COLLEGE OF COMPUTER AND INFORMATION SCIENCE

Academic Year 2025 – 2026

DIANA: A PREDICTIVE MODEL-BASED APPLICATION USING SELECTED BLOOD BIOMARKERS FOR CLUSTER-BASED IDENTIFICATION OF TYPE 2 DIABETES RISK IN MENOPAUSAL WOMEN

Marc Kennel M. ANGELES
Adrian Gabriell P. FRANCISCO
Sophia Nicole D. GREFALDO
Neoron G. LOPEZ

Thesis Adviser: Aurelia Sharlene Delos Santos

Submitted to the Faculty of Mapúa Malayan Colleges Laguna
In Partial Fulfillment of the Requirements for the degree of

Bachelor of Science in Computer Science

The thesis project attached hereto, entitled **"DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Cluster-Based Identification of Type 2 Diabetes Risk in Menopausal Women"** prepared and submitted by **Marc Kennel M. Angeles, Adrian Gabriell P. Francisco, Sophia Nicole D. Grefaldo** and **Neoron G. Lopez**, in partial fulfillment of the requirements for the degree of Bachelor of Science in Information Technology/Computer Science is hereby accepted.

Dennis A. Martillano
Panel Chair

Date Signed

Michael James F. Gnilo
Panel Member

Date Signed

Aurelia Sharlene O. Delos Santos
Adviser

Date Signed

Jonalyn G. Ebron
Program Chair, CS

Date Signed

Khristian G. Kikuchi
Dean, College of Computer and Information Science

Date Signed

Copyright

Mapúa Malayan Colleges Laguna, November 2025

The Adviser:

Aurelia Sharlene O. Delos Santos

The Authors:

Marc Kennel M. Angeles

Adrian Gabriell P. Francisco

Sophia Nicole D. Grefaldo

Neoron G. Lopez

Table of Contents

Copyright i

Table of Contents ii

List of Tables v

List of Figures vi

List of Equations vii

Definition of Terms ix

Introduction 1

Background of the Study 2

Objective of the Study 8

Rationale of the Study 11

Scope and Delimitations of the Study 14

Conceptual Framework 20

Review of Literature 23

Diabetes: Type 1 and Type 2 23

Diabetes Risk Factors 28

Menopausal 28

Biomarkers 30

Predictive Modeling 31

Data Visualization Techniques 33

Machine Learning 35
Synthesis 40
Methodology 43
Research Design 43
Research Locale 44
Population of the Study 45
Data Gathering Tools and Procedures 46
Software Methodology 49
Phase 1: Data Acquisition and Biomarker Preparation 50
Phase 2a: Model Development and Training 52
Phase 2b: Model Testing, Evaluation, and Comparison 54
Phase 3: Web Application Integration and Visualization Development 55
Phase 4: Technical Testing and Validation 56
Phase 5: Doctor's Evaluation 57
Data Analysis 58
References 69
Appendices 81

List of Tables

Table 1: Range Values for FBS, OGTT and HBA1C 30

Table 2: Range Values for FBS and RBS 31

Table 3: Dataset Composition: Blood Biomarkers and Demographic Variables 48

Table 4: Initial Clustering Label 66

Table 5: Data Dictionary 68

List of Figures

Figure 1: Conceptual Framework 22

Figure 2: General Prototyping Model 50

Figure 3: Data Acquisition and Biomarker Preparation Phase Flow 52

Figure 4: Model Development and Training 54

Figure 5: Model Testing, Evaluation and Comparison 55

Figure 6: Web Application Integration and Visualization Development Phase Flow 56

Figure 7: Technical and Validation Phase Flow 57

Figure 8: Doctor's Evaluation Phase Flow 58

List of Equations

Equation 1: Entropy of Y 59

Equation 2: The Conditional Entropy of Y given X 59

Equation 3: Information Gain of X with Respect to Y 60

Equation 4: Euclidean Distance Formula 62

Equation 5: Assignment Step Formula 62

Equation 6: Update Step Formula 63

Equation 7: Inertia or Objective Function J Formula 63

Equation 8: Formulas for Model Performance Evaluation 65

List of Appendices

Appendix A 81

Prototype of the Web Application Features 81

Appendix B 83

Letter of Requests 83

Appendix C 89

Example Test Results 89

Appendix D 90

Interview Pictures 90

Appendix E 93

Transcript of Interviews 93

Appendix F 117

Facebook Group: Usapang Perimenopause and Menopause 117

Approval Message for the admin through Facebook Messenger 118

Definition of Terms

Diabetes Mellitus (DM). a metabolic disorder characterized by high blood glucose levels. It includes several types such as Type 1 Diabetes, Type 2 Diabetes, Maturity-Onset Diabetes of the Young (MODY), gestational diabetes, neonatal diabetes, and diabetes secondary to other conditions or factors, such as hormonal disorders or prolonged use of steroids.

Type 1 Diabetes Mellitus (T1DM). A sub type of diabetes characterized by the destruction of pancreatic beta cells, usually caused by an autoimmune process. This destruction leads to little or no insulin production, resulting in a complete or near-complete lack of insulin in the body.

Type 2 Diabetes Mellitus (T2DM). A sub type of diabetes characterized by a gradual onset, in which a mismatch between insulin production and insulin sensitivity leads to a functional insulin deficiency. Insulin resistance, a key feature of T2DM, often arises from multiple factors, including obesity and aging.

Pre-diabetic. A condition in which blood glucose levels are higher than normal but not high enough to be classified as Type 2 Diabetes. It indicates an increased risk for developing diabetes and provides an opportunity for early intervention through lifestyle modification and monitoring of blood glucose

Hormonal Changes. alterations in the levels or activity of hormones in the body can affect various physiological processes. In the context of menopause, hormonal changes primarily refer to the decline in estrogen and progesterone levels, influencing metabolism, insulin sensitivity, and overall risk for conditions such as Type 2 Diabetes Mellitus (T2DM).

Insulin Restistance. A physiological condition in which the body's cells respond less effectively to insulin, reducing glucose uptake from the blood. This leads to higher circulating

blood glucose levels and increased insulin production, often contributing to the development of Type 2 Diabetes Mellitus (T2DM), particularly in populations with obesity, aging, or hormonal changes.

Fasting Blood Sugar (FBS). A laboratory test that measures the glucose level in a person's blood after an overnight fast, typically 8–12 hours. FBS is used to assess glycemic control and is a key biomarker in diagnosing and monitoring diabetes mellitus, including Type 2 Diabetes Mellitus (T2DM).

Hemoglobin A1c (Hba1C). A laboratory test that measures the average blood glucose levels over the past 2–3 months by determining the percentage of glucose bound to hemoglobin in red blood cells. HbA1c is commonly used to diagnose and monitor diabetes, providing an indicator of long-term glycemic control, including in patients with Type 2 Diabetes Mellitus (T2DM).

Lipid Profiles. A set of blood tests that measure the levels of specific lipids, including total cholesterol, high-density lipoprotein (HDL), low-density lipoprotein (LDL), and triglycerides. Lipid profiles are used to assess cardiovascular health and metabolic risk factors, including the risk of developing Type 2 Diabetes

Menopausal. The stage in a woman's life marked by the end of menstrual periods for at least twelve consecutive months, accompanied by hormonal changes, particularly a decline in estrogen levels. This transition, which includes perimenopause and postmenopause, is associated with physiological and metabolic changes that can affect insulin sensitivity and increase the risk of developing conditions such as Type 2 Diabetes Mellitus (T2DM).

Biomarkers. Biological indicators, such as Fasting Blood Sugar (FBS) and Hemoglobin A1c (HbA1c), Lipid Profiles, Age, BMI, Lifestyle, and Menopausal Status that are measurable

and used to assess the risk, presence, or progression of Type 2 Diabetes in an individual. Key variables in DIANA predictive model-based application to identify menopausal women at risk of type 2 diabetes

Predictive Modeling. Involves the use of statistical techniques, machine learning algorithms, or computational methods to analyze historical data and generate predictions about future outcomes or trends. In healthcare research, it is used to identify patterns, estimate risks, and forecast the likelihood of developing certain conditions based on measurable variables such as clinical biomarkers, demographic factors, or medical histories.

Machine Learning (ML). A branch of artificial intelligence that enables computer systems to learn patterns from data and make predictions or decisions without being explicitly programmed.

Cluster. A group of data points or individuals that share similar characteristics. In this study, clustering refers to grouping menopausal women based on their health and biomarker profiles to identify patterns and risk levels of Type 2 Diabetes.

Entropy. A measure of uncertainty or randomness in a dataset. In this study, entropy is used to quantify the amount of disorder in the biomarker data, helping to determine how informative each attribute is for predicting Type 2 Diabetes risk.

Information Gain (IG). Metrics are used to measure the effectiveness of a feature in reducing uncertainty in predicting an outcome. In this study, IG is computed for each biomarker to identify which attributes contribute most to predicting the risk of Type 2 Diabetes among menopausal women.

Introduction

Diabetes Mellitus (DM) is a chronic metabolic disorder characterized by persistently elevated blood glucose levels resulting from inadequate insulin production or decreased insulin sensitivity (Goyal et al., 2023). Globally, the burden of diabetes continues to rise. According to the International Diabetes Federation (IDF), more than 500 million adults were affected in 2021, with projections surpassing 700 million by 2045, reflecting a significant increase from 415 million in 2015. Among the different types, Type 2 Diabetes Mellitus (T2DM) is the most prevalent and is primarily associated with insulin resistance and impaired glucose regulation (Dhaliwal, 2025). The growing incidence of T2DM presents one of the foremost global public health challenges of the 21st century.

Nevertheless, the period surrounding menopause represents a critical stage in a woman's life that introduces an elevated risk for the development of metabolic disorders, including T2DM. The decline in ovarian function and subsequent decrease in estrogen levels significantly affect various metabolic processes, leading to changes in body fat distribution, insulin sensitivity, and lipid metabolism. These physiological alterations collectively contribute to an increased risk of developing T2DM and related cardiovascular complications among menopausal women compared to their premenopausal counterparts. The transition to menopause brings about substantial metabolic and hormonal shifts, including estrogen deficiency that reduces insulin sensitivity, redistribution of body fat toward visceral deposits that promote inflammation, and weight gain due to metabolic slowdown and muscle mass reduction. Furthermore, stress-induced hormonal fluctuations, sleep disturbances, and pre-existing conditions such as polycystic ovary syndrome (PCOS), gestational diabetes, family history, or premature menopause further elevate the risk (IDF, 2024; Cleveland Clinic, 2024). These factors not only heighten T2DM onset but also complicate

its management, contributing to dyslipidemia, hypertension, atherosclerosis, and genitourinary complications caused by hyperglycemia.

Background of the Study

Diabetes remains an escalating public health concern in the Philippines, with significant clinical and socioeconomic consequences. According to the International Diabetes Federation (2024), an estimated 4.2–4.7 million Filipino adults aged 20–79 is currently living with diabetes predominantly Type 2 Diabetes Mellitus (T2DM). This corresponds to a national prevalence of approximately 7.5%, a notable rise from the 4.1% reported in earlier national surveys (Azurin et al., 1986). Alarmingly, more than half of Filipino adults with diabetes remain undiagnosed, exposing millions to long-term complications such as cardiovascular disease, nephropathy, neuropathy, and retinopathy (IDF, 2024; Philippine Statistics Authority). Furthermore, with projections suggesting that cases may reach 7.5 million in the coming years, the urgency for accessible early-detection tools and preventive strategies continues to intensify.

Globally, T2DM poses a similar threat. The World Health Organization (2023) reports that the number of adults diagnosed with diabetes has nearly quadrupled in the past three decades. As a chronic metabolic disorder characterized by persistent hyperglycemia due to insulin resistance and inadequate insulin secretion, T2DM significantly reduces quality of life and places a heavy burden on healthcare systems worldwide (American Diabetes Association, 2022). Early detection remains challenging because many individuals remain asymptomatic for years, allowing the disease to progress silently before receiving appropriate care (Chen et al., 2023).

Among all demographic groups, menopausal women represent a particularly high-risk population. Menopause, defined as twelve consecutive months without menstruation (Cleveland

Clinic, 2024), accompanies hormonal fluctuations, most notably the decline in estrogen, which contributes to insulin resistance. While menopause itself does not directly cause diabetes, it significantly increases susceptibility to its development (Spencer, 2024). These hormonal changes amplify existing risk factors such as family history, dyslipidemia, and physical inactivity, increasing overall vulnerability to T2DM (NIH; Cleveland Clinic, 2023; Diabetes UK, 2023).

Type 2 Diabetes Mellitus (T2DM) continues to pose a significant public health challenge globally, with its prevalence steadily increasing across diverse populations. According to the World Health Organization (2023), the number of adults living with diabetes has nearly quadrupled over the past three decades. This chronic metabolic condition manifests with persistently high blood glucose levels due to insulin resistance and inadequate insulin secretion (International Diabetes Federation, 2023). The consequences of T2DM are profound, leading to complications such as cardiovascular diseases, kidney failure, neuropathy, and retinopathy conditions that drastically reduce quality of life and place heavy financial strain on healthcare systems worldwide (American Diabetes Association, 2022).

One of the major difficulties in managing T2DM is its silent progression. Many individuals remain undiagnosed for years because early symptoms are often subtle or absent, allowing the disease to advance before effective intervention occurs (Chen et al., 2023). This delayed detection underscores the need for innovative and precise predictive strategies, particularly among high-risk populations.

Despite this evidence, existing screening protocols rarely incorporate menopause-specific physiological changes when assessing diabetes risk. This leads to potential underestimation of

T2DM susceptibility in menopausal women and highlights a critical diagnostic gap that must be addressed.

In response, machine learning (ML) has emerged as a promising solution for precision risk detection. ML algorithms can analyze complex, multi-dimensional health data, capture subtle biomarker patterns, and improve predictive accuracy compared to traditional statistical methods (Alvarez et al., 2023). Recent studies demonstrate the effectiveness of ML models in distinguishing between normoglycemic, prediabetic, and high-risk individuals, reinforcing their potential for early detection (Patil et al., 2023). When applied to menopausal women, ML-driven predictive tools can incorporate hormonal, metabolic, and demographic factors to generate personalized risk profiles and support preventive care.

These clusters differ in progression rates, biomarker patterns, and risks for complications. Traditional screening methods rarely distinguish among these patterns, which leads to underdiagnosis or misclassification—particularly in populations undergoing physiological transitions. This complexity strengthens the need for advanced predictive technologies capable of recognizing subtle, cluster-specific biomarkers.

Moreover, beyond clinical settings, Filipino adults increasingly turn to social media particularly Facebook groups to seek health information, connect with peers, and share personal health experiences. In fact, Isip-Tan et al. (2020) demonstrate that Filipinos actively use Facebook communities to discuss health concerns and exchange advice, especially for chronic conditions such as diabetes. Likewise, Facebook has been described as a communication lifeline in the Philippines, enabling users to form supportive online communities centered on shared experiences (Congjuico, 2018). In addition, analyses of Filipino Facebook groups reveal that these spaces

function as modern social communities where members provide emotional support, shared knowledge, and collective understanding (Lapitan & Doromal, 2015).

Given this digital behavior, Facebook interest groups for menopausal women serve as accessible spaces where members openly discuss symptoms, share experiences, and seek guidance regarding menopause-related health concerns. These interactions offer meaningful insights into the lived experiences, risk factors, and health behaviors of menopausal women, making such groups a contextually relevant environment for gathering user-centered data.

Furthermore, a selected Facebook interest group named *Usapang Perimenopause at Menopause* and other interest groups were used as part of the study's research locale. This Facebook group, founded on April 20, 2023, serves as a space where members share experiences and knowledge to support one another in understanding the difficulties and challenges associated with the menopausal stage. Additionally, the group functions as a support community focused on menopause-related discussions, making it contextually suitable for gathering data on blood biomarkers, lifestyle patterns, and health awareness (Usapang Perimenopause at Menopause, 2023). By understanding the needs and health concerns expressed in these online communities, the study ensures that DIANA, a predictive model-based application, is grounded in real user context and responds directly to the population it aims to serve.

Given these challenges and opportunities, this study proposes DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Cluster-Based Identification of Type 2 Diabetes Risk in Menopausal Women. By incorporating information gain, clustering, and interactive visualization, DIANA model supports early identification of T2DM risk among

menopausal women and provides clinicians with an accessible, interpretable, and user-centered digital tool for preventive healthcare.

