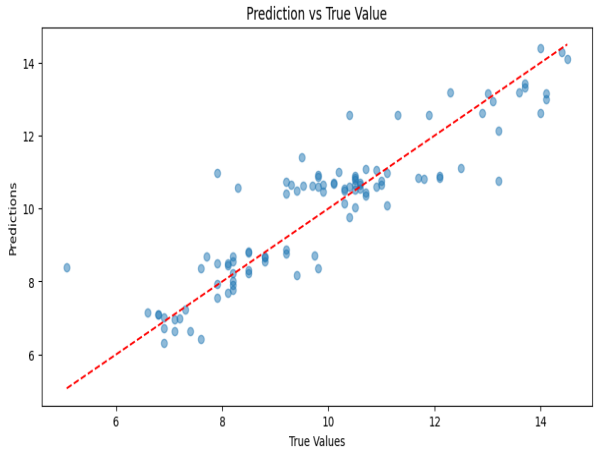
plot illustrates how the model's predictions are compared against the real data. If the true values are close to the predicted values than it suggests accuracy of the model.



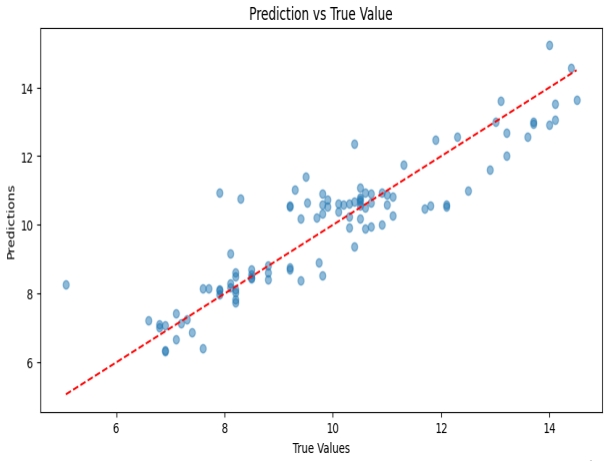
A



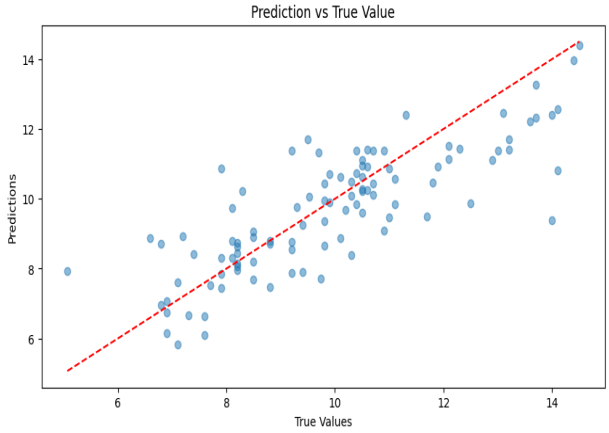
B



C



D



E

Figure 4.12: A prediction plot: An ideal model would align with the dashed red line. Red line is the predicted outcome and blue dots are the actual values. **A:** Prediction plot for linear regression model. **B:** Prediction plot for Neural network model. **C:** Prediction plot for Random forest regression model. **D:** Prediction plot for Gradient boosting regression model. **E:** Prediction plot for Support Vector Regression model.

4.4 Model performance:

Furthermore, MSE (Mean Squared Error) and R^2values are calculated to identify the accuracy and sensitivity of the model respectively. Lower the MSE value, greater is the accuracy of model and vice versa. R^2values (Coefficient of determination) close to 1 indicates better fit of the model and vice versa. R^2values also indices how well the independent features predict the dependent variable i.e. HbA1c level.

Gradient boosting regression model is the best with the highest coefficient of determination (R^2) of 0.817 and the lowest Mean Squared Error (MSE) of 0.796 of the models evaluated. This shows that out of all the models tested, Gradient Boosting fits best with the dataset.

Table 4-3: Summary of model performance

Model	**MSE**	**R^2**
Linear regression	0.84	0.776
Neural network regression	1.893	
Random forest regression	0.82	0.781

Gradient boosting regression	0.796	0.817
SVR	1.563	0.642

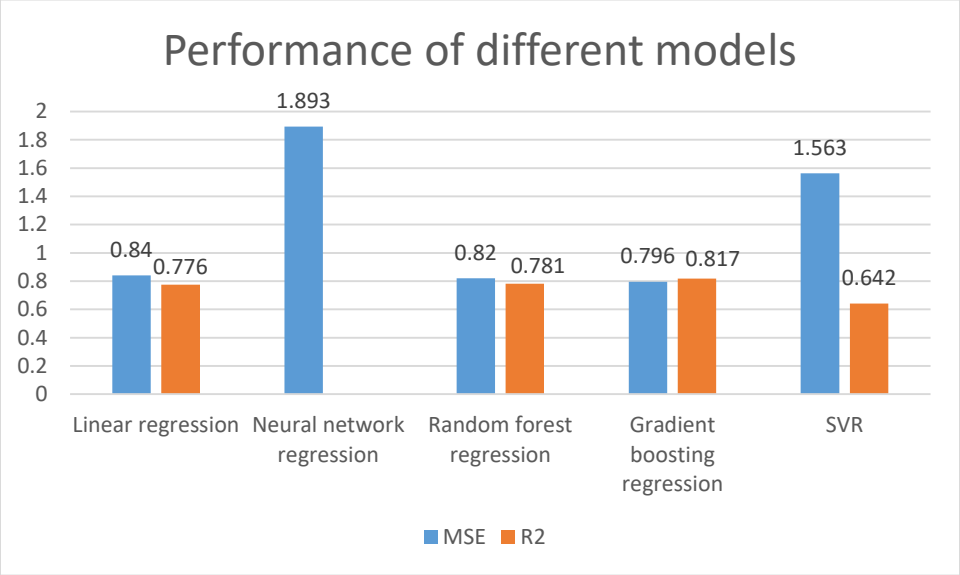


Figure 4.13: Graph showing performance of different models

Learning curve is typically used to evaluate the performance of the machine learning model. Since gradient boosting regression model performs well, for further validation of the performance learning curve is plotted as shown in the figure. The learning curve plot shows how a machine learning model performs as the quantity of training instances rises. The Mean Squared Error (MSE) is displayed on the Y-axis, with lower values denoting greater performance, and the number of training samples is represented on the X-axis. The cross-validation score is shown by the green line, while the training score is represented by the red line. Both errors are high at first, which suggests under fitting. Both errors decrease with an increase of training examples, indicating increased performance. Both the cross-validation and training errors level out after continuing to decline and stable, suggesting strong generalization. Although the model is marginally overfitting, it is generally doing well on untested data, as seen by the narrow difference between the training and validation errors.

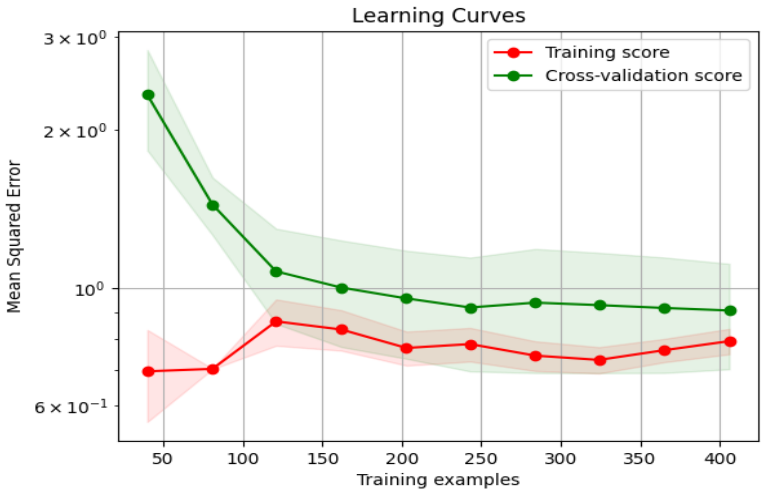


Figure 4.14: Learning curve for gradient boost regression model

4.5 Validation of Gradient boosting regression model on new unseen data:

The HbA1c level prediction by Gradient Boosting Regression model is validated by predicting the HbA1c value of new Patients and the HbA1c level is classified into 4 categories as HbA1c level indicates the level of glycemic control. Lower HbA1c level indicates better glycemic control and vice versa. The classification is based on the threshold values of HbA1c level given by ADA. Validation is shown in the figure.

Table 4-4: Reference HbA1c values indicating level of glycemic control

Sr.no.	Glycemic control	HbA1c value
1.	Excellent	<6.5 (Green)
2.	Good	6.6-8.0 (Blue)
3.	Fair	8.1-9.5 (Orange)
4.	Poor	>9.5 (Red)

```
Enter the feature values:
Enter Age: 52
Enter BMI: 27.21
Enter Duration of Diagnosis: 7
Enter Any other family member suffering from diabetes : 1
Enter Suffering from hypertension: 0
Enter Random Blood glucose levels: 223
Enter How would you rate your current level of diabetes management: 3
Enter How frequently do you check: 2
Enter Do you spare some time for physical activity to manage your diabetes: 1
Enter How frequently you practice physical activity: 3
Enter Do you consume sweets or sugary drink and how often: 0
Enter Hw often you eat fast food: 1
Enter What is the quality of your sleep: 1
Enter Are you feeling symptoms of depression or anxiety : 1
Enter How comfortable are you using technology to manage your diabetes: 3
Predicted HbA1c value: 10.25
Category: Poor
```

Figure 4.15: Gradient boost regression model predicting HbA1c level using new unseen data

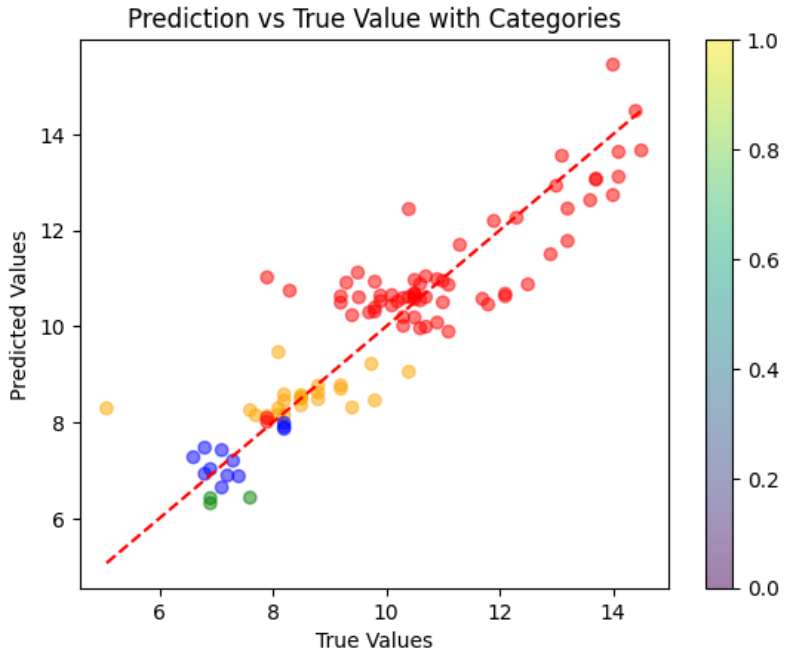


Figure 4.16: Predicted HbA1c level values using gradient boost regression model are classified into 4 categories of glycemic control

4.6 Primer designing:

The primers of GCK gene was designed using Primer 3 plus and was confirmed using UCSC in-silico PCR results.

Table 4-5: GCK Primer characteristics

Primer	Sequence 5' to 3'	Length	GC% content	TM	Product length
GCK Forward	ACCATTACCTCCCTGAGTCT	20bp	50%	60.4	208
GCK Reverse	TTCCCATAGATGCCTTCCAC	20bp	50%	60.4	

UCSC In-Silico PCR

>chr7:44354394-44354601 208bp ACCATTACCTCCCTGAGTCT TTCCCATAGATGCCTTCCAC
ACCATTACCTCCCTGAGTCTccttttgtttttatctataaaactagaaaa
atattctttccataggaatgttgttggaaataataaaacattatattaca
agctctagtcattgttgatgtttaacaggtaacagtgataattatttgtc
ttctcattaatgaagaaaaggattattaatcatagaggGTGGAAGGCATC
TATGGGAA

Figure 4.17: UCSC in-silico PCR results for GCK primers

4.6.1 Primer optimization:

Primers were optimized using gradient PCR. PCR results were confirmed using gel electrophoresis.

Ladder GCK at 62^0 annealing temperature

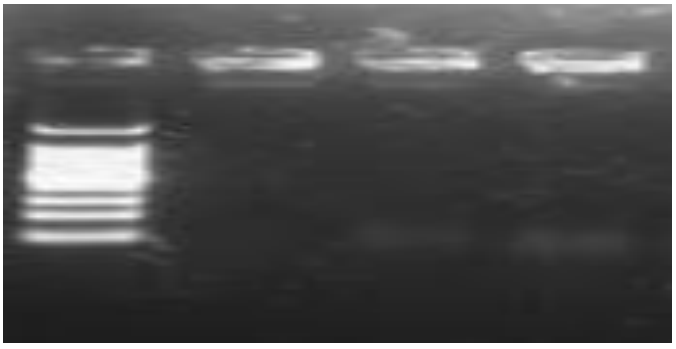


Figure 4.18: The gel electrophoresis result of gradient Pcr optimized GCK gene.