Nevertheless, several gaps persist in current research and predictive tools. Predictive tools for assessing diabetes risk among menopausal women remain underexplored, and existing models show notable limitations in both visualization and implementation. This raises the general research question: How can DIANA, a predictive model-based application utilizing selected blood biomarkers, effectively identify cluster-based risk levels of Type 2 Diabetes among menopausal women through biomarker analysis, clustering techniques, and user-friendly visualizations? Notably, there are three critical gaps:

Limited identification of key biomarkers for predicting Type 2 Diabetes risk in menopausal women, as current models often overlook critical indicators such as Fasting Blood Sugar (FBS), Hemoglobin A1c (HbA1c), and lipid profiles. Given the need for improved tools in predicting Type 2 Diabetes risk, there is an evident limited identification of key biomarkers that can significantly contribute to risk assessment in menopausal women. Current predictive models often fail to utilize critical indicators such as Fasting Blood Sugar (FBS), Hemoglobin A1c (HbA1c), and lipid profiles, which are essential for early detection and intervention. Several studies emphasize the importance of these biomarkers in menopausal populations. Cybulska et al. (2025) highlight that adiponectin, which correlates with FBS, HbA1c, and triglycerides, is underutilized in existing risk prediction models for postmenopausal women. Similarly, Sharma et. al., (2020) report that insulin resistance and associated glycemic and lipid abnormalities increase after menopause, underlining the need for their inclusion in predictive modeling. A study of Tamakoshi at. al., (2006), further demonstrate that menopause is an independent risk factor for elevated fasting plasma glucose, while Liu et. al., (2023), show strong associations between insulin

resistance and key biomarkers during the menopausal transition. A study of Malti and Gopalakrishnan (2007) confirm correlations between serum adiponectin, blood lipids, and HbA1c in type 2 diabetic postmenopausal women. Integrating these biomarkers into a predictive model, such as DIANA, could fill current gaps, providing more accurate risk assessment and personalized preventive strategies for menopausal women.

Lack of predictive modeling approaches that can classify menopausal women into meaningful clusters and accurately predict both their cluster membership and likelihood of developing Type 2 Diabetes. The current tools for monitoring Type 2 Diabetes risk among menopausal women remain insufficient, particularly in terms of risk profiling and visualization. Existing predictive models rarely classify individuals by risk categories or apply clustering techniques capable of revealing hidden subgroups within the population. Instead, many systems rely on static tables or basic charts, limiting the ability of healthcare providers and patients to interpret patterns, assess risk, and make informed decisions. Futhermore, several studies show that clustering combined with interactive visualization significantly enhances the understanding of heterogeneous health profiles. Kavakiotis et al. (2017) demonstrated that machine learning models employing clustering can uncover subgroups with different diabetes risk levels, thus improving early detection strategies. Similarly, Weng et al. (2017) emphasized that visual tools such as heatmaps and interactive plots help clinicians interpret complex predictive outputs more effectively, supporting personalized and targeted interventions. Given these limitations, the integration of clustering and interactive visualization into the DIANA predictive model-based application addresses the gap by enabling more accurate risk stratification, identifying high-risk subgroups, and supporting timely, evidence-based preventive care.

Absence of accessible and interactive web-based tools that present predictive outcomes and clustering results through clear visualizations such as bar charts, heatmaps, and cluster plots. Many existing predictive tools lack accessible web-based platforms capable of presenting risk predictions and clustering results through user-friendly, interactive visuals. Instead, most systems focus primarily on predictive accuracy while overlooking usability, user engagement, and real-world applicability. As a result, even technically accurate models may fail to be utilized in actual clinical or community settings due to poor user experience and limited interpretability. The need for usability evaluation in health applications is well-documented. A study of Maramba, et. al., (2019) reported that digital health tools without structured usability testing often struggle to achieve adoption, regardless of their technical quality. Likewise, Tubaishat et al. (2021) emphasized that incorporating Likert-scale surveys and user feedback mechanisms provides essential insights into usability, satisfaction, and engagement, leading to more effective and user-centered system design. By evaluating the DIANA model through a Likert-scale survey, the developed web application is not only scientifically robust but also practical, accessible, and user-friendly for both healthcare providers and menopausal women at risk of Type 2 Diabetes.

Objective of the Study

The general objective of this study is to develop DIANA, a predictive model-based application that utilizes selected blood biomarkers to identify cluster-based risk levels of Type 2 Diabetes among menopausal women. This study aims to develop a system that analyzes biomarker patterns, predicts individual risk through clustering techniques, and presents results through user-friendly visualizations. By doing so, DIANA supports menopausal women in understanding their

health status, assists clinicians in interpreting key risk-contributing factors, and bridges the gap between data-driven prediction and practical clinical decision-making.

Specifically, this study aims to:

1. **Determine the most informative biomarker attributes by computing Information Gain (IG) and analyzing the entropy of the dataset to select features that significantly contribute to predicting Type 2 Diabetes risk among menopausal women.**

   The study will develop a predictive model utilizing a machine learning techniques trained on the identified blood biomarkers which are the Fasting Blood Sugar (FBS), Hemoglobin A1c (HBA1C), Lipid Profiles and the non-blood biomarkers which are the Age, Weight, BMI and Lifestyle associated with type 2 diabetes in menopausal women. Unlike the traditional risk assessment that relies on limited statistical correlations, machine learning can capture complex and nonlinear interactions among biomarkers through algorithms such as regression analysis, decision trees, and neural networks. The model will identify patterns that enhance prediction accuracy. This data-driven approach enables a more personalized and precise risk evaluation offering healthcare professionals a tool for early detection and targeted prevention compared to conventional methods

2. **Cluster users based on biomarker, demographic, and lifestyle data, and develop a predictive modeling approach that determines their cluster membership while estimating their likelihood of Type 2 Diabetes risk using machine learning techniques applied to data gathered from interest groups.**

   The DIANA predictive model-based application will function as an advanced platform designed to develop a predictive modeling approach that classifies users into distinct clusters and predicts both their cluster membership and likelihood of Type 2

Diabetes risk. This modeling framework will employ machine learning techniques that integrate biomarker data, demographic profiles, and lifestyle factors gathered from targeted interest groups. By processing these multidimensional inputs, the system will generate precise and individualized risk classifications specifically tailored for menopausal women.

This will generate structured, personalized data summaries that consolidate biomarker readings, demographic characteristics, and relevant lifestyle information into a coherent clinical profile. By integrating predictive modeling results with these detailed user profiles, the system converts raw clinical data into actionable, evidence-based insights. This enables healthcare practitioners to recognize deviations, risk elevations, or emerging metabolic concerns more effectively

Through this platform, healthcare professionals including physicians, endocrinologists, and clinical researchers will have access to comprehensive predictive outputs that reflect the user's overall metabolic status. These outputs will be derived from key biomarkers such as Hemoglobin A1c (HbA1c), Fasting Blood Sugar (FBS), and lipid profiles, allowing clinicians to examine how each physiological indicator contributes to the user's assigned cluster and estimated diabetes risk level. The clustering mechanism will further help in identifying patterns or subgroups within the menopausal population, supporting more targeted clinical interpretation.

3. **Visualization of risk predictions and clustering outputs will be enabled through a web-based application that integrates the developed predictive model and presents results using bar charts, heatmaps, and cluster plots.**

This serve as an interactive platform designed to develop a predictive modeling approach that classifies users into clusters and predicts both their cluster membership and likelihood of Type 2 Diabetes risk, using machine learning techniques applied to biomarker, demographic, and lifestyle data gathered from interest groups. Through this system, healthcare professionals such as physicians and endocrinologists will be able to visualize individualized risk levels among menopausal women based on selected blood biomarkers.

Designed with a user-centered approach, the platform aims to strengthen the capacity of medical practitioners to detect at-risk individuals at earlier stages. In doing so, the DIANA system supports improved preventive care, encourages timely intervention strategies, and promotes proactive health management specifically tailored for the needs of menopausal women.

Additionally, the web application will offer personalized data summaries, enabling healthcare professionals to monitor trends in key biomarkers such as Hemoglobin A1c (HbA1c), Fasting Blood Sugar (FBS), and lipid profiles. By integrating clustering results with predictive indicators, the DIANA system bridges the gap between raw diagnostic data and actionable clinical insights. Its user-friendly interface will further support practitioners in identifying high-risk individuals earlier, promoting preventive care and proactive health management for menopausal women.

Rationale of the Study

In recent years, the increasing prevalence of Type 2 Diabetes Mellitus (T2DM) has posed a growing challenge to public health, particularly among menopausal women who experience hormonal and metabolic changes that heighten their risk of developing the disease. Despite

advancements in diagnostics, current screening methods such as Fasting Blood Sugar (FBS) and Hemoglobin A1c (HbA1c), primarily serve as a reactive function identifying diabetes only after its onset (IDF, 2024). As a result, there is a pressing need for predictive tools that enable early detection and prevention of T2DM before severe complications arise.

Recent studies have shown the potential of machine learning-based predictive modeling in healthcare, allowing for the identification of hidden patterns within biomedical data that traditional diagnostic methods often overlook (Kopitar et al., 2020; Mohd Rizal et al., 2024). However, many existing models lack population-specific considerations and effective visualization tools that can help healthcare professionals interpret risk levels more efficiently.

The proposed study addresses this gap through the development of DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Cluster-Based Identification of Type 2 Diabetes Risk in Menopausal Women. DIANA predictive model-based application aims to integrate traditional diagnostic markers with novel blood biomarkers to provide a more comprehensive and individualized risk assessment. Additionally, the inclusion of an interactive dashboard and visualization feature allows healthcare professionals to easily monitor, interpret, and act upon predictive results. The groups what will benefit from this study are:

Healthcare providers: including medical professionals will benefit from this study. With DIANA, they can review and validate the predictive results generated by the web application, allowing them to better assess patient risk, interpret biomarker patterns, and support informed clinical decision-making.

Hospitals and clinics: will also benefit as the application enhances patient care and monitoring systems. By integrating predictive analytics into their operations, healthcare facilities

can reduce diagnostic delays, improve preventive care strategies, and strengthen patient record management, particularly in populations that are traditionally underserved in risk assessments.

Menopausal women: are the primary beneficiaries of this study. Through personalized risk assessments and improved record management, they can gain better awareness of their health status. This allows for timely lifestyle modifications, preventive measures, and early medical interventions, ultimately lowering the likelihood of developing Type 2 Diabetes and its complications.

Public Health Sector: This will gain from this study's contribution to targeted disease prevention efforts. By addressing a vulnerable group, the study supports strategies aimed at reducing the prevalence and burden of diabetes at both community and national levels. In the context of the Philippines, where diabetes remains a growing concern, this study offers an innovative tool that aligns with public health initiatives to strengthen early detection and prevention strategies.

Moreover, this study leverages the use of predictive modeling techniques and data visualization to create a robust tool for identifying early risks of Type 2 Diabetes in menopausal women including T2DM Subgroups clustering. By doing so, DIANA serves as a valuable decision-support system that enhances clinical judgment, promotes preventive care, and assists in timely intervention. The application also enables medical practitioners to translate complex biomarker data into understandable insights that support early diagnosis and lifestyle-based risk management.

In alignment with the United Nations Sustainable Development Goal (SDG) 3: Good Health and Well-Being, this study contributes to the promotion of healthier lives and the prevention of noncommunicable diseases such as diabetes. By developing DIANA, the researchers

aim to support healthcare professionals and empower menopausal women through technology-driven early detection, improved monitoring, and informed clinical decision-making ultimately fostering a healthier and more proactive healthcare system in the Philippines.

Scope and Delimitations of the Study

This study focuses on the development of DIANA, a predictive model-based application that utilizes selected *blood biomarkers* to classify current diabetes risk in menopausal women (T2D).

This study has the following scope:

- *Identification and Selection of Blood Biomarkers***:** The identification and selection of blood biomarkers will focus on determining those significantly associated with current Type 2 Diabetes status among menopausal women. This process will involve evaluating clinically accessible biomarkers such as Fasting Blood Sugar (FBS) and Hemoglobin A1c (HbA1c), which are known to play vital roles in glucose metabolism and hormonal regulation during menopause. In addition, the study will consider biomarkers that have shown relevance in distinguishing between the established clustering of T2DM subgroups such as Severe Insulin-Deficient Diabetes (SIDD), Severe Insulin-Resistant Diabetes (SIRD), Mild Obesity-Related Diabetes (MOD), and Mild Age-Related Diabetes (MARD) as identified in recent diabetes stratification research (Prasad et al., 2018;Veelen et al., 2021; Yang et al., 2025). By acknowledging these subgroup clusters, the selection process ensures that the biomarkers reflect not only general diabetes indicators but also the heterogeneity of T2DM presentations.

- *Development of Predictive Model Using Machine Learning:* Using the selected biomarkers, the researchers will develop a predictive classification model employing machine learning algorithms to classify the current diabetes risk status of menopausal women. The study will explore models such as logistic regression, random forest, or support vector machines to determine which approach yields the most accurate predictive performance. Model training and validation will be performed using available biomarker datasets. Statistical evaluation methods such as accuracy, sensitivity, specificity, and AUC (Area Under the Curve) will be applied to measure the model's performance and reliability.
- *Integration into a Web Application***:** The study will integrate the developed predictive model into a web-based application named DIANA, which will be designed exclusively for use by healthcare professionals, particularly doctors specializing in endocrinology or internal medicine. The web application will serve as a decision-support tool, allowing physicians to input relevant biomarker data from menopausal patients and receive a predictive assessment of Type 2 Diabetes risk. The system will focus on assisting in clinical evaluation and preventive care planning rather than providing automated medical diagnoses or treatment recommendations. To enhance usability, the interface will include visual representations such as graphs, charts, and color-coded indicators to support clear interpretation and facilitate evidence-based decision-making in clinical settings.
- *Evaluation of Model Accuracy and Application Usability:* The predictive model's performance will be assessed using computational evaluation metrics such as accuracy, sensitivity, specificity, and Area Under the Curve (AUC) to determine its

predictive reliability. Meanwhile, the usability of the DIANA web application will be evaluated through a structured usability test involving selected healthcare professionals. The assessment will focus on key factors such as system functionality, ease of navigation, clarity of risk presentation, and reliability of results. Feedback gathered from participating medical professionals will serve as the basis for refining the application's interface and ensuring that it effectively supports clinical decision-making in the early identification of Type 2 Diabetes risk among menopausal women.

- *Web Application Functionalities:* The DIANA predictive model will be incorporated into a dedicated web application designed solely for healthcare professionals particularly Doctors. This system will operate within secure medical networks and be accessible only to authorized users. By entering patient biomarker data in a prescribed format, doctors can obtain real-time diabetes risk assessments for menopausal women. The web application will feature multiple sections that provide key functions, including data input, result visualization, and risk analysis, ensuring an efficient and user-friendly experience for clinical use.
    - *Dashboard Tab*– serves as the central overview interface of the DIANA application. It provides a real-time summary of all patient-related data stored in the system. At the top of the dashboard, users can view the total number of registered patients, recent additions, and summary statistics that reflect the system's current data load. Below these key metrics will be showing the graphical representations of the collective blood biomarker levels gathered from all patient entries. These graphs visualize the overall

trends in biomarkers such as fasting plasma glucose, HbA1c, and Estradiol, among others. This allows the user to observe trends across their patient population and detect potential increases in diabetes risk prevalence among menopausal women. This feature supports data-driven monitoring and decision-making.

- *Patient history Tab*– acts as archive of the stored data and organizes all patient records systematically. Each record entry includes essential details such as the patient's name, age, and the date of the latest added assessment. When user clicks or hovers over a patient record, the interface opens a detailed profile view displaying the patient's full information. This includes the complete name, historical biomarker readings, and a line graph that overlays previous biomarker results with the most recent assessment. This visual overlay provides an immediate comparison between the patient's historical biomarker readings and current assessment (e.g., changes in FPG or HbA1c values), enabling the clinician to contextualize current diabetes risk within the patient's biomarker trend. Furthermore, this tab presents the risk assessment result generated by DIANA model showing the current probability score (0–100%) and risk category (Low/Moderate/High) indicating likelihood of current undiagnosed Type 2 Diabetes or prediabetes. This enables clinicians to prioritize diagnostic confirmation and early intervention.
- *Analytics Tab* – provides an interactive visualization interface that allows the users to interpret predictive insights derived from the patient data

processed by the type 2 diabetes risk prediction model. It displays two primary components: the risk factor importance chart, which ranks input variables such as Age, BMI, Blood Pressure, Glucose Level, and Physical Activity according to their computed contribution to the model's predictive output and the BMI vs Glucose Correlation chart, which plots the relationship between body mass index and glucose level among patients to identify trends or potential risk associations. Data for this visualization are retrieved from the system database, processed by the backend analytics engine, and dynamically rendered on the frontend using visualization libraries. This module enables clinicians and researchers to easily assess which factors are most influential on developing diabetes risk and to explore correlations among physiological indicators, supporting data driven medical interpretations and decision making within the DIANA system.