4.6.2 Real time PCR results:

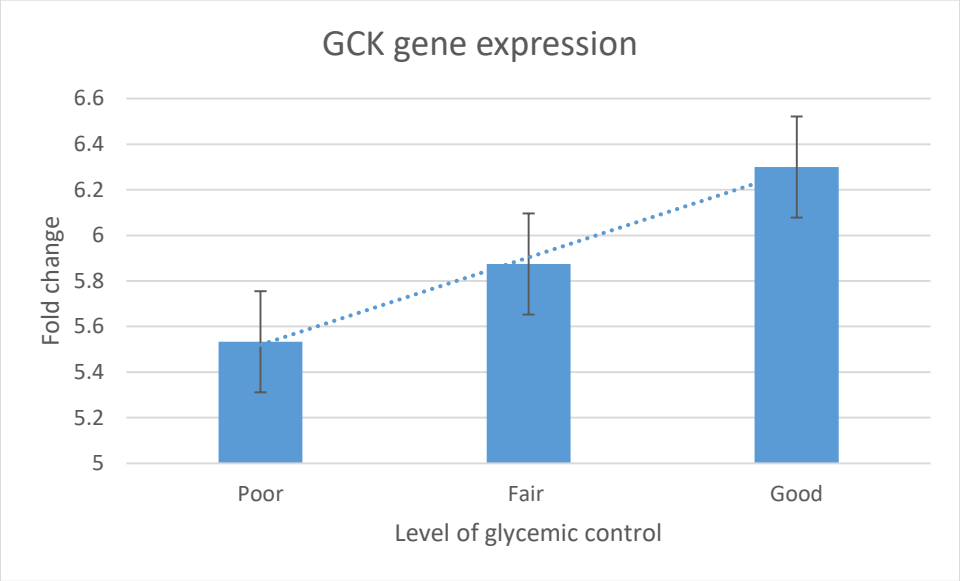


Figure 4.19: The expression of GCK gene in diabetic patient with poor glycemic control is significantly higher as compared to diabetic patients with good glycemic control with significant $p<0.0001$ and $n=10$

DISCUSSION

The study shows that machine learning (ML) models based on clinical and demographic data can predict HbA1c readings, which simplifies diabetes care by keeping glycemic levels under control. The ultimate goal is to improve glycemic management by accurately predicting HbA1c levels. HbA1c is a key biomarker of glycemic management and average blood glucose levels (Krhač & Lovrenčić, 2019). HbA1c levels must be accurately predicted in order to control diabetes.

Machine learning algorithms that predict HbA1c values are useful for diabetes management and early diagnosis of problems (Dankwa-Mullan et al., 2019; Kasula, 2023). After training numerous models, the Gradient Boosting Regression (GBR) technique outperforms them all in terms of accuracy. Jahani et.al found that machine learning regression models outperformed simple logistic regression models(Jahani & Mahdavi, 2016). The GBR model is the most effective at collecting variance in the dataset without introducing fitting noise. It exhibits the ideal combination of low Mean Squared Error (MSE) and high R^2.

Random Forest Regression also performs well, with slightly lower R^2 and somewhat higher MSE than GBR. The high performance of linear regression indicates that the data's relationships may be linear. In contrast, Neural Network Regression (NNR) and Support Vector Regression (SVR) perform significantly worse on both criteria. This could mean that SVR's sensitivity to feature scaling and parameter tuning is more important, or that the default settings are inappropriate for the data. Previous research has also shown that Gradient Boosting Regression outperforms other methods in terms of performance, accuracy, and effectiveness for predicting diabetes diagnosis and care. The highest performance of GBR algorithms indicates intricate connections between the features utilized to forecast HbA1c levels. This model's accuracy can be further refined.

HbA1c prediction using ML models has important clinical consequences. These models can help healthcare providers anticipate exact HbA1c levels based on clinical data from patients. Accurate predictions made using machine learning models help to improve glycemic control and spot problems early on, allowing specialists to recommend proactive management options to patients. However, despite the numerous benefits, ML-driven prediction and diabetes management present several obstacles. The most difficult aspect of applying ML models is accounting for individual

variation; these models do not adequately account for genetic and clinical variability between people. It is necessary to link these ML systems with the healthcare infrastructure and take into account external environmental elements that influence patients' health (Domingo-Lopez et al., 2022). Furthermore, ML models raise various ethical problems about patients' data security, which need to be addressed during model design.

Although the GBR model is accurate, the study has certain drawbacks. The training dataset is modest, and increasing it would improve accuracy even more. The data on self-management behavior were gathered through face-to-face interviews, which may have introduced biases. Additional study should concentrate on incorporating other data sources, such as genetic data and continuous glucose monitoring data. The model's practical utility should be evaluated by assessing its effect on clinical outcomes.

In the context of diabetes care, glycemic control modulation is of essential importance, as it has a direct impact on the expression of glucose metabolism genes. One such gene, glucokinase (GCK), regulates glucose homeostasis by promoting the phosphorylation of glucose to glucose-6-phosphate, a vital stage in glycolysis. This study investigates the differential expression of the GCK gene in diabetes individuals with differing degrees of glycemic control, as defined by HbA1c levels.

The bar graph results show a clear trend: GCK gene expression is significantly higher in individuals with strong glycemic control (lower HbA1c levels) versus those with poor glycemic control (higher HbA1c levels). This observation is consistent with the known role of GCK in glucose metabolism. In settings of good glycemic control, effective glucose metabolism is required, hence enhanced GCK expression supports the increased demand for glucose phosphorylation.

By adding GCK expression data, the prediction accuracy of HbA1c levels improves, providing a more sophisticated tool for monitoring and managing diabetic patients. The incorporation of gene expression data into predictive models marks a huge step forward in individualized diabetes care. Identifying specific molecular markers, such as GCK, allows us to better understand disease causes and customize therapies more precisely.

CONCLUSION

Finally, the study emphasizes Gradient Boosting Regression's strong performance in predicting HbA1c levels, indicating that it has the potential to enhance glycemic management in diabetics. The application of artificial intelligence and machine learning in diabetes treatment has great promise for improving patient outcomes and healthcare efficiency. To fully achieve these benefits, additional research should focus on establishing strong models, tackling practical implementation issues, and controlling ethical concerns such data privacy and bias. It is also critical to expand databases to include a wide range of demographic and clinical factors, as well as to ensure that healthcare practitioners are properly trained to use these tools. With these initiatives, AI-driven methods to diabetes treatment have the potential to fundamentally revolutionize the field, resulting in more tailored care and better patient outcomes. This work also emphasizes the significance of GCK gene expression as a measure of glycemic control in diabetes patients. The considerable changes in expression levels across different levels of glycemic control highlight the gene's significance in glucose homeostasis. The inclusion of these findings into a predictive model for HbA1c levels improves the model's accuracy while also opening up new paths for tailored diabetes care techniques.