- *Export Tab*– enables users to download datasets and analytical reports generated within the system for documentation, research, or further offline analysis. It provides three main export functionalities: export participant data, export analytics report, and filtered export. The export participant data section allows users to download the complete dataset of participant records, including demographic details, biomarker values, and prediction results, in either CSV or Excel format for compatibility with the data analysis tools. The export analytics report feature generates a comprehensive summary of the model's analytical outputs, including factor importance, correlation analysis, and predictive insights, which can be downloaded as formatted

report file. Meanwhile the filtered export option allows users to selectively export data according to specific parameters such as menopausal stage and diabetes risk level, enabling focused examination of subsets of the dataset. All export processes are handled by the backend, where the system compiles and formats the requested data, then generates a downloadable file. This module ensures efficient data management, facilitates result sharing and supports further statistical evaluation or validation of the predictive model outside the DIANA platform.

The delimitations of the study are:

- *Limitation to Blood Biomarkers:* The study will exclusively utilize blood-based biomarkers as indicators for diabetes risk. Other potential diagnostic sources such as imaging data, genomic markers, or microbiome profiles will not be included. This delimitation is set to maintain the practicality and accessibility of the study, as blood biomarkers are commonly used in clinical settings, cost-effective, and easily obtainable through routine laboratory testing.
- *Absence of Treatment or Management Recommendations:* The predictive model developed in this study is not intended to function as a diagnostic or medical decision-making tool. Instead, it serves as a decision-support system to assist healthcare providers and menopausal women in recognizing potential diabetes risk. The DIANA application will only provide probabilistic or risk-based outputs derived from biomarker data, and all final diagnoses should still be conducted by licensed medical professionals.

- *Dataset Size and Demographic Limitation:* The dataset used to train and validate the predictive model may be limited in sample size, geographical coverage, and participant diversity. This may influence the model's ability to generalize predictions to broader populations. Consequently, while the model aims to achieve high accuracy within the study parameters, its predictive performance may vary when applied to different demographic groups or clinical populations.
- The dataset used may be limited in size and scope, potentially affecting generalizability to broader populations.
- *Temporal Scope & Prediction Clarification:* DIANA classifies current diabetes/prediabetes risk at a single timepoint using machine learning. The term "predictive" refers to classifying current disease state, not forecasting future onset. No longitudinal follow-up data are employed; therefore, DIANA is a risk classification tool for current-state screening, not a prospective forecasting system.

Conceptual Framework

The conceptual framework, shown in Figure 1, illustrates the overall process of how the proposed predictive classification model-based application identifies menopausal women with undiagnosed Type 2 Diabetes or prediabetes based on their biomarker profiles. It begins with the recognition of the key inputs necessary for analysis, including patient demographics such as age, menopausal status, and lifestyle factors. It also incorporates selected blood biomarkers which include HbA1c, Fasting Plasma Glucose, and Lipid Profiles. This serves as biochemical indicators for assessing the metabolic and inflammatory states associated with diabetes risk. Additionally,

patient history data such as previous diagnoses and family history of diabetes provide contextual information to improve the classification accuracy of the system.

The process stage comprises data collection and preprocessing, followed by application of machine learning algorithms such as logistic regression, random forest, or neural networks. These analyze the relationships between biomarkers, background risk factors, and current diabetes status. The model produces a risk probability score (0–100%) indicating the current likelihood that a patient has undiagnosed Type 2 Diabetes or prediabetes based on her present biomarkers. A score of 72% means: given her current biomarker profile, there is a 72% probability she currently has or is at high risk of prediabetes/diabetes, not that she will develop it in future years. Risk is stratified into Low (0–33%), Moderate (34–66%), and High (67–100%) for clinical decision-making.

The proposed model-application system converts biomedical input into actionable risk probability estimates, supporting evidence-based clinical decision-making. By stratifying current diabetes risk across clinically meaningful categories (Low/Moderate/High), the application helps prioritize diagnostic confirmation, lifestyle intervention, and clinical surveillance based on current biomarker patterns.

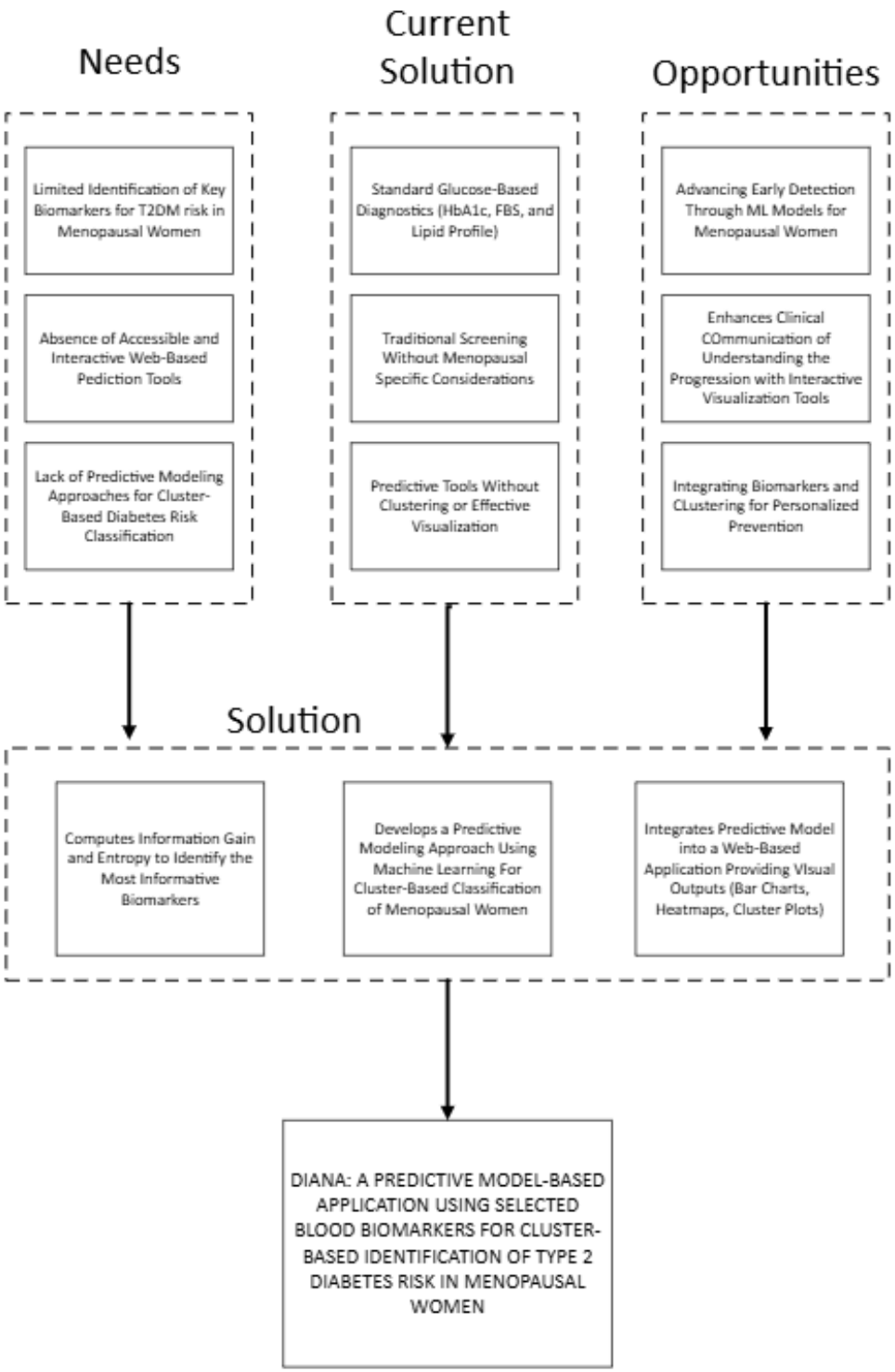


Figure 1: Conceptual Framework

Review of Literature

This chapter explores the literature and studies that form the theoretical basis of the research. Also, it presents distinctive foreign and local kinds of literature and studies compiled from online journal resources used by this study. The study utilized this chapter as its foundation and guide toward proper research. The following pieces of literature gave the study a better understanding and a broader perspective on the topic proposed.

Diabetes: Type 1 and Type 2

Diabetes Mellitus (DM) is a chronic metabolic disorder characterized by elevated blood glucose levels resulting from insufficient insulin production or ineffective insulin utilization. Insulin, a hormone produced by the pancreas, regulates glucose metabolism and maintains the body's energy balance. Globally, the prevalence of diabetes has risen sharply, with the World Health Organization (2024) reporting that approximately 14% of adults aged 18 and older were diagnosed in 2022, representing a 7% increase since 1990. The International Diabetes Federation (IDF) projects that the number of adults aged 20 to 79 with diabetes may surpass 200 million by 2040, emphasizing its growing public health burden.

Type 1 Diabetes

Type 1 diabetes develops due to autoimmune destruction of pancreatic beta cells, resulting in insufficient insulin production. Affected individuals rely on external insulin to maintain glucose homeostasis. Although typically diagnosed in childhood or adolescence, prevalence among adults is rising, highlighting the need for lifelong management (Ezzatvar et al., 2023; Mobasseri et al., 2020).

Type 2 Diabetes

On the other hand, Type 2 diabetes primarily results from insulin resistance, in which the body produces insulin, but cells fail to respond effectively, gradually causing hyperglycemia. This condition is strongly linked to lifestyle factors such as high-calorie diets, obesity, and physical inactivity. Hormonal changes during menopause further exacerbate insulin resistance, placing women at heightened risk. Studies show that early menopause increases the odds of developing Type 2 Diabetes by 24%, while menopause before age 45 significantly elevates risk (HR = 1.31; 95% CI: 1.05–1.64) (Muka et al., 2017; Yazdkhasti et al., 2024). These findings underscore the importance of early preventive strategies for menopausal women.

Pre-Diabetic Stage

In addition to these types, the pre-diabetic stage represents an intermediate condition between normal glucose tolerance and diabetes, marked by mildly elevated blood glucose levels. This stage serves as a critical window for intervention, as lifestyle modification can often prevent progression to full diabetes. Many individuals remain undiagnosed due to the absence of symptoms. Among menopausal women, fluctuating hormone levels accelerate insulin resistance, increasing progression risk. Biomarker-based predictive models, such as FBS, HbA1c, lipid profiles, BMI, and lifestyle factors, can help identify women at this stage and support timely intervention (Cybulska et al., 2023; Xiao et al., 2024).

Characterization of Type 2 Diabetes Subgroups

Recent research has shifted the conceptualization of adult-onset diabetes away from a one-size-fits-all "Type 2" label toward a more nuanced, data-driven taxonomy. Work in this area combines clinical phenotyping, cluster analysis, molecular profiling, and evidence synthesis to define discrete subgroups that differ in pathophysiology, risk of complications, and likely therapeutic response. The papers you supplied form a tight, complementary body of literature: Ahlqvist et al. (2018) introduced the cluster approach and demonstrated its clinical relevance; follow-on molecular work (Schrader et al., 2022) linked those clusters to distinct epigenetic signatures and complication risk; translational reviews (Veelen et al., 2021) explored how subgroups could guide medication strategies; and a recent systematic review and meta-analysis (Ao et al., 2025) aggregated clinical and laboratory differences across studies. Together they provide a framework for stratified risk prediction and personalized management of diabetes.

Foundational data-driven taxonomy (Ahlqvist et al., 2018).

Ahlqvist and colleagues applied unsupervised clustering to routinely available clinical variables to identify reproducible subgroups of adult-onset diabetes. Using six variables commonly measured at diagnosis, their data-driven approach partitioned patients into distinct clusters (commonly reported as severe autoimmune diabetes, severe insulin-deficient diabetes, severe insulin-resistant diabetes, mild obesity-related diabetes, and mild age-related diabetes). Crucially, these clusters showed different trajectories for glycemic control, required treatments, and risks for micro- and macrovascular complications. The study's key contribution was empirical: showing that a simple algorithmic reclassification of

heterogenous diabetes presentations yields groups with prognostic and therapeutic relevance. Methodologically, it demonstrated the power of clustering on clinical phenotypes and emphasized externally validating cluster solutions across cohorts.

Molecular and epigenetic stratification

In a study of Schrader et. al., (2022), they show clustering concept by interrogating epigenetic profiles across the novel subgroups. Their analysis revealed subgroup-specific DNA methylation patterns and epigenetic signatures that correlated with later development of diabetic complications. This result supports the idea that clinical clusters reflect underlying, biologically meaningful differences rather than purely phenotypic or demographic variation. By linking epigenetic marks to subgroup identity and outcome, the study provided molecular validation for the cluster model and suggested possible mechanisms (and biomarkers) for subgroup-specific complication risks.

Therapeutic implications and personalized strategies

In a study of Veelen et al., (2021), reviewed how subgroup classification might inform medication strategies, mapping subgroup pathophysiology (insulin deficiency vs. insulin resistance, obesity-related mechanisms, age-related processes) to drug classes and therapeutic goals. The review argued for tailoring therapy to the dominant mechanistic defect for example, prioritizing insulin-secretagogue or insulin replacement strategies in insulin-deficient subgroups and insulin-sensitizing or incretin-based approaches in insulin-resistant groups while cautioning that randomized evidence for subgroup-guided therapy is still limited. The authors emphasized that subgrouping could accelerate precision medicine in

diabetes but also stressed the need for prospective trials and consideration of comorbidities, age, and patient preferences.

Synthesis across cohorts — systematic review and meta-analysis

In a study of Ao et al., (2025), they synthesized clinical and laboratory characteristics of the novel diabetes subgroups across available studies, quantifying how subgroups differ in biomarkers, demographics, and complication rates. Their meta-analytic aggregation strengthened the robustness of several subgroup-defining features (e.g., measures of insulin resistance and beta-cell function) and identified heterogeneity across cohorts that related to study design, population ancestry, and variable definitions. The review highlighted consistent patterns but also underscored variability that complicates direct transfer of cluster algorithms between settings.

Meanwhile, in the Philippines, Diabetes represents a growing public health concern. Cando et al. (2024) estimate that in 2021, 4.3 million Filipinos were diagnosed with diabetes, while 2.8 million remained undiagnosed due to limited access to healthcare and screening. Contributing factors include urban lifestyle changes, high-carbohydrate diets, sedentary routines, and limited awareness of menopausal health, particularly among women (Araneta et al., 2020; Tan, 2015; Fuller-Thomson, 2017).

Overall, despite global and local research on Type 1 and Type 2 Diabetes, most studies focus on the general population, leaving high-risk groups such as menopausal women underrepresented. This highlights the need for predictive tools, like DIANA, that target specific risk factors in this population.

Diabetes Risk Factors

Several studies have identified key factors that significantly influence the likelihood of developing Type 2 Diabetes (T2DM). A study of Bi et al. (2012) shows that the rise of T2DM results not only from obesity and family history but also from lifestyle and environmental contributors such as physical inactivity, unhealthy diet, smoking, and alcohol consumption. These behaviors promote insulin resistance and metabolic imbalance, which increases diabetes risk.

Furthermore, according to Wu et al. (2014) emphasize that engaging in regular physical activity and maintaining a healthy diet substantially reduces the risk of T2DM. Individuals who perform moderate exercise for at least 150 minutes per week show a lower risk compared to those who are inactive. Diets rich in whole grains, vegetables, and legumes offer protective benefits, while excessive intake of processed foods and sugary beverages increases risk.

In addition, among menopausal women, these risk factors become more critical due to hormonal changes that alter glucose metabolism and fat distribution. The decline in estrogen levels affects insulin sensitivity and contributes to central obesity, further increasing the likelihood of developing T2DM. Identifying these risks through selected blood biomarkers provides a measurable approach to predicting diabetes onset during menopause, supporting early intervention and prevention strategies.

Menopausal

Menopause is a significant physiological transition characterized by reduced ovarian function and declining estrogen levels, which directly influence glucose metabolism, lipid regulation, and insulin sensitivity. As estrogen levels decrease, women become more vulnerable

to metabolic disorders such as Type 2 Diabetes Mellitus (T2DM) (Carr, 2020). Moreover, hormonal changes during this stage often lead to increased central adiposity and altered fat distribution, both of which contribute to insulin resistance.

In addition, hormonal fluctuations also impair pancreatic beta-cell function, reducing insulin secretion and further elevating diabetes risk. Women who experience early or premature menopause are particularly susceptible, as studies reveal that early menopause increases the likelihood of developing T2DM by approximately 24%, while menopause before age 45 further amplifies the risk (Muka et al., 2017; Yazdkhasti et al., 2024).

Furthermore, advancements in predictive modeling and biomarker-based approaches provide valuable tools for identifying menopausal women at risk of diabetes. Recent studies demonstrate that combining biomarkers such as fasting blood sugar (FBS), HbA1c, and lipid profiles with machine learning algorithms yields high predictive accuracy in assessing diabetes risk (Chatterjee et al., 2023). Similarly, local research explores prototype systems that integrate biomarker analysis for diabetes prediction in clinical settings (Campugan et al., 2025).

Overall, menopause shows a crucial role in the development of Type 2 Diabetes. Thus, integrating key biomarkers including FBS, HbA1c, lipid profiles, BMI, age, menopausal status, and lifestyle factors into predictive models supports early detection and timely intervention. This approach aligns with the goal of the present study, which focuses on developing DIANA, a predictive model-based application designed to identify menopausal women at risk of Type 2 Diabetes.

Biomarkers

Biomarkers are measurable biological indicators that provide critical insights into an individual's physiological and metabolic state. In diabetes research, they play a vital role in identifying metabolic disturbances, monitoring disease progression, and predicting future risk (World Health Organization, 2023). This predictive potential is especially important for menopausal women, who experience hormonal and metabolic changes that increase susceptibility to Type 2 Diabetes Mellitus (T2DM) (Cybulska et al., 2023; Wang et al., 2018). Integrating biomarker assessments with computational models can enhance early detection and support personalized preventive strategies.

T2DM is commonly diagnosed using standard glucose-based tests such as Fasting Blood Sugar (FBS), Hemoglobin A1c (HbA1c), and Oral Glucose Tolerance Test (OGTT) (The Medical City, 2023; NYU Langone Health, 2023). Local experts also recommend including lipid profiles, renal and liver function tests (Chem 12/15), along with other assessments such as thyroid hormone and urinalysis, to evaluate related metabolic risks (personal communication, 2025; Villa, 2017). Table 1 shows the commonly used range values for FBS, OGTT, and HbA1c for detecting diabetes and prediabetes (Gutierrez, 2020). An example HbA1c test result is provided in Appendix C.

Test	Normal	Prediabetes	Diabetes
FBS (mg.dL)	<100	100 – 125	>126
75g OGTT	<140	140 – 199	>200
HBA1C (%)	<5.7	5.7 – 6.4	>6.5

Table 1: Range Values for FBS, OGTT and HBA1C

Furthermore, according to the interviewed specialist (personal communication, 2025), the range values for Fasting Blood Sugar (FBS) and Random Blood Sugar (RBS) are presented in the table below, with the full interview transcript provided in Appendix E.

Fasting Blood Sugar (FBS)	> 126 mg/dL
Random Blood Sugar (RBS)	> 210 mg/dL

Table 2: Range Values for FBS and RBS

In addition to glucose measures, lipid profile parameters including triglycerides, LDL, HDL, and apolipoprotein B play a significant role in assessing insulin resistance and diabetes risk, particularly in menopausal women (Giandalia et al., 2021). A recent study by Jasim et al. (2025) highlighted correlations between triglyceride/HDL ratios and the triglyceride-FBS index as early indicators of impaired glucose metabolism. The detailed correlation table from their study is included in Appendix D.

Moreover, the study also considers non-biochemical factors that influence diabetes risk, such as age, menopausal status, body mass index (BMI), and lifestyle behaviors, which are critical for developing accurate predictive models tailored to the target population.

Overall, by focusing on these clinically accessible and population-relevant biomarkers, the current study aims to develop a predictive model that identifies menopausal women at risk of developing Type 2 Diabetes, supporting early intervention and evidence-based preventive strategies**.**

Predictive Modeling

Predictive modeling shows an essential role in identifying individuals at risk for Type 2 Diabetes Mellitus (T2DM), especially in populations with varying metabolic and hormonal

characteristics. Traditional regression-based models have long been utilized to determine relationships among clinical and demographic variables, yet they often rely on simplified assumptions that limit their capacity to capture complex health interactions. Despite this, predictive modeling remains a cornerstone of preventive medicine, providing structured risk assessment tools that aid in early diagnosis and disease management.

Furthermore, in the Philippines, A study of Campugan & Aguaras (2025) conduct a local predictive modeling study involving 947 Filipino adults aged 24–79 years to classify diabetes status using accessible clinical biomarkers. By employing binomial logistic regression and decision-tree analysis on variables such as Body Mass Index (BMI), Low-Density Lipoprotein (LDL), Hemoglobin A1c (HbA1c), and triglycerides, their study identifies BMI as the most influential predictor ($\chi^2 = 104.44$, $p < 0.001$), followed by HbA1c ($\chi^2 = 51.80$, $p < 0.001$), triglycerides, and LDL. The logistic model achieves a strong explanatory power (McFadden $R^2 = 0.80$; Nagelkerke $R^2 = 0.85$), while decision-tree analysis confirms BMI and HbA1c as critical classifiers. These results highlight the potential of predictive modeling using low-cost and clinically measurable biomarkers, particularly in resource-limited healthcare systems. The researchers recommend integrating such predictive tools into local health programs for early diabetes screening and risk management, an approach directly relevant to the design of the DIANA system.

Moreover, international research supports the value of biomarker-based predictive modeling in menopausal populations. A study of Chatterjee et al. (2023) demonstrate that integrating biomarkers such as fasting blood sugar (FBS), HbA1c, and lipid profiles with machine learning algorithms enhances predictive accuracy for identifying high-risk women undergoing menopausal transition. Similarly, a 10-year prospective cohort study involving 300,000 Chinese

women finds that earlier menopause correlates with a heightened risk of T2DM, emphasizing the metabolic impact of hormonal decline (Zhao et al., 2022). In Japan, the Toranomon Hospital Health Management Center Study reports that postmenopausal women exhibit elevated fasting glucose and impaired insulin regulation compared to premenopausal counterparts (Nishida et al., 2021).

Overall, these findings validate the effectiveness of predictive modeling using clinical biomarkers in assessing diabetes risk among menopausal women. Consequently, the integration of such data-driven models into digital health tools, such as DIANA, presents an innovative step toward early detection and intervention tailored to the needs of this high-risk population.

Data Visualization Techniques

Data visualization plays a significant role in translating complex health information into insights that both clinicians and patients can easily understand. Literature shows that visual representations significantly reduce cognitive load, making it easier for users to identify patterns, assess trends, and interpret large datasets. A study of Knaflic et al., (2021) and McNutt et al., (2022), shows systematic review of data visualization in public health wherein that clear visual tools improve comprehension, trust, and decision-making among both experts and non-experts, emphasizing that visualization influences perceptions, attitudes, and behavior When applied to clinical settings, visualization becomes even more essential because healthcare professionals routinely interpret biomarker changes, risk classifications, and longitudinal trends that are better understood when presented visually rather than through plain numerical outputs.

In addition, studies on clinical visualization techniques highlight the importance of simplicity and alignment with clinician workflows. A study of Sun et al., (2024) shows recent

scoping review found that tables, scatterplot-line timelines, event timelines, and structured text displays are among the most commonly used formats for presenting individual patient health data because they allow clinicians to quickly identify abnormalities and track physiological changes over time. These visual tools reduce cognitive burden, especially when integrated into user-centered design frameworks, which emphasize iterative feedback, prototyping, and usability testing to ensure that visualizations remain intuitive within busy clinical environments.

Furthermore, in predictive modeling, effective visualization serves as a bridge between machine learning outputs and clinical interpretation. Predictive models often generate probability scores or complex risk values that can be difficult to understand without context. To address this, visual explanation tools such as feature-importance charts, risk contribution plots, and color-coded indicator bars help translate model outputs into interpretable insights. In a study of Van Belle and Van Calster (2015), they demonstrated how patient-specific contribution charts and colorized risk bars enhance model transparency by showing how each variable influences a prediction. This increases clinician trust and supports evidence-based decision-making.

Moreover, interactive visual analytics frameworks further elevate interpretability by linking prediction explanations with real patient data. Tools like VBridge demonstrate how interactive visualization enables users to explore model predictions, examine high-impact features, and understand underlying data relationships through a hierarchical interface (Li et al., 2021). Wherein systems allow clinicians to interact with both raw data and machine learning explanations, improving their ability to evaluate risk estimates, and verify the model's reasoning. Model-agnostic interpretability tools like Petal-X also show that interactive visual explanations outperform traditional risk charts by helping users compare the influence of modifiable and non-modifiable risk factors without sacrificing trust or transparency (Desai et al., 2024).

Furthermore, visualization dashboards have also been widely used in chronic disease monitoring, including diabetes research. A study of German et al., (2024) developed an interactive Tableau dashboard integrating sociodemographic, biomarker, and geographic data to explore disparities in Type 2 Diabetes outcomes. Wherein the dashboard enabled users to filter and analyze data at individual, neighborhood, and population levels, allowing for targeted interventions and more effective monitoring. These findings suggest that visualization-driven systems significantly enhance engagement, interpretation, and decision-making across different levels of healthcare.

Overall, literature consistently shows that effective visualization techniques enhance usability, improve clarity of predictive outputs, and support more reliable decision-making. By combining intuitive charts, interactive elements, and explainable model outputs, visualization becomes a critical component of modern predictive tools.

Machine Learning

The integration of machine learning (ML) techniques in healthcare has transformed the landscape of predictive analytics, allowing for the detection of subtle, nonlinear relationships among clinical and behavioral variables. Unlike conventional regression methods, ML algorithms can handle large, multidimensional datasets and uncover hidden patterns that enhance early disease prediction. This advancement has proven particularly effective in chronic diseases like Type 2 Diabetes Mellitus (T2DM), where early identification of at-risk individuals is crucial for timely intervention and management.

Furthermore, recent comparative studies highlight the superior predictive performance of algorithms such as Random Forest, XGBoost, LightGBM, and Support Vector Machines (SVM) over traditional models, particularly as datasets grow in complexity (Kopitar et al., 2020). Among

these, LightGBM exhibits high stability in variable selection, while regression-based models continue to offer interpretability an important aspect for clinical decision-making. These findings underscore the importance of selecting algorithms that balance predictive strength with explainability in healthcare contexts.

In addition, a study by Abdulhadi and Al-Mousa (2021) applies six ML classification methods such as Random Forest, Naïve Bayes, K-Nearest Neighbor, Decision Tree, SVM, and Neural Networks on both the PIMA diabetes dataset and original questionnaire data. Results reveal that Random Forest achieves the highest accuracy (94.10%), identifying variables such as age, family history, physical activity, and gestational diabetes as the strongest predictors. This research demonstrates that integrating both clinical biomarkers and lifestyle factors yields more accurate predictions of diabetes onset.

Moreover, in menopausal populations, Xiaoxue et al. (2024) develop a risk prediction model for metabolic syndrome using machine learning techniques. Their findings demonstrate that ML-based approaches can effectively capture the complex interplay between hormonal decline and metabolic dysfunction, validating the feasibility of applying predictive analytics to women undergoing menopausal transition. This study reinforces the notion that traditional screening tools often overlook hormonal and metabolic variables unique to this demographic.

In addition, further advances in ML applications emphasize early prediction of key biomarkers such as HbA1c. Innovative frameworks now aim to predict glycemic deterioration before clinical thresholds are reached, enabling proactive management and personalized intervention strategies. Moreover, ensemble learning methods, including Gradient Boosting and Random Forest, demonstrate robust performance due to their ability to combine multiple weak

learners and provide variable importance measures that enhance interpretability (Mohd Rizal et al., 2024).

However, despite their advantages, ML-based models face several challenges that hinder clinical adoption. Vabalas et al. (2019) highlight issues related to interpretability, data imbalance, and external validation, which affect model generalizability and clinical trust. Neural networks and ensemble algorithms, while powerful, often lack transparency posing difficulties for healthcare practitioners who require clear, evidence-based reasoning in diagnosis and treatment recommendations. These limitations indicate the need for models that are not only accurate but also user-friendly, interpretable, and validated across diverse populations.

Entropy and Information Gain for Feature Selection.

In machine learning, feature selection is the process of identifying the most informative attributes from a larger set of candidate variables. This process is essential for building efficient, interpretable, and accurate predictive models. Entropy and Information Gain (IG) represent foundational information-theoretic measures that quantify the relevance and discriminative power of individual features in classification tasks (Sreehari et al., 2024). Entropy measures the degree of uncertainty or randomness in a dataset. For a classification variable with multiple classes, entropy is calculated based on the probability distribution of those classes, representing the average information needed to predict class membership (Kaliappan et al., 2024). Higher entropy indicates greater disorder and unpredictability, while lower entropy suggests that one class dominates, making predictions more straightforward. In medical diagnosis, entropy quantifies the level of heterogeneity within patient data, guiding clinicians and researchers toward more refined risk stratification. Furthermore, Information Gain measures how much an attribute helps

reduce uncertainty about the target class by comparing the entropy of the class label before and after considering a specific feature. A higher IG value means that the feature provides greater discriminatory power and is more useful in separating the classes (Sreehari et al.,2024).

Moreover, In healthcare and diabetes prediction research, entropy and IG have been proven valuable for identifying the most clinically relevant biomarkers from potentially large sets of measurements. Kaliappan et al. (2024) demonstrated that Information Gain-based feature selection effectively reduces dimensionality and improves classification accuracy in diabetes datasets by prioritizing attributes such as glucose levels, BMI, and age. Similarly, Sreehari et al. (2024) applied information gain alongside chi-square and recursive feature elimination techniques to analyze critical factors in diabetes mellitus prediction, achieving improved F1 scores and accuracy by focusing on the most discriminative features. In a study of Sirmayanti et al. (2025) further enhanced diabetes prediction performance by integrating advanced feature selection strategies based on a grey wolf optimizer with an autophagy mechanism, showing that targeted selection of informative biomarkers can significantly improve model reliability and interpretability.

In addition, the context of menopausal women specifically, entropy and IG analysis can reveal which biomarkers such as Fasting Blood Sugar, HbA1c, lipid profiles, and anthropometric indicators are most discriminative for distinguishing diabetes risk across different hormonal and metabolic states. This targeted feature selection ensures that downstream machine learning models focus on variables that meaningfully contribute to risk stratification, supporting both scientific rigor and practical usability in clinical settings.

Clustering

Clustering has emerged as a central unsupervised machine learning technique for patient stratification in diabetes research, offering the ability to identify meaningful subgroups beyond traditional diagnostic criteria. Taurbekova et al. (2024) conducted a comprehensive systematic review and cross-sectional analysis, demonstrating that cluster analysis most notably employing k-means and hierarchical methods consistently uncovers patient subtypes with distinct clinical and metabolic characteristics that correlate with varying complication risks.

These findings have been echoed in population-specific research; for instance, Li et al. (2024) applied data-driven clustering to Chinese community diabetes populations, revealing subgroups with diverse profiles for metabolic markers, complication risks, and prevalence of comorbidities. Similarly, Lu et al. (2025) identified that in Black/African American cohorts, the severe insulin-deficient cluster is associated with a heightened risk for adverse outcomes, emphasizing the necessity of precise, data-driven risk assessments for high-risk demographic groups. In Indian populations, Tripathi et al. (2024) compared multiple clustering and phenotyping approaches and confirmed that clustering allows for finer subclassification, which in turn supports the prediction of remission rates and tailored interventions.

Within DIANA, cluster analysis thus serves a critical function by enabling the model to group menopausal women according to shared biomarker patterns and clinical characteristics. This approach aids in visualizing and interpreting heterogeneous risk profiles, providing clinicians with actionable insights for personalized diabetes risk management. Integrating clustering with predictive analytics enhances both the

interpretability and precision of digital health solutions designed for at-risk populations, particularly when applied to the nuanced context of menopausal women's health.