References

[1] Abbasi, A., Peelen, L. M., Corpeleijn, E., van der Schouw, Y. T., Stolk, R. P., Spijkerman, A. M. W., van der A, D. L., Moons, K. G. M., Navis, G., Bakker, S. J. L., & Beulens, J. W. J. (2012). Prediction models for risk of developing type 2 diabetes: Systematic literature search and independent external validation study. *BMJ (Clinical Research Ed.), 345*, e5900. https://doi.org/10.1136/bmj.e5900

[2] Abd-Elraouf, M. S. E.-D. (2020). Factors Affecting Glycemic Control in Type II Diabetic Patients. *The Egyptian Journal of Hospital Medicine*, *81*(2), 1457–1461. https://doi.org/10.21608/ejhm.2020.114454

[3] Afroz-Hossain, A., Dawkins, M., & Myers, A. K. (2019). Sleep and Environmental Factors Affecting Glycemic Control in People with Type 2 Diabetes Mellitus. *Current Diabetes Reports*, *19*(7), 40. https://doi.org/10.1007/s11892-019-1159-9

[4] Ahola, A. J., & Groop, P.-H. (2013). Barriers to self-management of diabetes. *Diabetic Medicine: A Journal of the British Diabetic Association*, *30*(4), 413–420. https://doi.org/10.1111/dme.12105

[5] Akinci D'Antonoli, T. (2020). Ethical considerations for artificial intelligence: An overview of the current radiology landscape. *Diagnostic and Interventional Radiology (Ankara, Turkey)*, *26*(5), 504–511. https://doi.org/10.5152/dir.2020.19279

[6] Al-Khawaldeh, O. A., Al-Hassan, M. A., & Froelicher, E. S. (2012). Self-efficacy, self-management, and glycemic control in adults with type 2 diabetes mellitus. *Journal of Diabetes and Its Complications*, *26*(1), 10–16. https://doi.org/10.1016/j.jdiacomp.2011.11.002

[7] Almutairi, N., Hosseinzadeh, H., & Gopaldasani, V. (2020). The effectiveness of patient activation intervention on type 2 diabetes mellitus glycemic control and self-management behaviors: A systematic review of RCTs. *Primary Care Diabetes*, *14*(1), 12–20. https://doi.org/10.1016/j.pcd.2019.08.009

[8] Amann, J., Blasimme, A., Vayena, E., Frey, D., Madai, V. I., & Precise4Q consortium. (2020). Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, *20*(1), 310. https://doi.org/10.1186/s12911-020-01332-6

[9] Amarasekara, A. A. T. D., Fongkaew, W., Wimalasekera, S. W., Turale, S., & Chanprasit, C. (2015). Cross-sectional study of glycemic control among adults with type 2 diabetes. *Nursing & Health Sciences*, *17*(2), 223–228. https://doi.org/10.1111/nhs.12179

[10] Ansari, R. M., Harris, M. F., Hosseinzadeh, H., & Zwar, N. (2023). Application of Artificial Intelligence in Assessing the Self-Management Practices of Patients with Type 2 Diabetes. *Healthcare (Basel, Switzerland)*, *11*(6), 903. https://doi.org/10.3390/healthcare11060903

[11] Antonetti, D. A., Silva, P. S., & Stitt, A. W. (2021). Current understanding of the molecular and cellular pathology of diabetic retinopathy. *Nature Reviews. Endocrinology*, *17*(4), 195–206. https://doi.org/10.1038/s41574-020-00451-4

[12] Banday, M. Z., Sameer, A. S., & Nissar, S. (2020). Pathophysiology of diabetes: An overview. *Avicenna Journal of Medicine*, *10*(4), 174–188. https://doi.org/10.4103/ajm.ajm_53_20

[13] Bekele, B. B., Negash, S., Bogale, B., Tesfaye, M., Getachew, D., Weldekidan, F., & Balcha, B. (2021). Effect of diabetes self-management education (DSME) on glycated

hemoglobin (HbA1c) level among patients with T2DM: Systematic review and meta-analysis of randomized controlled trials. *Diabetes & Metabolic Syndrome*, *15*(1), 177–185. https://doi.org/10.1016/j.dsx.2020.12.030

[14] Benhalima, K., Standl, E., & Mathieu, C. (2011). The importance of glycemic control: How low should we go with HbA1c? Start early, go safe, go low. *Journal of Diabetes and Its Complications*, *25*(3), 202–207. https://doi.org/10.1016/j.jdiacomp.2010.03.002

[15] Boniol, M., Dragomir, M., Autier, P., & Boyle, P. (2017). Physical activity and change in fasting glucose and HbA1c: A quantitative meta-analysis of randomized trials. *Acta Diabetologica*, *54*(11), 983–991. https://doi.org/10.1007/s00592-017-1037-3

[16] Chaki, J., Thillai Ganesh, S., Cidham, S. K., & Ananda Theertan, S. (2022). Machine learning and artificial intelligence based Diabetes Mellitus detection and self-management: A systematic review. *Journal of King Saud University - Computer and Information Sciences*, *34*(6, Part B), 3204–3225. https://doi.org/10.1016/j.jksuci.2020.06.013

[17] Che, M., Zhou, Q., Lin, W., Yang, Y., Sun, M., Liu, X., Liu, H., & Zhang, C. (2023). Healthy Lifestyle Score and Glycemic Control in Type 2 Diabetes Mellitus Patients: A City-Wide Survey in China. *Healthcare*, *11*(14), 2037. https://doi.org/10.3390/healthcare11142037

[18] Chen, L., Magliano, D. J., & Zimmet, P. Z. (2011). The worldwide epidemiology of type 2 diabetes mellitus—Present and future perspectives. *Nature Reviews. Endocrinology*, *8*(4), 228–236. https://doi.org/10.1038/nrendo.2011.183

[19] Cheng, L. J., Wang, W., Lim, S. T., & Wu, V. X. (2019). Factors associated with glycaemic control in patients with diabetes mellitus: A systematic literature review. *Journal of Clinical Nursing*, *28*(9–10), 1433–1450. https://doi.org/10.1111/jocn.14795

[20] Cichosz, S. L., Johansen, M. D., & Hejlesen, O. (2015). Toward Big Data Analytics: Review of Predictive Models in Management of Diabetes and Its Complications. *Journal of Diabetes Science and Technology*, *10*(1), 27–34. https://doi.org/10.1177/1932296815611680