Synthesis

Numerous studies illustrate the progression of diabetes research, highlighting global prevalence, risk factors, and predictive approaches for Type 2 Diabetes Mellitus (T2DM) in menopausal women. Specifically, Diabetes Mellitus is a chronic metabolic disorder characterized by elevated blood glucose due to insufficient insulin production or impaired insulin utilization. While Type 1 Diabetes arises from autoimmune destruction of pancreatic beta cells, Type 2 Diabetes primarily results from insulin resistance, influenced by lifestyle factors such as diet, obesity, and physical inactivity. In the Philippines, diabetes represents a growing public health concern, with both diagnosed and undiagnosed cases increasing steadily. Importantly, women undergoing menopause are particularly vulnerable due to lifestyle factors and limited awareness of menopausal health.

Blood biomarkers and predictive modeling (for example, predicting Type 2 Diabetes risk or complications in subpopulations such as menopausal women), this literature offers several actionable lessons: use multi-dimensional phenotyping (insulin resistance markers, beta-cell function proxies, adiposity measures), consider clustering as a way to reduce heterogeneity in both model development and validation, and evaluate whether subgroup membership improves predictive performance for outcomes of interest beyond conventional risk factors. The epigenetic results suggest that adding molecular features could increase discrimination for complication risk, but systematic, cost-effective biomarker selection and validation (as Ao et al. recommend) will be essential before routine adoption

The cluster-based reclassification of adult-onset diabetes represents a promising shift toward biologically informed precision medicine. Ahlqvist et al. established clinically meaningful subgroups; Schrader et al. provided molecular validation; Veelen et al. mapped therapeutic implications; and Ao et al. synthesized the emerging evidence base. Together they create a framework for stratified risk prediction, targeted therapy, and future research — while clearly indicating the need for broader validation, mechanistic work, and prospective trials to confirm that subgroup-driven care improves patient-centered outcomes.

Moreover, menopause represents a critical physiological transition that significantly affects glucose metabolism, lipid profiles, and insulin sensitivity. During this period, hormonal changes, especially early or premature menopause, increase the risk of developing T2DM. Additionally, changes in body fat distribution and central adiposity further exacerbate insulin resistance. Taken together, these findings underscore the need for targeted preventive strategies that consider the unique metabolic and hormonal profiles of menopausal women.

Consequently, biomarkers such as Fasting Blood Sugar (FBS), Hemoglobin A1c (HbA1c), and lipid profiles, along with age, body mass index (BMI), menopausal status, and lifestyle factors, play a critical role in assessing diabetes risk. Moreover, local clinical recommendations include additional tests such as Chem 12/15 panels, thyroid hormone, and urinalysis to evaluate metabolic function. By integrating these biomarkers into computational models, researchers enable early identification of at-risk menopausal women and support personalized interventions.

In addition, predictive modeling and machine learning techniques gain increasing relevance in identifying individuals at risk for T2DM. For instance, studies show that models using biomarkers like BMI, HbA1c, triglycerides, and LDL achieve strong predictive performance, while algorithms such as Random Forest and Gradient Boosting effectively capture complex

interactions among clinical and demographic variables. Furthermore, incorporating menopausal status and lifestyle behaviors enhances model accuracy for this specific population. Together, these approaches highlight the potential of predictive tools to support early detection and risk management, addressing limitations in traditional diabetes screening.

When considered collectively, the reviewed literature emphasizes that although biomarkers, predictive modeling, and machine learning each provide valuable insights, none fully address the need for a comprehensive, user-friendly tool for identifying menopausal women at risk of T2DM. Therefore, these gaps form the foundation for the present study, which introduces DIANA, a predictive model-based application that integrates clinically relevant biomarkers, demographic data, and lifestyle factors to identify high-risk menopausal women, enabling timely interventions and evidence-based preventive strategies.

Methodology

Research Design

This study employs a quantitative research design to develop and evaluate a predictive classification model for identifying menopausal women at risk of Type 2 Diabetes using selected blood biomarkers. The quantitative framework is appropriate because it enables the collection and analysis of numerical biomarker data, measurement of model performance through statistical metrics, and objective evaluation of system usability through structured instruments. The study uses non-probability purposive sampling to select clinical data, contextual participants, and medical evaluators based on their relevance to the research objectives and availability of complete, structured biomarker records.

The research involves collecting clinical biomarker data from Philippine hospital records, developing a machine learning-based predictive model through feature selection (entropy and Information Gain), supervised classification, and clustering, and integrating the model into a web-based application. The system will be evaluated by licensed physicians to assess its clinical applicability, usability, and interpretability in supporting diabetes risk detection for menopausal women.

This design aligns with the study's objectives of identifying the most informative biomarker attributes, developing ML-based classification and clustering methods for risk prediction and group profiling, and validating the system's clinical utility in the Philippine healthcare context.

Research Locale

This study will be conducted through collaboration with selected healthcare institutions in the Philippines that maintain systematic electronic health record systems or structured medical documentation for women aged 45 to 60 years who have transitioned through menopause. Priority will be given to hospitals with established departments in obstetrics-gynecology (OB-GYN) and endocrinology, as these medical specialties routinely document the metabolic and hormonal biomarkers central to this study including fasting blood sugar, HbA1c, lipid panels, and related metabolic indicators. The researchers are actively coordinating with three prospective hospital partners to obtain institutional clearance for data access. Final determination of participating institutions will depend on successful acquisition of formal authorization, confirmation of adequate patient record availability meeting inclusion criteria (targeting approximately 1,000 to 2,000 de-identified records), and alignment with the study's ethical and methodological requirements. The selected healthcare facilities will provide a clinically representative sample of Filipino menopausal women, ensuring that the predictive model reflects the demographic, metabolic, and hormonal characteristics specific to this population.

The "Usapang Perimenopause at Menopause" Facebook interest group serves as the locale for user acceptance testing of the DIANA web application. This online community of Filipino women actively discussing menopause-related health topics provides access to the target end-user population who will evaluate the application's usability, clarity, and practical relevance. Volunteer members who meet the study's inclusion criteria will be invited to test the application and provide structured feedback on its interface design, information presentation, and usefulness for personal health monitoring. Engagement and data collection from the group will only commence upon

receipt of formal permission and cooperation from group administrators, ensuring compliance with ethical standards and data privacy regulations.

The clinical evaluation phase of the study will be conducted in the practices or offices of licensed endocrinologists and OB-GYN specialists participating as expert evaluators. These settings will allow the clinicians to systematically review and assess the DIANA application's usability, clinical validity, and relevance for routine patient care. Their feedback will be critical for determining the clinical acceptability and practical integration of the DIANA system into real-world healthcare workflows in the Philippine context.

Population of the Study

The study population consists of three distinct groups that contribute to different phases of the research, selected using non-probability purposive sampling based on their relevance to the study objectives.

The primary modeling dataset comprises de-identified clinical records of menopausal women aged 45 to 60 years obtained from partner hospitals in the Philippines, targeting approximately 1,000 to 2,000 complete records with documented biomarker data (Fasting Blood Sugar, HbA1c, lipid parameters, age, BMI, menopausal status, and family history). Hospitals and records will be selected to include institutions that routinely collect these biomarkers and maintain sufficient data quality for machine learning analysis.

Members of the "Usapang Perimenopause at Menopause" Facebook interest group will participate as end-user evaluators during the user acceptance testing phase of the study. At least 30 participants will be purposively selected from women who meet the study's age and menopausal status criteria. These participants, purposively selected from women who meet the study's age and

menopausal status criteria, will interact with the DIANA web application and provide feedback on its usability, understandability, and relevance to their health management needs through a structured survey. Their involvement ensures that the application is designed to meet the practical needs and preferences of its intended users in the Filipino context.

Finally, licensed physicians with expertise in endocrinology, obstetrics-gynecology, or internal medicine will participate as clinical evaluators of the DIANA web application during Phase 5 of the study. At least two (2) expert validators will be purposively selected from the pool of doctors previously interviewed during the earlier phases of the research. These medical professionals will be purposively sampled based on their clinical experience with menopausal women and Type 2 Diabetes, and will assess the clinical applicability, usability, and interpretability of the system's risk predictions and visualizations using a structured Likert-scale survey. Their feedback ensures that the tool aligns with clinical practice needs and supports informed decision-making in the Philippine healthcare context.

Data Gathering Tools and Procedures

The study will utilize de-identified patient records from selected hospitals in the Philippines as the primary source of data for model development. These records will contain clinical biomarker measurements and demographic information necessary for training the predictive classification model.

Blood Biomarkers. The following blood-based clinical biomarkers will be collected from patient records: Fasting Blood Sugar (FBS), Glycated Hemoglobin (HbA1c), Triglycerides (TG), Low-Density Lipoprotein Cholesterol (LDL-C), High-Density Lipoprotein Cholesterol (HDL-C), and Total Cholesterol (TC). These biomarkers were selected based on their documented

association with metabolic dysfunction and Type 2 Diabetes risk, as identified in the literature review and validated through consultations with medical experts.

Non-Blood Biomarkers and Demographic Variables. In addition to blood biomarkers, the following non-blood clinical indicators and demographic variables will be extracted from patient records: Age, Body Mass Index (BMI), Menopausal Status, and Family History of Diabetes. These variables provide essential contextual information that influences diabetes risk and will serve as supplementary features for the predictive model.

All data will be extracted from hospital electronic health records or physical medical charts through formal coordination with hospital administrations. Patient records will be anonymized prior to extraction to ensure compliance with data privacy regulations and ethical research standards.

Variable	Type	Coding / Unit	Source	Missing-Data Rule / Notes
Fasting Blood Sugar (FBS)	Continuous	mg/dL	Hospital lab record	Records missing FBS are excluded from model training.
Hemoglobin A1c (HbA1c)	Continuous	%	Hospital lab record	Records missing HbA1c are excluded from model training.
Triglycerides (TG)	Continuous	mg/dL	Hospital lab record	Retained if core glycemic and lipid fields are complete.
Low-Density Lipoprotein (LDL-C)	Continuous	mg/dL	Hospital lab record	Retained if core glycemic and lipid fields are complete.

High-Density Lipoprotein (HDL-C)	Continuous	mg/dL	Hospital lab record	Retained if core glycemic and lipid fields are complete.
Total Cholesterol (TC)	Continuous	mg/dL	Hospital lab record	Retained if core glycemic and lipid fields are complete.
Age	Continuous	Years	Patient demographic record	Records missing age are excluded from the final dataset.
Body Mass Index (BMI)	Continuous	kg/m²	Computed from height/weight	Exclude records with missing or implausible BMI values.
Menopausal Status	Categorical	Perimenopausal/ Postmenopausal	Clinical record	Only menopausal women (45–60 years) are retained.
Family History of Diabetes	Categorical	Yes / No	Clinical / family history	Records with undocumented family history are treated as missing and excluded from modeling.
Glycemic Class Label (Outcome Y)	Categorical	Non-Diabetic/ Prediabetic/ Diabetic	Derived from FBS & HbA1c	Derived using clinical cut-offs defined in Chapter 2; records with inconsistent or missing labels removed.

Table 3: Dataset Composition: Blood Biomarkers and Demographic Variables

In the hospital dataset used for model development, structured fields are available for blood biomarkers, age, BMI, menopausal status, and family history of diabetes, which are all included

as predictive features. In contrast, lifestyle-related factors such as detailed diet, physical activity patterns, and smoking history are primarily discussed in the Review of Related Literature and medical interviews but are not consistently encoded as structured variables in the hospital records, so they are not directly used as input features in the current predictive models.

Software Methodology

The development of the DIANA predictive model is anchored in a rapid prototyping methodology, which empowers iterative improvement based on the active involvement of stakeholders and end-users. This approach allows for the swift creation of functional prototypes, facilitating feedback collection from healthcare professionals, and enabling ongoing refinement of both the computational model and its interface. By systematically advancing through each stage from requirement gathering and data handling to technical and clinical validation the methodology ensures the solution remains responsive to practical healthcare needs, data privacy standards, and clinical effectiveness. The phased structure provides clarity, traceability, and adaptability, guiding the project from inception to real-world readiness.

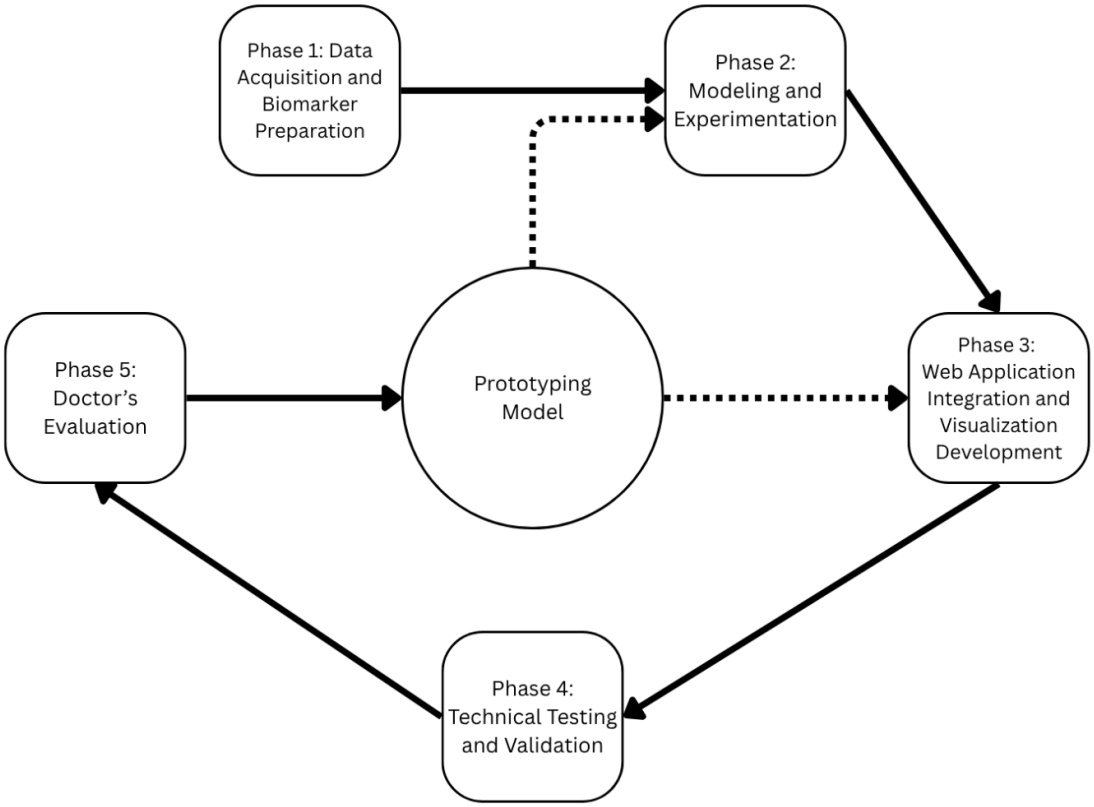


Figure 2: General Prototyping Model

Phase 1: Data Acquisition and Biomarker Preparation

This phase focuses on collecting, cleaning, and preparing clinical data from partner hospitals to build the dataset used for feature selection and model development. De-identified records of menopausal women aged 45–60 will be obtained through formal coordination, targeting approximately 1,000–2,000 records with documented glycemic and lipid results. Only records with complete core biomarkers and key demographic fields will be retained to ensure data quality and ethical compliance.

The dataset will include blood biomarkers (Fasting Blood Sugar, HbA1c, Triglycerides, LDL-C, HDL-C, Total Cholesterol) and non-blood variables (Age, BMI, Menopausal Status,

Family History of Diabetes) as the core predictors, with lifestyle-related variables (such as smoking history, recent physical activity, and comorbidities like hypertension or heart disease) considered as candidate inputs when they are available in structured categorical or numerical form in the records. Each variable will be classified by data type (continuous or categorical), checked for outliers and unit inconsistencies, and evaluated for completeness. Variables, including any lifestyle-related fields, with at least 70% non-missing values will be prioritized for feature selection, model training, and interpretation, while highly incomplete or unstructured lifestyle information will be used only to descriptively characterize clusters and contextualize the findings, and may be excluded from the final predictive model.