[21] Cunningham, A. T., Crittendon, D. R., White, N., Mills, G. D., Diaz, V., & LaNoue, M. D. (2018). The effect of diabetes self-management education on HbA1c and quality of life in African-Americans: A systematic review and meta-analysis. *BMC Health Services Research*, *18*(1), 367. https://doi.org/10.1186/s12913-018-3186-7

[22] Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., & Rhee, K. (2019). Transforming Diabetes Care Through Artificial Intelligence: The Future Is Here. *Population Health Management*, *22*(3), 229–242. https://doi.org/10.1089/pop.2018.0129

[23] Deepa, R., & Sivasamy, A. (2023). Advancements in early detection of diabetes and diabetic retinopathy screening using artificial intelligence. *AIP Advances*, *13*(11), 115307. https://doi.org/10.1063/5.0172226

[24] Del Parigi, A., Tang, W., Liu, D., Lee, C., & Pratley, R. (2019). Machine Learning to Identify Predictors of Glycemic Control in Type 2 Diabetes: An Analysis of Target HbA1c Reduction Using Empagliflozin/Linagliptin Data. *Pharmaceutical Medicine*, *33*(3), 209–217. https://doi.org/10.1007/s40290-019-00281-4

[25] Demir, S., Nawroth, P. P., Herzig, S., & Ekim Üstünel, B. (2021). Emerging Targets in Type 2 Diabetes and Diabetic Complications. *Advanced Science (Weinheim,*

Baden-Wurttemberg, Germany), *8*(18), e2100275. https://doi.org/10.1002/advs.202100275

[26] Diaz C, J. L., Villa-Tamayo, M. F., Moscoso-Vasquez, M., & Colmegna, P. (2023). Simulation-driven optimization of insulin therapy profiles in a commercial hybrid closed-loop system. *Computer Methods and Programs in Biomedicine*, *242*, 107830. https://doi.org/10.1016/j.cmpb.2023.107830

[27] Dillmann, W. H. (2019). Diabetic Cardiomyopathy. *Circulation Research*, *124*(8), 1160–1162. https://doi.org/10.1161/CIRCRESAHA.118.314665

[28] Domingo-Lopez, D. A., Lattanzi, G., H J Schreiber, L., Wallace, E. J., Wylie, R., O'Sullivan, J., Dolan, E. B., & Duffy, G. P. (2022). Medical devices, smart drug delivery, wearables and technology for the treatment of Diabetes Mellitus. *Advanced Drug Delivery Reviews*, *185*, 114280. https://doi.org/10.1016/j.addr.2022.114280

[29] Ellahham, S. (2020). Artificial Intelligence: The Future for Diabetes Care. *The American Journal of Medicine*, *133*(8), 895–900. https://doi.org/10.1016/j.amjmed.2020.03.033

[30] Fan, Y., Long, E., Cai, L., Cao, Q., Wu, X., & Tong, R. (2021). Machine Learning Approaches to Predict Risks of Diabetic Complications and Poor Glycemic Control in Nonadherent Type 2 Diabetes. *Frontiers in Pharmacology*, *12*, 665951. https://doi.org/10.3389/fphar.2021.665951

[31] Farajollahi, B., Mehmannavaz, M., Mehrjoo, H., Moghbeli, F., & Sayadi, M. J. (2021). Diabetes Diagnosis Using Machine Learning. *Frontiers in Health Informatics*, *10*(1), Article 1. https://doi.org/10.30699/fhi.v10i1.267

[32] Fatima, W., Khan, F., Ali, A., & Verma, A. K. (2020, October 1). *Association of GCK Gene Promoter Polymorphism and their Role in its MRNA Expression among Type 2 Diabetes Mellitus Patients. | Journal of Clinical & Diagnostic Research | EBSCOhost*. https://doi.org/10.7860/JCDR/2020/42876.14135

[33] Galicia-Garcia, U., Benito-Vicente, A., Jebari, S., Larrea-Sebal, A., Siddiqi, H., Uribe, K. B., Ostolaza, H., & Martín, C. (2020). Pathophysiology of Type 2 Diabetes Mellitus. *International Journal of Molecular Sciences*, *21*(17), 6275. https://doi.org/10.3390/ijms21176275

[34] Haeusler, R. A., Camastra, S., Astiarraga, B., Nannipieri, M., Anselmino, M., & Ferrannini, E. (2015). Decreased expression of hepatic glucokinase in type 2 diabetes. *Molecular Metabolism*, *4*(3), 222–226. https://doi.org/10.1016/j.molmet.2014.12.007

[35] Haghighatpanah, M., Nejad, A. S. M., Haghighatpanah, M., Thunga, G., & Mallayasamy, S. (2018). Factors that Correlate with Poor Glycemic Control in Type 2 Diabetes Mellitus Patients with Complications. *Osong Public Health and Research Perspectives*, *9*(4), 167–174. https://doi.org/10.24171/j.phrp.2018.9.4.05

[36] Hakeem, R., & Fawwad, A. (2010). Diabetes in Pakistan: Epidemiology, Determinants and Prevention. *Journal of Diabetology*, *1*(3), 3.

[37] Heisler, M., Piette, J. D., Spencer, M., Kieffer, E., & Vijan, S. (2005). The relationship between knowledge of recent HbA1c values and diabetes care understanding and self-management. *Diabetes Care*, *28*(4), 816–822. https://doi.org/10.2337/diacare.28.4.816

[38] Houle, J., Lauzier-Jobin, F., Beaulieu, M.-D., Meunier, S., Coulombe, S., Côté, J., Lespérance, F., Chiasson, J.-L., Bherer, L., & Lambert, J. (2016). Socioeconomic status

and glycemic control in adult patients with type 2 diabetes: A mediation analysis. *BMJ Open Diabetes Research & Care*, *4*(1), e000184. https://doi.org/10.1136/bmjdrc-2015-000184

[39] Hu, C., Zhang, R., Wang, C., Yu, W., Lu, J., Ma, X., Wang, J., Jiang, F., Tang, S., Bao, Y., Xiang, K., & Jia, W. (2010). Effects of GCK, GCKR, G6PC2 and MTNR1B Variants on Glucose Metabolism and Insulin Secretion. *PLOS ONE*, *5*(7), e11761. https://doi.org/10.1371/journal.pone.0011761

[40] Iparraguirre-Villanueva, O., Espinola-Linares, K., Flores Castañeda, R. O., & Cabanillas-Carbonell, M. (2023). Application of Machine Learning Models for Early Detection and Accurate Classification of Type 2 Diabetes. *Diagnostics (Basel, Switzerland)*, *13*(14), 2383. https://doi.org/10.3390/diagnostics13142383

[41] Ishaque, A., Shahzad, F., Muhammad, F. H., Usman, Y., & Ishaque, Z. (2016). Diabetes risk assessment among squatter settlements in Pakistan: A cross-sectional study. *Malaysian Family Physician: The Official Journal of the Academy of Family Physicians of Malaysia*, *11*(2–3), 9–15.