A glycemic status label (non-diabetic, prediabetic, diabetic) will be assigned to each record using established FBS and HbA1c cut-offs summarized in Chapter 2, and this label will serve as the class variable *Y* for both feature selection and supervised learning. Records with inconsistent or missing information for defining this label will be removed from the analytic dataset. Continuous predictors will then be discretized into clinically meaningful categories (for example, normal, borderline, and high ranges for FBS, HbA1c, lipids, and BMI) to support entropy and Information Gain computation.

Using the cleaned and discretized dataset, entropy and Information Gain will be applied to rank all candidate attributes according to how strongly they help distinguish the glycemic classes. The procedure is as follows:

1. Compute the overall entropy $H(Y)$ of the class label using the full dataset.
2. For each attribute Xj, compute the conditional entropy $H(Y|Xj)$ based on its discrete categories or bins.

3. Calculate the Information Gain $IG(Y,Xj)=H(Y)-H(Y|Xj)$ and rank attributes from highest to lowest *IG*.
4. Use the top-ranking attributes as the core feature set for Phase 2 model training and for generating "risk factor importance" visualizations in the DIANA Analytics tab.

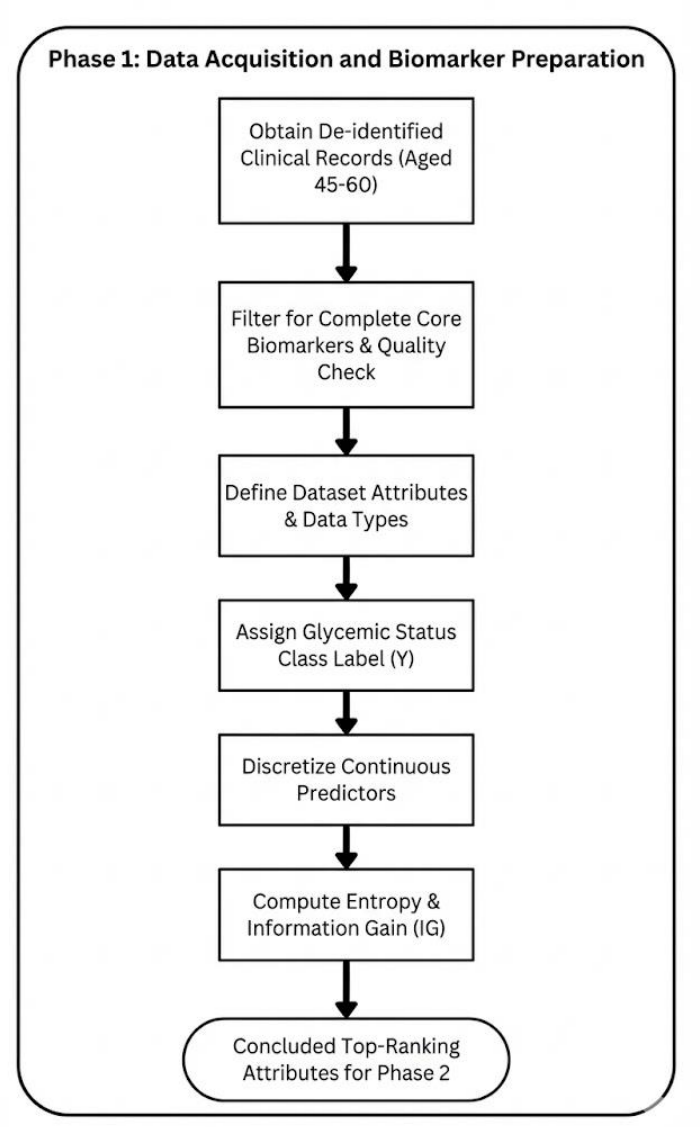


Figure 3: Data Acquisition and Biomarker Preparation Phase Flow

Phase 2a: Model Development and Training

This phase focuses on building the predictive model using the prepared dataset from Phase 1. The process begins with feature selection using entropy and Information Gain to identify the most informative attributes from clinical biomarkers and demographic variables. The selected features, which include key blood biomarkers such as Fasting Blood Sugar (FBS), Hemoglobin A1c (HbA1c), lipid profiles, and non-blood variables like age, BMI, and menopausal status, serve as inputs to machine learning algorithms.

Supervised classification models including Logistic Regression, Random Forest, and XGBoost are trained using the selected features. The cleaned dataset will be randomly partitioned using stratified sampling into a training set (70%) and an independent test set (30%). The 30% subset will serve as the actual held-out test set and will not be used during model development or cross-validation. Within the 70% training portion, k-fold cross-validation will be applied to tune hyperparameters, compare algorithms, and obtain stable internal performance estimates before selecting the final model to evaluate the unseen test set. Model training emphasizes techniques that balance predictive accuracy with clinical interpretability and computational efficiency to facilitate practical integration into medical decision-making tools.

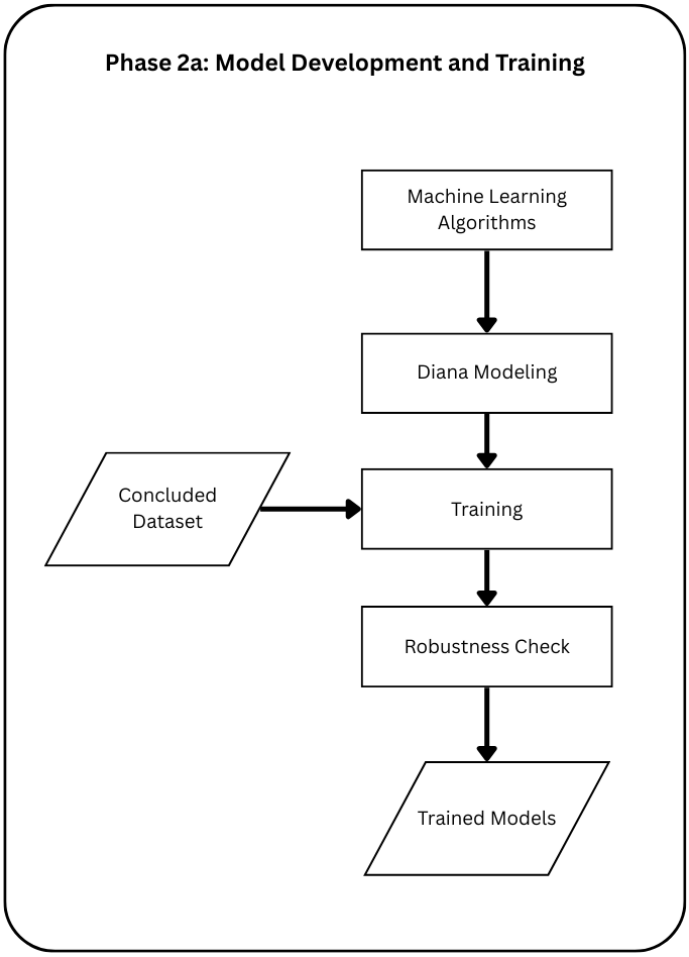


Figure 4: Model Development and Training

Phase 2b: Model Testing, Evaluation, and Comparison

Phase 2b emphasizes the rigorous validation and comparison of trained models to ensure clinical relevance and reliability. Models are evaluated using standard metrics such as accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). Special focus is placed on AUC-ROC, given its importance in balancing sensitivity and specificity in a medical context where accurate discrimination between at-risk and non-risk patients is critical, making it a widely used and clinically relevant metric in medical machine learning.

Beyond statistical performance, models are also assessed for clinical interpretability and feasibility of implementation in healthcare settings. The final model selection considers a combination of predictive performance, ease of interpretation by clinicians, and computational efficiency for real-time application. The testing procedures include evaluation on held-out datasets and cross-validation to ensure consistent performance across different patient subgroups.

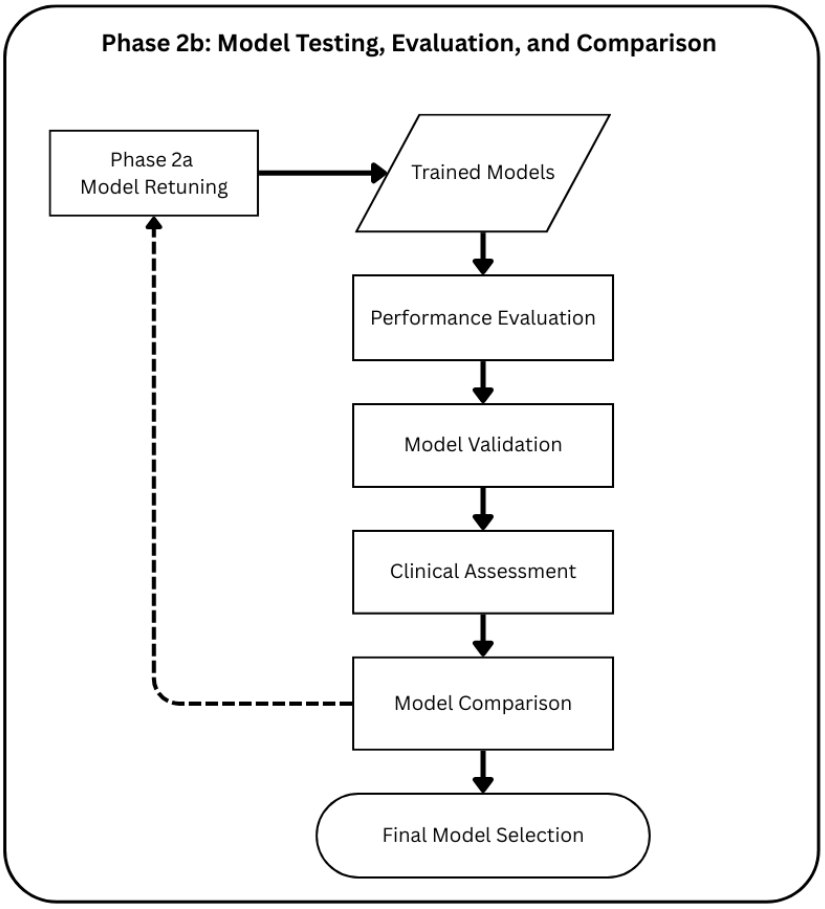


Figure 5: Model Testing, Evaluation and Comparison

Phase 3: Web Application Integration and Visualization Development

This phase focuses on integrating the trained predictive model into a web-based application using a suitable web framework. The application will feature an interactive dashboard for risk prediction, biomarker visualization, and patient history tracking. Core functionalities will include

a patient management system, risk prediction interface with probability outputs, and data visualization tools to display biomarker trends and risk levels. Secure authentication and role-based access control will also be implemented to ensure data confidentiality and appropriate system access for healthcare professionals.

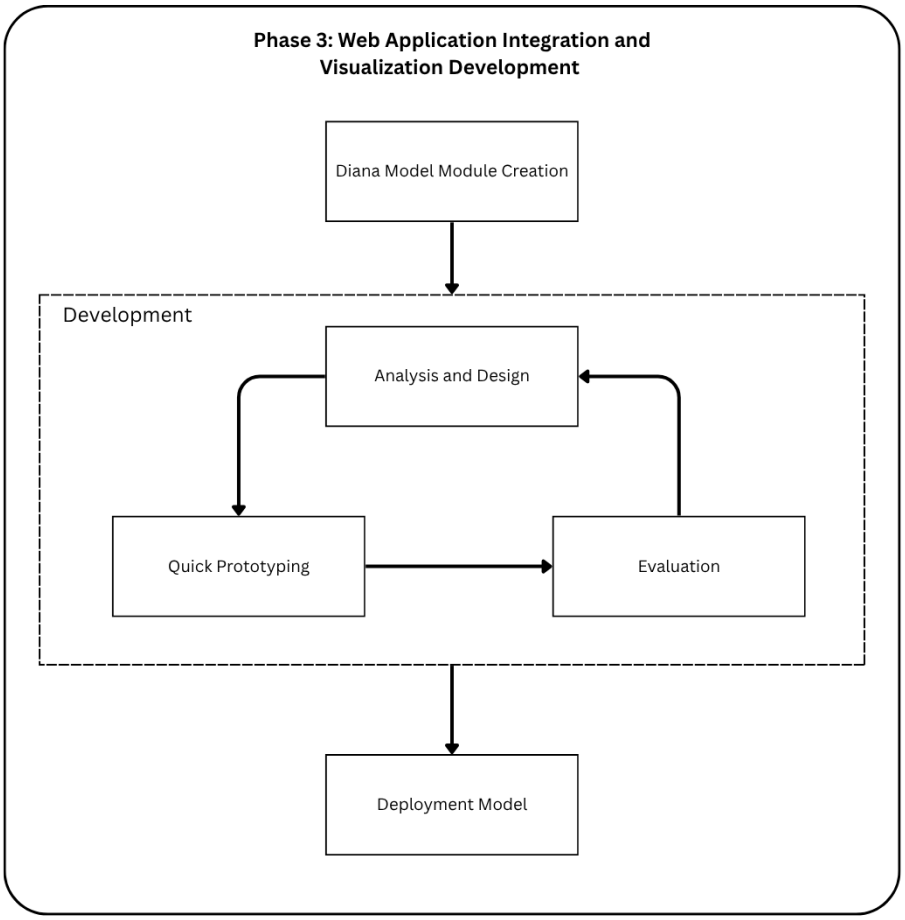


Figure 6: Web Application Integration and Visualization Development Phase Flow

Phase 4: Technical Testing and Validation

This phase encompasses comprehensive evaluation of the system's technical performance and reliability. Functional testing will verify feature accuracy and system performance. Performance testing will measure response times and system stability under typical usage

conditions. Additionally, the predictive model accuracy will be validated using the test dataset reserved during Phase 2, ensuring the system meets required standards for clinical deployment.

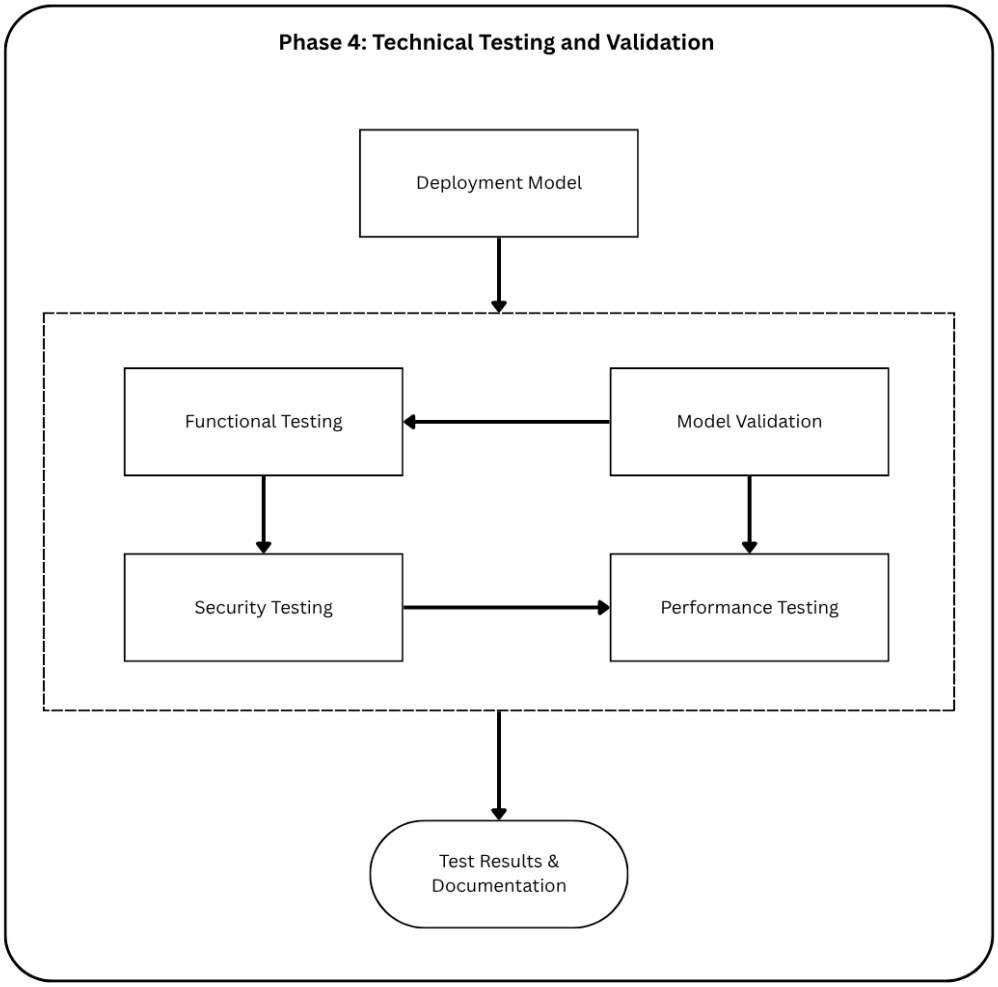


Figure 7: Technical and Validation Phase Flow

Phase 5: Doctor's Evaluation

This phase involves conducting evaluation sessions with licensed physicians to assess the clinical appropriateness of the model's risk predictions and the usability of the web application. Feedback will be gathered regarding the accuracy of risk categorization, interpretability of visualizations, and compatibility with clinical workflows. Based on this feedback, necessary

refinements will be implemented to enhance the application's effectiveness and relevance for clinical use. This iterative process ensures the system aligns with real-world healthcare needs.

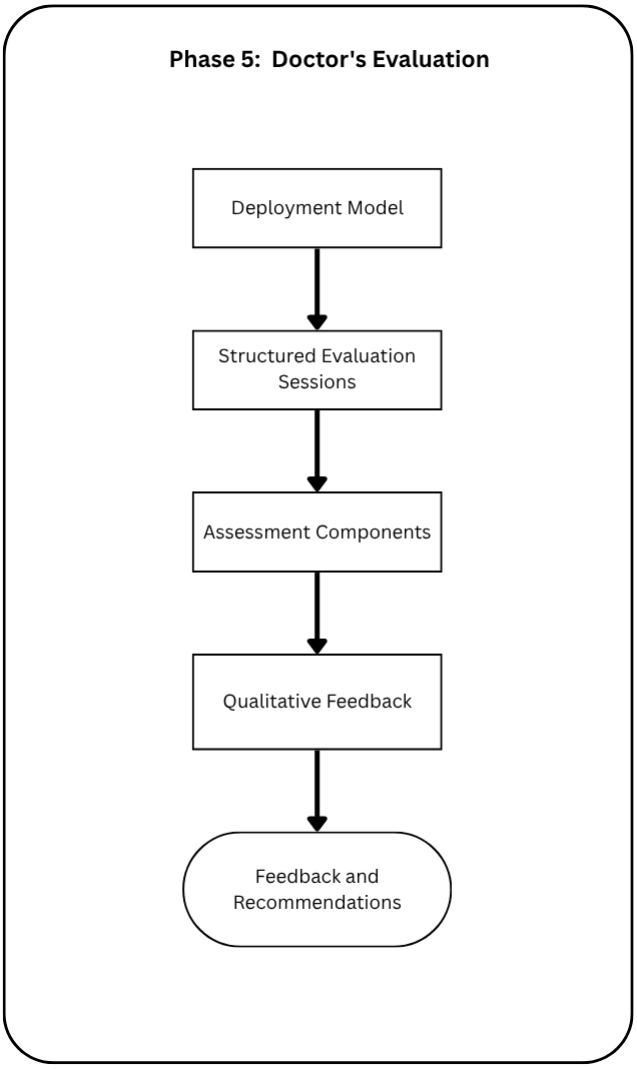


Figure 8: Doctor's Evaluation Phase Flow

Data Analysis

The data analysis phase will involve training and evaluating machine learning algorithms to develop a predictive classification model for identifying menopausal women at risk of Type 2

Diabetes. The collected biomarker data will be processed, split into training and testing sets, and used to compare the performance of multiple supervised learning algorithms.

Feature Selection using Entropy and Information Gain. Before training the predictive models, the study will perform feature selection to identify which biomarkers and related variables are most informative for classifying current glycemic status among menopausal women. Using the cleaned and discretized dataset from Phase 1, entropy and Information Gain will be computed for each candidate attribute that meets the predefined completeness threshold of at least 70% non-missing values.

Let *Y* denote the *class* label (non-diabetic, prediabetic, diabetic) and let C be the set of possible classes. The entropy of *Y* is defined as

$$H(Y) = -\sum_{c \in C} p(c) \log_2 p(c)$$

Equation 1: Entropy of Y

where $p(c)$ is the proportion of records belonging to class c. For a given attribute X with discrete values or bins V, the conditional entropy of Y given X is

$$H(Y \mid X) = \sum_{v \in V} p(v)\, H(Y \mid X = v)$$

Equation 2: The Conditional Entropy of Y given X

Where $p(v)$ is the proportion of records with $X = v$, and $H(Y \mid X = v)$ *is* the entropy of the class labels within that subset. The Information Gain of X with respect to Y is then

$$\boldsymbol{IG\,(Y, X) = H(Y) - H(Y|X)}$$

Equation 3: Information Gain of X with Respect to Y

Which measures how much knowing the value of *X* reduces uncertainty about the glycemic class.

In this study, Information Gain will be computed for each biomarker and non-blood variable (e.g., Fasting Blood Sugar, HbA1c, lipid parameters, age, BMI, menopausal status, family history, and any sufficiently complete lifestyle fields). The analysis will proceed as follows:

1. Compute using the overall distribution of glycemic classes in the dataset.
2. For each attribute Xj, compute the conditional entropy $H(Y \mid Xj)$ and then the Information Gain $IG(Y, Xj) = H(Y) - H(Y \mid Xj)$.
3. Rank all attributes from highest to lowest $IG(Y, Xj)$.
4. Use the top-ranking attributes as the core feature set for model training in Phase 2 and as the basis for the risk-factor importance visualizations in the DIANA Analytics tab.

Machine Learning Algorithms. The study will apply supervised machine learning algorithms to develop a predictive classification model for identifying menopausal women at current risk of Type 2 Diabetes. Each model will be trained using the feature set selected through the entropy and Information Gain procedure, ensuring that only the most informative biomarkers and related variables are used as inputs. The cleaned dataset will be randomly partitioned using stratified sampling into a training set (70%) and an independent test set (30%). The 30% subset will serve as the actual held-out test set and will not be used during model development or cross-validation. Within the 70% training portion, k-fold cross-validation will be applied to tune

hyperparameters, compare algorithms, and obtain stable internal performance estimates before selecting the final model.

Candidate algorithms include Logistic Regression, Random Forest, and XGBoost, with Support Vector Machines (SVM) considered if preliminary results indicate potential benefit. Logistic Regression is included for its interpretability and clinically meaningful probability outputs, Random Forest for its ability to model nonlinear relationships and handle interactions among biomarkers, and XGBoost for its strong performance in structured healthcare data and its capacity to capture complex patterns. Each algorithm will be trained on the IG-selected attributes and evaluated using accuracy, classification error rate *(1−accuracy)*, precision, recall (sensitivity), F1-score, and AUC-ROC. These metrics will be used to compare candidate models and select the final classifier for integration into the DIANA web application based on predictive performance, clinical interpretability, and computational efficiency.

Clustering Analysis. In addition to supervised classification, the study will apply clustering to group menopausal women into risk-related profiles based on the same feature set selected through the entropy and Information Gain procedure. This unsupervised analysis aims to reveal patterns in biomarkers and related attributes that may not be captured by classification alone and to support more interpretable risk stratification in the DIANA web application.

The primary clustering technique will be k-means, applied to standardized versions of the selected features. Several candidate values of k will be examined using the elbow method and silhouette scores to identify several clusters that provide a good balance between within-cluster compactness and between-cluster separation. The final clustering solution will be profiled in terms of average biomarker values and class label distributions, and these cluster profiles will be

visualized in the DIANA Analytics tab to help clinicians compare risk groups and relate them to the supervised model's predictions.

The distance metric most commonly employed in K-means is the Euclidean distance. Formally, let $X = \{x1, x2, \ldots, xn\}$ denote the set of *n* data points in a *d*-dimensional space, and let $\{\mu1, \mu2, \ldots, \mu k\}$ represent the centroids of the *k* clusters. The assignment of each data point *xi* to a cluster *Cj* is determined by minimizing the Euclidean distance:

$$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$$

Equation 4: Euclidean Distance Formula

where:

- $xi = (xi1, xi2, \ldots, xid)$ is the ith data point,
- $\mu j = (\mu j1, \mu j2, \ldots, \mu jd)$ is the centroid of the jth cluster.

At each iteration, the K-means algorithm operates in two main steps:

1. **Assignment Step:** Each data point *xi* is assigned to the cluster *Cj* whose centroid μj is nearest, as measured by $d(xi, \mu j)$:

$$S_i^{(t)} = \left\{ x_p : \left\| x_p - m_i^{(t)} \right\|^2 \leq \left\| x_p - m_j^{(t)} \right\|^2 \ \forall j, 1 \leq j \leq \right.$$

Equation 5: Assignment Step Formula

This means for each point, choose the cluster whose centroid is the closest (usually Euclidean distance).

2. **Update Step:** The centroid μj of each cluster is recalculated as the mean of all points assigned to that cluster:

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j$$

Equation 6: Update Step Formula

This means the new centroid is the mean of all points assigned to that cluster.

The objective of K-means is to minimize the total within-cluster sum of squared errors (SSE), also referred to as the inertia or the objective function *J*:

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x} \in S_i} \left\| \mathbf{x} - \boldsymbol{\mu}_i \right\|^2$$

Equation 7: Inertia or Objective Function J Formula

where $\| xi - \mu j \|_2$ denotes the squared Euclidean distance between *xi* and its corresponding cluster centroid μj.

Model Performance Metrics. These metrics collectively serve as the model selection criteria for choosing the final classifier to be deployed in DIANA, balancing overall accuracy, classification error, and the correct identification of menopausal women at higher risk for undiagnosed Type 2 Diabetes or prediabetes. The performance of each trained model will be assessed using standard machine learning evaluation metrics. Accuracy will measure the overall proportion of correct predictions, and the Classification Error Rate (calculated as 1 − Accuracy)

will quantify the proportion of incorrect predictions while Precision will evaluate the model's ability to correctly identify women at risk (positive cases) among all predicted positive cases. Recall (Sensitivity) will assess the model's ability to detect all actual positive cases, minimizing false negatives, which is critical in healthcare applications where missing at-risk individuals can have serious consequences. The F1-Score, which balances precision and recall, will provide a single metric for comparing models. Additionally, the Area Under the ROC Curve (AUC-ROC) will be calculated to evaluate the model's ability to discriminate between diabetic/prediabetic and non-diabetic cases across varying probability thresholds. A model with an AUC above 0.80 will be considered acceptable for clinical applications.

These probability scores, ranging from 0 to 1 and displayed as 0–100% in the application, represent the model's estimated confidence that a given menopausal patient currently has undiagnosed Type 2 Diabetes or prediabetes based on her biomarker profile. All final performance metrics (accuracy, classification error rate, precision, recall, F1-score, and AUC-ROC) will be computed on this unseen 30% test set, providing an unbiased estimate of the model's real-world performance and the reliability of DIANA's risk predictions.

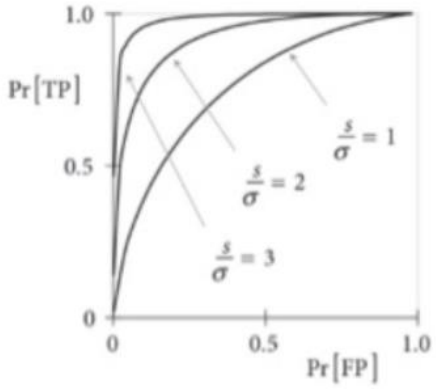
$$\boldsymbol{Accuracy} = \frac{\Sigma\ (TP + TN)}{\Sigma\ (TP + TN + FP + FN)}$$

$$\boldsymbol{Precision} = \frac{\Sigma\ TP}{\Sigma\ (TP + FP)}$$

$$\boldsymbol{Recall} = \frac{\Sigma\ TP}{\Sigma\ (TP + FN)}$$

$$\boldsymbol{F1\ Score} = \frac{\Sigma\ 2TP}{\Sigma\ (2TP + FP + FN)}$$

$$\boldsymbol{AUC} = \sum_{i=1}^{n-1} \frac{(Xi + 1 - Xi) \times)\ (Yi + Yi + 1)}{2}$$



Equation 8: Formulas for Model Performance Evaluation

Model Selection and Validation. The final model will be selected based on a combination of predictive accuracy, clinical interpretability, and computational efficiency. Models will be compared using the metrics described above, and the best-performing algorithm will be chosen for integration into the DIANA web application. To ensure robustness and prevent overfitting, cross-validation techniques will be applied during training, allowing the model to be tested on multiple subsets of the data. This approach ensures that the selected model generalizes well to unseen data and maintains reliable performance in real-world clinical scenarios.

Initial Cluster Labeling. The initial cluster labels (SOIRD, SIDD, MARD, MIDD) were identified based on relevant research and literature. These assignments reflect commonly recognized subgroups in diabetes stratification. For this study, the clusters will be further checked

and validated by the licensed physicians and endocrinologists we interviewed, ensuring each label accurately matches clinical patterns seen in our target population.

Cluster	Label	Defining Features
SOIRD	Severe Obesity-Related and Insulin-Resistant Diabetes	Highest BMI, highest HOMA-β, highest HOMA-IR; moderate HbA1c; youngest age at diagnosis
SIDD	Severe Insulin-Deficient Diabetes	Highest HbA1c, lowest HOMA-β; relatively high HOMA-IR; moderate BMI and age
MARD	Mild Age-Associated Diabetes Mellitus	Oldest age at diagnosis; moderate BMI and HbA1c; moderate insulin release and resistance
MIDD	Mild Insulin-Deficient Diabetes	Lowest BMI, HbA1c, HOMA-IR; moderate age and HOMA-β

Table 4: Initial Clustering Label

Variable Definitions and Metadata for the DIANA Study Dataset. The following table presents a proposed list of variables and their definitions for potential inclusion in the DIANA Machine Learning dataset. These variables represent key clinical, demographic, and behavioral measures of interest considered relevant to the research's aims. Please note that this is not the final dataset, but rather a compilation of variables under consideration for collection and analysis in future phases of the study.

Field Name	Type	Description
FBS	Integer	Fasting Blood Sugar (mg/dL). Value represents the participant's fasting plasma glucose measured after at least 8 hours of fasting.

HbA1c	Integer	Glycated Hemoglobin (HbA1c, %). Value represents the average blood glucose control over the past 2–3 months.
Triglycerides	Integer	Serum Triglycerides (mg/dL). Value represents the concentration of triglycerides in blood after overnight fasting.
LDL-C	Integer	Low-Density Lipoprotein Cholesterol (mg/dL). Value indicates calculated LDL cholesterol, an atherogenic lipid fraction.
HDL-C	Integer	High-Density Lipoprotein Cholesterol (mg/dL). Value represents protective HDL cholesterol levels.
Total Cholesterol	Integer	Total Serum Cholesterol (mg/dL). Value represents the sum of all cholesterol types in blood.
BMI	Integer	Body Mass Index (kg/m^2). Value calculated as weight in kilograms divided by the square of height in meters.
AGE	Integer	Age (years) of participant at the time of study enrollment.
Menopausal Status	Binary	Menopausal status: 0 = premenopausal, 1 = postmenopausal. Indicates if participant has ceased having menstrual periods for 12 consecutive months.
Family History of Diabetes	Binary	Has any biological parent or sibling been diagnosed with diabetes? 0 = no, 1 = yes.
Smoking_History	Binary	Have you smoked at least 100 cigarettes in your entire life? 0 = no, 1 = yes.
Hypertension	Binary	Has a healthcare provider ever told you that you have hypertension or high blood pressure? 0 = no, 1 = yes.

Heart_disease	Binary	Have you ever been diagnosed with coronary heart disease, angina, or myocardial infarction? 0 = no, 1 = yes.
PhysActivity	Binary	Physical activity in the past 30 days not including job-related activity: 0 = no, 1 = yes.

Table 5: Data Dictionary

References

Ahlqvist, E., et. al., (2018). Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. *The Lancet Diabetes & Endocrinology*, *6*(5), 361–369. https://doi.org/10.1016/s2213-8587(18)30051-2

Anklam, C. et. al. (2021, September 16). Oxidative and Cellular Stress Markers in Postmenopause Women with Diabetes: The Impact of Years of Menopause. *Journal of Diabetes Research*. Doi: 10.1155/2021/3314871.