[42] Iynedjian, P. B. (2009). Molecular Physiology of Mammalian Glucokinase. *Cellular and Molecular Life Sciences*, *66*(1), 27. https://doi.org/10.1007/s00018-008-8322-9

[43] Jahani, M., & Mahdavi, M. (2016). Comparison of Predictive Models for the Early Diagnosis of Diabetes. *Healthcare Informatics Research*, *22*(2), 95–100. https://doi.org/10.4258/hir.2016.22.2.95

[44] Jun, J., Lee, Y., & Oh, Y. jin. (2016). A Factor Analysis Study on Blood Glucose Control in Diabetics Mellitus Patients(1):Focus on Blood Glucose Control and Lifestyle Factors. *Korean Journal of Community Nutrition*, *14*(2), 236–244.

[45] Juneja, D., Gupta, A., & Singh, O. (2022). Artificial intelligence in critically ill diabetic patients: Current status and future prospects. *Artificial Intelligence in Gastroenterology*, *3*(2), 66–79. https://doi.org/10.35712/aig.v3.i2.66

[46] Kasula, B. Y. (2023). Machine Learning Applications in Diabetic Healthcare: A Comprehensive Analysis and Predictive Modeling. *International Numeric Journal of Machine Learning and Robots*, *7*(7), Article 7. https://injmr.com/index.php/fewfewf/article/view/19

[47] Katagiri, H., Asano, T., Ishihara, H., Inukai, K., Anai, M., Miyazaki, J., Tsukuda, K., Kikuchi, M., Yazaki, Y., & Oka, Y. (1992). Nonsense mutation of glucokinase gene in late-onset non-insulin-dependent diabetes mellitus. *Lancet (London, England)*, *340*(8831), 1316–1317. https://doi.org/10.1016/0140-6736(92)92494-z

[48] Kautzky-Willer, A., Leutner, M., & Harreiter, J. (2023). Sex differences in type 2 diabetes. *Diabetologia*, *66*(6), 986–1002. https://doi.org/10.1007/s00125-023-05891-x

[49] Kerr, D., Abbasi, M., Bevier, W., Glantz, N., Larez, A., & Sabharwal, A. (2024). Patterns of Timing and Intensity of Physical Activity and HbA1c Levels in Hispanic/Latino Adults With or at Risk of Type 2 Diabetes. *Journal of Diabetes Science and Technology*, *18*(1), 106–112. https://doi.org/10.1177/19322968221105531

[50] Khan, M. A. B., Hashim, M. J., King, J. K., Govender, R. D., Mustafa, H., & Al Kaabi, J. (2020). Epidemiology of Type 2 Diabetes—Global Burden of Disease and

Forecasted Trends. *Journal of Epidemiology and Global Health*, *10*(1), 107–111. https://doi.org/10.2991/jegh.k.191028.001

[51] Kodakandla, K., Maddela, G., & Shahid, M. (n.d.). *Factors influencing sleep quality and its impact on glycemic control in patients with type II diabetes mellitus—A hospital based cross sectional study*.

[52] Kohnert, K.-D., Heinke, P., Vogt, L., & Salzsieder, E. (2015). Utility of different glycemic control metrics for optimizing management of diabetes. *World Journal of Diabetes*, *6*(1), 17–29. https://doi.org/10.4239/wjd.v6.i1.17

[53] Kolb, H., & Martin, S. (2017). Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. *BMC Medicine*, *15*(1), 131. https://doi.org/10.1186/s12916-017-0901-x

[54] Krhač, M., & Lovrenčić, M. V. (2019). Update on biomarkers of glycemic control. *World Journal of Diabetes*, *10*(1), 1–15. https://doi.org/10.4239/wjd.v10.i1.1

[55] Lai, H., Huang, H., Keshavjee, K., Guergachi, A., & Gao, X. (2019). Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocrine Disorders*, *19*(1), 101. https://doi.org/10.1186/s12902-019-0436-6

[56] Leong, A., & Wheeler, E. (2018). Genetics of HbA1c: A case study in clinical translation. *Current Opinion in Genetics & Development*, *50*, 79–85. https://doi.org/10.1016/j.gde.2018.02.008

[57] Li, J., Huang, J., Zheng, L., & Li, X. (2020). Application of Artificial Intelligence in Diabetes Education and Management: Present Status and Promising Prospect. *Frontiers in Public Health*, *8*, 173. https://doi.org/10.3389/fpubh.2020.00173

[58] Liu, Y. (2020). Artificial Intelligence-Based Neural Network for the Diagnosis of Diabetes: Model Development. *JMIR Medical Informatics*, *8*(5), e18682. https://doi.org/10.2196/18682

[59] Lorenzo-Almorós, A., Cepeda-Rodrigo, J. M., & Lorenzo, Ó. (2022). Diabetic cardiomyopathy. *Revista Clinica Espanola*, *222*(2), 100–111. https://doi.org/10.1016/j.rceng.2019.10.012

[60] Marques, I. P., Madeira, M. H., Messias, A. L., Martinho, A. C.-V., Santos, T., Sousa, D. C., Figueira, J., & Cunha-Vaz, J. (2021). Different retinopathy phenotypes in type 2 diabetes predict retinopathy progression. *Acta Diabetologica*, *58*(2), 197–205. https://doi.org/10.1007/s00592-020-01602-9

[61] Martínez, Y. V., Campbell, S. M., Hann, M., & Bower, P. (2016). The individual contribution and relative importance of self-management and quality of care on glycaemic control in type 2 diabetes. *Salud Publica De Mexico*, *58*(4), 404–411. https://doi.org/10.21149/spm.v58i4.8020

[62] Matschinsky, F. M. (1990). Glucokinase as glucose sensor and metabolic signal generator in pancreatic beta-cells and hepatocytes. *Diabetes*, *39*(6), 647–652. https://doi.org/10.2337/diab.39.6.647

[63] Matschinsky, F. M., Zelent, B., Doliba, N., Li, C., Vanderkooi, J. M., Naji, A., Sarabu, R., & Grimsby, J. (2011). Glucokinase Activators for Diabetes Therapy. *Diabetes Care*, *34*(Supplement_2), S236–S243. https://doi.org/10.2337/dc11-s236

[64] Matsutani, A., Janssen, R., Donis-Keller, H., & Permutt, M. A. (1992). A polymorphic (CA)n repeat element maps the human glucokinase gene (GCK) to

chromosome 7p. *Genomics*, *12*(2), 319–325. https://doi.org/10.1016/0888-7543(92)90380-b

[65] Milluzzo, A., Maugeri, A., Barchitta, M., Sciacca, L., & Agodi, A. (2021). Epigenetic Mechanisms in Type 2 Diabetes Retinopathy: A Systematic Review. *International Journal of Molecular Sciences*, *22*(19), 10502. https://doi.org/10.3390/ijms221910502

[66] Muller, Y. L., Piaggi, P., Hoffman, D., Huang, K., Gene, B., Kobes, S., Thearle, M. S., Knowler, W. C., Hanson, R. L., Baier, L. J., & Bogardus, C. (2014). Common genetic variation in the glucokinase gene (GCK) is associated with type 2 diabetes and rates of carbohydrate oxidation and energy expenditure. *Diabetologia*, *57*(7), 1382–1390. https://doi.org/10.1007/s00125-014-3234-8

[67] Murdoch, B. (2021). Privacy and artificial intelligence: Challenges for protecting health information in a new era. *BMC Medical Ethics*, *22*(1), 122. https://doi.org/10.1186/s12910-021-00687-3

[68] Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Frontiers in Surgery*, *9*, 862322. https://doi.org/10.3389/fsurg.2022.862322

[69] Nelson, K. M., McFarland, L., & Reiber, G. (2007). Factors influencing disease self-management among veterans with diabetes and poor glycemic control. *Journal of General Internal Medicine*, *22*(4), 442–447. https://doi.org/10.1007/s11606-006-0053-8

[70] Pamungkas, R. A., Hadijah, St., Mayasari, A., & Nusdin, N. (2017). FACTORS ASSOCIATED WITH POOR GLYCEMIC CONTROL AMONG TYPE 2 DIABETES MELLITUS IN INDONESIA. *Belitung Nursing Journal*, *3*(3), 272–280. https://doi.org/10.33546/bnj.61

[71] Paolillo, S., Marsico, F., Prastaro, M., Renga, F., Esposito, L., De Martino, F., Di Napoli, P., Esposito, I., Ambrosio, A., Ianniruberto, M., Mennella, R., Paolillo, R., & Gargiulo, P. (2019). Diabetic Cardiomyopathy: Definition, Diagnosis, and Therapeutic Implications. *Heart Failure Clinics*, *15*(3), 341–347. https://doi.org/10.1016/j.hfc.2019.02.003

[72] Reddy, S., Fox, J., & Purohit, M. P. (2019). Artificial intelligence-enabled healthcare delivery. *Journal of the Royal Society of Medicine*, *112*(1), 22–28. https://doi.org/10.1177/0141076818815510

[73] Riangkam, C., Sriyuktasuth, A., Pongthavornkamol, K., Kusakunniran, W., & Sriwijitkamol, A. (2021). Effects of a mobile health diabetes self-management program on HbA1C, self-management and patient satisfaction in adults with uncontrolled type 2 diabetes: A randomized controlled trial. *Journal of Health Research*, *36*(5), 878–888. https://doi.org/10.1108/JHR-02-2021-0126

[74] Sagoo, M. K., & Gnudi, L. (2020). Diabetic Nephropathy: An Overview. *Methods in Molecular Biology (Clifton, N.J.)*, *2067*, 3–7. https://doi.org/10.1007/978-1-4939-9841-8_1

[75] Saltiel, A. R., & Kahn, C. R. (2001). Insulin signalling and the regulation of glucose and lipid metabolism. *Nature*, *414*(6865), 799–806. https://doi.org/10.1038/414799a

[76] Samsu, N. (2021). Diabetic Nephropathy: Challenges in Pathogenesis, Diagnosis, and Treatment. *BioMed Research International*, *2021*, 1497449. https://doi.org/10.1155/2021/1497449

[77] Sarnowski, C., Leong, A., Raffield, L. M., Wu, P., de Vries, P. S., DiCorpo, D., Guo, X., Xu, H., Liu, Y., Zheng, X., Hu, Y., Brody, J. A., Goodarzi, M. O., Hidalgo, B. A., Highland, H. M., Jain, D., Liu, C.-T., Naik, R. P., O'Connell, J. R., … National Heart, Lung, and Blood Institute TOPMed Consortium. (2019). Impact of Rare and Common Genetic Variants on Diabetes Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for Precision Medicine Program. *American Journal of Human Genetics*, *105*(4), 706–718. https://doi.org/10.1016/j.ajhg.2019.08.010

[78] Selby, N. M., & Taal, M. W. (2020). An updated overview of diabetic nephropathy: Diagnosis, prognosis, treatment goals and latest guidelines. *Diabetes, Obesity & Metabolism*, *22 Suppl 1*, 3–15. https://doi.org/10.1111/dom.14007

[79] Singh, P., Singh, S. P., & Singh, D. S. (2019). AN INTRODUCTION AND REVIEW ON MACHINE LEARNING APPLICATIONS IN MEDICINE AND HEALTHCARE. *2019 IEEE Conference on Information and Communication Technology*, 1–6. https://doi.org/10.1109/CICT48419.2019.9066250

[80] Soranzo, N. (2011). Genetic determinants of variability in glycated hemoglobin (HbA(1c)) in humans: Review of recent progress and prospects for use in diabetes care. *Current Diabetes Reports*, *11*(6), 562–569. https://doi.org/10.1007/s11892-011-0232-9

[81] Steele, A. M., Wensley, K. J., Ellard, S., Murphy, R., Shepherd, M., Colclough, K., Hattersley, A. T., & Shields, B. M. (2013). Use of HbA1c in the Identification of Patients

with Hyperglycaemia Caused by a Glucokinase Mutation: Observational Case Control Studies. *PLOS ONE*, *8*(6), e65326. https://doi.org/10.1371/journal.pone.0065326

[82] Stoffel, M., Patel, P., Lo, Y. M., Hattersley, A. T., Lucassen, A. M., Page, R., Bell, J. I., Bell, G. I., Turner, R. C., & Wainscoat, J. S. (1992). Missense glucokinase mutation in maturity-onset diabetes of the young and mutation screening in late-onset diabetes. *Nature Genetics*, *2*(2), 153–156. https://doi.org/10.1038/ng1092-153

[83] Su, D., Michaud, T. L., Estabrooks, P., Schwab, R. J., Eiland, L. A., Hansen, G., DeVany, M., Zhang, D., Li, Y., Pagán, J. A., & Siahpush, M. (2019). Diabetes Management Through Remote Patient Monitoring: The Importance of Patient Activation and Engagement with the Technology. *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association*, *25*(10), 952–959. https://doi.org/10.1089/tmj.2018.0205

[84] Tan, Y., Zhang, Z., Zheng, C., Wintergerst, K. A., Keller, B. B., & Cai, L. (2020). Mechanisms of diabetic cardiomyopathy and potential therapeutic strategies: Preclinical and clinical evidence. *Nature Reviews. Cardiology*, *17*(9), 585–607. https://doi.org/10.1038/s41569-020-0339-2

[85] Tay, J. H. T., Jiang, Y., Hong, J., He, H., & Wang, W. (2021). Effectiveness of lay-led, group-based self-management interventions to improve glycated hemoglobin (HbA1c), self-efficacy, and emergency visit rates among adults with type 2 diabetes: A systematic review and meta-analysis. *International Journal of Nursing Studies*, *113*, 103779. https://doi.org/10.1016/j.ijnurstu.2020.103779

[86] Taylor, R. (2013). Type 2 diabetes: Etiology and reversibility. *Diabetes Care*, *36*(4), 1047–1055. https://doi.org/10.2337/dc12-1805

[87] Utinane, Z., & Tīcmane, G. (n.d.). THE IMPACT OF PHYSICAL ACTIVITY ON HEMOGLOBIN A1c (HbA1c) LEVEL. *DAUGAVPILS UNIVERSITĀTES 63. STARPTAUTISKĀS ZINĀTNISKĀS KONFERENCES RAKSTU KRĀJUMS*, 34.

[88] van der Heijden, A. A., Nijpels, G., Badloe, F., Lovejoy, H. L., Peelen, L. M., Feenstra, T. L., Moons, K. G. M., Slieker, R. C., Herings, R. M. C., Elders, P. J. M., & Beulens, J. W. (2020). Prediction models for development of retinopathy in people with type 2 diabetes: Systematic review and external validation in a Dutch primary care setting. *Diabetologia*, *63*(6), 1110–1119. https://doi.org/10.1007/s00125-020-05134-3

[89] Wan, T. T. H. (2021). Predictive Analytics for the KMAP-O Model in Design and Evaluation of Diabetes Care Management Research. *Health Services Research and Managerial Epidemiology*, *8*, 23333928211023220. https://doi.org/10.1177/23333928211023220

[90] Wang, H., Liu, L., Zhao, J., Cui, G., Chen, C., Ding, H., & Wang, D. W. (2013). Large scale meta-analyses of fasting plasma glucose raising variants in GCK, GCKR, MTNR1B and G6PC2 and their impacts on type 2 diabetes mellitus risk. *PloS One*, *8*(6), e67665. https://doi.org/10.1371/journal.pone.0067665

[91] Willaing, I., Rogvi, S.-A., Bøgelund, M., Almdal, T., & Schiøtz, M. (2013). Recall of HbA1c and self-management behaviours, patient activation, perception of care and diabetes distress in Type 2 diabetes. *Diabetic Medicine: A Journal of the British Diabetic Association*, *30*(4), e139-142. https://doi.org/10.1111/dme.12121

[92] Xing, X.-Y., Wang, X.-Y., Fang, X., Xu, J.-Q., Chen, Y.-J., Xu, W., Wang, H.-D., Liu, Z.-R., & Tao, S.-S. (2022). Glycemic control and its influencing factors in type 2

diabetes patients in Anhui, China. *Frontiers in Public Health*, *10*, 980966. https://doi.org/10.3389/fpubh.2022.980966

[93] Yang. (n.d.). *Analysis on Influencing Factors for Achieving Glycemic Control in Patients with Type 2 Diabetes Mellitus in Shanghai*. Retrieved May 19, 2024, from https://www.jeom.org/en/article/id/8a250ab3-8566-4bee-9db8-cb773001faba?viewType=HTML

[94] Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nature Reviews Endocrinology*, *14*(2), 88–98. https://doi.org/10.1038/nrendo.2017.151

turnitin

Prof. Dr. Attya Bhatti
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Research Dte Reseach Dte
Assignment title:	Biology
Submission title:	Zanib Thesis
File name:	Zainab_Thesis_Plg.docx
File size:	929.91K
Page count:	65
Word count:	15,916
Character count:	93,532
Submission date:	21-Aug-2024 03:01AM (UTC-0700)
Submission ID:	2429495797

Abstract

Diabetes mellitus is a chronic metabolic disorder that affects millions of people worldwide and has been related to serious complications such as cardiovascular disease, neuropathy, nephropathy, and retinopathy. Effective diabetes care is dependent on maintaining glycemic control, which is frequently measured using glycated hemoglobin (HbA1c) values. HbA1c is an important biomarker that represents average blood glucose levels over the previous two to three months and is used to determine long-term glycemic management and the risk of problems.

The goal of this research is to create a reliable and accurate HbA1c prediction model utilizing machine learning (ML) technologies. The objectives include assessing clinical and demographic data to find significant variables, creating and comparing multiple machine learning models, and determining the best effective model for HbA1c prediction. In addition, the study investigates the clinical consequences of precise HbA1c forecasts, as well as the obstacles and ethical problems involved with machine learning in health care.

The study indicated that Gradient Boosting Regression (GBR) outperformed other models, with the lowest Mean Squared Error (MSE) and greatest R^2 values. Random Forest Regression also fared well, although Neural Network Regression (NNR) and Support Vector Regression (SVR) were less effective due to their sensitivity to feature scaling and parameter adjustment. Accurate HbA1c forecasts can assist healthcare practitioners anticipate levels and enhance glycemic control, but issues like individual variability and data security must be addressed.

Development of prediction model for the level of glycemic control as a measure of diabetes self-management

Zanib Thesis

Prof. Dr. Attya Bhatti
Dept. of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB), NUST Islamabad

ORIGINALITY REPORT

9%	2%	9%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Seyed Arash Forough, Fereidoon Moghadas Nejad, Ali Khodaii. "Development of a predictive model for the compressive relaxation modulus of asphalt mixtures", Road Materials and Pavement Design, 2015 Publication	2%
2	Yosefa Birati, Joan R. Bloch, Amy McKeever, Beth D. Chiatti. "Relationships Among Neighborhood Poverty, Access to Healthy Food, and Diabetes Self-Management in Women Who Received Perinatal Nurse Home Visits", Journal of Obstetric, Gynecologic & Neonatal Nursing, 2021 Publication	2%
3	Submitted to Malta College of Arts,Science and Technology Student Paper	1%
4	"Minutes of The 43rd General Assembly of The European Association for The Study of Diabetes", Diabetologia, 2008 Publication	1%

5	"Abstracts of the IDF Congress in Paris 2003", Diabetologia, 2003 Publication	1 %
6	Robert Fried, Richard M. Carlton. "Type 2 Diabetes - Cardiovascular and Related Complications and Evidence-Based Complementary Treatments", Routledge, 2018 Publication	1 %
7	uilis.unsyiah.ac.id Internet Source	1 %
8	Seth Kwabena Amponsah, Emmanuel Kwaku Ofori, Yashwant V. Pathak. "Current Trends in the Diagnosis and Management of Metabolic Disorders", CRC Press, 2023 Publication	1 %

Exclude quotes	Off	Exclude matches	< 1%
Exclude bibliography	Off		