Ao, N. et. al., (2025). Clinical and laboratory characteristics of novel diabetes subgroups: A systematic review and meta-analysis. *Scientific Reports*, *15*(1), 38585. https://doi.org/10.1038/s41598-025-22556-4

Auro, K. et. al. (2014). A metabolic view on menopause and ageing. *Nature Communications, 5*, 4708. https://doi.org/10.1038/ncomms5708

Azurin, J.C, et. al. (1986). Diabetes mellitus survey in the Philippines. *Philippine Journal of Public Health, 24(1)*, 1-29

Bi, Y. et. al. (2012). Advanced research on risk factors of type 2 diabetes. *Diabetes/Metabolism Research and Reviews*. *28:2*. 32-39. Doi: https://doi.org/10.1002/dmrr.2352

Campugan, M. P., & Aguaras, J. L. (2025). *Predictive modeling for diabetes classification using clinical biomarkers among Filipino adults*. Philippine Journal of Health Informatics, 14(2), 45–57.

Cando L. et. al. (2024). Current status of diabetes mellitus care and management in the Philippines. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 18(2)*. Doi: https://doi.org/10.1016/j.dsx.2024.102951

Chappel, B. (2022, October 13). Menopause and Diabetes. *DiabetesUK*. Retrieved from https://www.diabetes.org.uk/living-with-diabetes/life-with-diabetes/menopause

Chatterjee, S., et. al., (2023). Machine learning-based prediction of Type 2 Diabetes using biomarker and lifestyle data among women in midlife transition. *Journal of Medical Systems, 47(8)*, 112–124. Doi: https://doi.org/10.1007/s10916-023-01985-7

Chen, C. (2005). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377. https://doi.org/10.1002/asi.20317

Cleveland Clinic (2023, November 8). Type 2 Diabetes. *ClevelandClinic*. Retrieved from https://my.clevelandclinic.org/health/diseases/21501-type-2-diabetes

Cleveland Clinic (2024, June 24). Menopause. *ClevelandClinic*. Retrieved from https://my.clevelandclinic.org/health/diseases/21841-menopause

Congjuico, T. S. (2018). Beyond the "social" in social media: Facebook as communication lifeline. *Diliman Review, 61(1),* 87–103.

Cybulska A. et. al. (2023, March 27). Diagnostic markers of insulin resistance to discriminate between prediabetes and diabetes in menopausal women. *European Review for Medical*

and Pharmacological Sciences, 27(6), 2453-2468. Doi: https://doi.org/10.26355/eurrev_202303_31779

Cybulska, A. et al. (2023). Biomarkers in metabolic disease: Diagnostic and predictive applications. *Frontiers in Endocrinology, 14*, 118–128.

Derrou, S., et. al. (2021). The Profile of Autoimmunity in Type 1 Diabetes Patients. *Annals of African Medicine. 20(1)*. 19-23. Doi: 10.4103/aam.aam_8_20

Dhaliwal, S. (2025, January 10). Type 2 diabetes. *MedlinePlus*. Retrieved from https://medlineplus.gov/ency/article/000313.htm

Ferris, E. (2023, April 3). Preparing for Menopause: Understanding the Signs and Symptoms in All Three Stages. *Summa Health*. Retrieved from https://www.summahealth.org/flourish/entries/2023/04/preparing-for-menopause-understanding-the-signs-and-symptoms-in-all-three-stages

Fuller-Thompson E. et. al., (2017). Diabetes among non-obese Filipino Americans: Findings from a large population-based study. *Canadian Journal of Public Health / Revue Canadienne De Santé Publique, 108(1)*, E36-E42.

Giandalia, A. et. al. (2021). Adipokines and endothelial dysfunction in postmenopausal women with obesity and type 2 diabetes. *Journal of Endocrinological Investigation*, *44(6)*, 1221–1232. https://doi.org/10.1007/s40618-020-01436-7

Glazier, E & Ko, E. (2023, April 10). Pregnancy still possible during perimenopause. *UCLA Health*. Retrieved from https://www.uclahealth.org/news/article/pregnancy-still-possible-during-perimenopause

Goyal, R. et. al., (2023, January 23). Type 2 Diabetes. *National Library of Medicine*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK513253/

Gonzales, M. et. al. (2023). Menopause and metabolic risk: A review of predictive models for type 2 diabetes. *Menopause and Metabolic Health, 10*(1), 12–25.

Goycheva, P. et. al. (2023). Antioxidant enzyme activity and lipid peroxidation in postmenopausal women with type 2 diabetes. *BMC Women's Health. 23.* 47. Doi: https://doi.org/10.1186/s12905-023-02156-2

Gregory, G., et. al. (2022, October 7). Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study. *The Lancet Diabetes & Endocrinology. 10(10).* 741-760. Doi: 10.1016/S2213-8587(22)00218-2

Hu, F. et. al. (2004). Inflammatory Markers and Risk of Developing Type 2 Diabetes in Women. *Diabetes. 53 (3)*. 693-700. Doi: https://doi.org/10.2337/diabetes.53.3.693

International Diabetes Federation (n.d). Diabetes in The Philippines (2024). *International Diabetes Federation*. Retrieved from https://idf.org/our-network/regions-and-members/western-pacific/members/the-philippines/

Isip-Tan, M. I., et. al., (2020). Use of Facebook to serve the information needs of persons with diabetes in the Philippines amid the COVID-19 pandemic. *Journal of the ASEAN Federation of Endocrine Societies, 35(1),* 8–14. https://doi.org/10.15605/jafes.035.01.03

Jasum, O. H., et. al., (2022, February 28). Significance of Lipid Profile Parameters in Predicting Pre-Diabetes. *Archive of Razi Institute*. *77(1)*. 277-284. Doi: 10.22092/ARI.2021.356465.1846

Jimeno, C. A., et. al., (2019). Diabetes in the Philippines: A review of current epidemiology, management practices, and challenges. *Journal of the ASEAN Federation of Endocrine Societies, 34*(1), 2–10. Doi: https://doi.org/10.15605/jafes.034.01.01

Kaliappan, J., Saravana Kumar, I. J., Sundaravelan, S., Anesh, T., Rithik, R. R., Singh, Y., Vera-Garcia, D. V., Himeur, Y., Mansoor, W., Atalla, S., & Srinivasan, K. (2024). Analyzing classification and feature selection strategies for diabetes prediction across diverse diabetes datasets. *Frontiers in artificial intelligence*, *7*, 1421751. https://doi.org/10.3389/frai.2024.1421751

Lai, H. et. al., (2019, October 15). Predictive models for diabetes mellitus using machine learning techniques. *BMC Endocrine Disorders. 19.* 101. Doi: https://doi.org/10.1186/s12902-019-0436-6

Lapitan, A. M., & Doromal, N. A. (2015). The illusions of community in Facebook. *LPU Laguna Journal of Arts and Sciences Communication Research. 2(1).*

Lee, C. et. al. (2006). Adipokines, inflammation, and visceral adiposity across the menopausal transition: A prospective study. *The Journal of Clinical Endocrinology & Metabolism, 91*(9), 3438–3445. https://doi.org/10.1210/jc.2006-0517

Li, B., et. Al., (2024). Clinical characteristics and complication risks in data-driven clusters among Chinese community diabetes populations. *Journal of Diabetes.* Doi: https://onlinelibrary.wiley.com/doi/10.1111/1753-0407.13596

Lu, B., et. al., (2025). Data-driven cluster analysis reveals increased risk for severe insulin-deficient diabetes in Black/African Americans. *The Journal of Clinical Endocrinology & Metabolism*. Doi: https://academic.oup.com/jcem/article/110/2/387/7724237

Li H, et. al., (2014). Machine learning-based prediction of diabetic patients using blood routine data. *Methods*. *229.* 156-162. doi: 10.1016/j.ymeth.2024.07.001.

Liu, S. et. al. (2007). A Prospective Study of Inflammatory Cytokines and Diabetes Mellitus in a multiethnic cohort of postmenopausal women. *Archives of Internal Medicine*. *167 (15)*. 1676-1685. Doi: https://doi.org/10.1001/archinte.167.15.1676

Ma, X. et. al. (2024, June 14). Serum selenium and fasting blood glucose: a cross- sectional study in women of different menopause status. *BMC Women's Health. 341 (2024)*. Doi: https://doi.org/10.1186/s12905-024-03200-1

Malti, H., & Gopalakrishnan, V. (2007). Relationship of serum adiponectin with blood lipids, HbA1c, and hs-CRP in type II diabetic postmenopausal women. *Metabolism: Clinical and Experimental, 56(2),* 256–260. Doi: https://doi.org/10.1016/j.metabol.2006.09.028

Mayo Clinic (2025, February 27). Type 2 diabetes. *MayoClinic*. Retrieved from https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/symptoms-causes/syc-20351193

Mayo Clinic (2024, March27). Type 1 diabetes. *MayoClinic*. Retrieved from https://www.mayoclinic.org/diseases-conditions/type-1-diabetes/symptoms-causes/syc-20353011

Mobasseri, M., et. al. (2020, March 3). Prevalence and incidence of type 1 diabetes in the world: a systematic review and meta-analysis. *National Library of Medicine. 10(2).* 98-115. Doi: 10.34172/hpp.2020.18

Mohd Rizal, M. F., et. al., (2024) A Scoping Review of Supervised Machine Learning Techniques in Predicting the Prevalence of Type 2 Diabetes Mellitus. *Medicine & Health*, *19*(2), 380–399. https://doi.org/10.17576/mh.2024.1902.03

Nam, H., & Choi, H. (2023). Age at menopause and development of type 2 diabetes in Korean women: a nationwide cohort study. *JAMA Network Open, 6(2),* Doi: https://doi.org/10.1001/jamanetworkopen.2023.29415

National Institute of Diabetes and Digestive and Kidney Disease (2022). Risk Factors for Type 2 Diabetes. *National Institute of Diabetes and Digestive and Kidney Disease (NIH).* Retrieved from https://www.niddk.nih.gov/health-information/diabetes/overview/risk-factors-type-2-diabetes

Nabovati, E., et. al., (2022, October 31). Design, development, and usability evaluation of a smartphone-based application for nutrition management in patients with type II diabetes.

Journal of Diabetes & Metabolic Disorders. 22. 315-323. Doi: https://doi.org/10.1007/s40200-022-01140-x.

Nishida, C., et. al., (2021). Association between menopausal status and glucose metabolism: The Toranomon Hospital Health Management Center Study. *Diabetes & Metabolism Journal, 45(6),* 835–843. Doi: https://doi.org/10.4093/dmj.2021.0001

NYU Langone Health (n,d). Diagnosing Type 2 Diabetes. *NYULangone*. Retrieved from https://nyulangone.org/conditions/type-2-diabetes

Office on Women's Health (2025, March 17). Menopause Basics. *OASH.* Retrieved from https://womenshealth.gov/menopause/menopause-basics

Office on Women's Health (2025, May 13). Navigating the Road to Menopause. *OASH*. Retrieved from https://womenshealth.gov/nwhw/menopause

Omer, R. (2025, September 1). Menopause: the three stages. *Top Doctors*. Retrieved from https://www.topdoctors.co.uk/medical-articles/menopause-the-three-stages/

Paschou S., et. al., (2019). Diabetes in menopause: risks and management. *Current Vascular Pharmacology, 17(3)*, 263-272. https://doi.org/10.2174/1570161116666180625124405

Sapra, A. & Bhandari, P. (2023, June 21). Diabetes. *National Library of Medicince*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK551501/

Schrader, S., et. al., (2022). Novel subgroups of Type 2 diabetes display different epigenetic patterns that associate with future diabetic complications. *Diabetes Care*, *45*(7), 1621–1630. https://doi.org/10.2337/dc21-2489

Sharma, S. et. al., (2020). Cardiometabolic risk in pre- and post-menopausal women with special reference to insulin resistance: A cross-sectional study. *Journal of Mid-Life Health*, *11(2)*, 78–83. Doi: https://doi.org/10.4103/jmh.jmh_53_19

Sirmayanti, Pulung Hendro Prastyo, & Mahyati. (2025). Enhancing diabetes prediction performance using feature selection based on grey wolf optimizer with autophagy mechanism. *Computer Methods and Programs in Biomedicine Update, 8*, 100207. https://doi.org/10.1016/j.cmpbup.2025.100207

Slopien, et al., (2018). Menopause and diabetes: EMAS clinical guide. *Maturitas. 117*. 6-10. Doi: https://doi.org/10.1016/j.maturitas.2018.08.009

Spencer, C. (2024, November 11). Navigating diabetes through the menopause transition. *My Menopause Center*. Retrieved from https://www.mymenopausecentre.com/gp-resources/navigating-diabetes-through-the-menopause-transition-what-you-need-to-know-about-perimenopause-postmenopause-and-hrt/?

Sreehari, E., & Babu, L. D. D. (2024). Critical Factor Analysis for prediction of Diabetes M ellitus using an Inclusive Feature Selection Strategy. *Applied Artificial In telligence, 38(1).*

Sy, R. G., et al. (2022). Current practices in diabetes management in the Philippines: Insights from the 2021 National Survey on Non-Communicable Diseases. *Philippine Journal of Internal Medicine, 60*(2), 45–56.

Tamakoshi, K., et. al., (2006). Menopause, but not age, is an independent risk factor for fasting plasma glucose levels in nondiabetic women. *Menopause*, *13(1),* 93–98. Doi: https://doi.org/10.1097/01.gme.0000204467.48730.04

Tan, G. (2015). Diabetes Care in the Philippines. *Annals of Global Health. 81(6)*. 863-869. Doi: https://doi.org/10.1016/j.aogh.2015.10.004

Taurbekova, B., et. al., (2024). Cluster analysis in diabetes research: A systematic review enhanced by a cross-sectional study. *Frontiers in Endocrinology*. Doi: https://pmc.ncbi.nlm.nih.gov/articles/PMC12112067/

The Medical City (n.d). Diabetes Testing in the Philippines. *themedicalcity.* Retrieved from https://www.themedicalcity.com/tmc-institutes/wai/patient-services/preventive-medicine/diabetes-test

Tripathi, P., et. al., (2024). Comparison of clustering and phenotyping approaches for subclassification of type 2 diabetes and its association with remission in Indian population. *Scientific Reports*. Doi: https://www.nature.com/articles/s41598-024-71126-7

UNAIDS. The Gap Report 2014: People aged 50 years and older. Geneva, Switzerland. (2014). UNAIDS. *Uniaids.* https://www.unaids.org/sites/default/files/media_asset/12_Peopleaged50yearsandolder.pdf

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Ageing 2019: Highlights (ST/ESA/SER.A/430). Retrieved from https://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeing2019-Highlights.pdf

Veleen, A., et. al., (2021). Type 2 diabetes subgroups and potential medication strategies in relation to effects on insulin resistance and beta cell function: A step toward personalised diabetes treatment?, *Molecular Metabolism*, *46*, 2021, Doi: https://doi.org/10.1016/j.molmet.2020.101158

Villa, M. (2017, September 5). What is my Blood Sugar Target. *St. Luke's Medical Center*. Retrieved from https://www.stlukes.com.ph/health-library/health-articles/what-is-my-blood-sugar-target

Vladimirovich, K. (2020). Visualization of information in the educational Process: Current trends. (2020). *Systematic Reviews in Pharmacy*, *11(04),* 01 – 05. Doi: https://doi.org/10.31838/srp.2020.4.01

Wang, N. et. al. (2015). Follicle-stimulating hormone associates with prediabetes and diabetes in postmenopausal women. *Acta Diabetologica, 53*(2), 227–236. https://doi.org/10.1007/s00592-015-0765-6

World Health Organization (2024, October 16). Menopause. *World Health Organization*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/menopause

World Health Organization (2024, November 14). Diabetes. *World Health Organization*. Retrieved from https://www.who.int/news-room/fact-sheets/detail/diabetes

Wu, Y., Ding, Y., Tanaka, Y., & Zhang, W. (2014). *Risk Factors Contributing to Type 2 Diabetes and Recent Advances in the Treatment and Prevention*. *International Journal of Medical Sciences, 11*(11), 1185-1200. https://doi.org/10.7150/ijms.10001

Wyne, K. (2025, March 4). The link between menopause and diabetes. *Health & Discovery*. Retrieved from https://health.osu.edu/wellness/exercise-and-nutrition/link-between-menopause-and-diabetes

Xiaoxue, W., et. al. (2024). Risk prediction model of metabolic syndrome in perimenopausal women based on machine learning. *International Journal of Medical Informatics.* Doi: https://doi.org/10.1016/j.ijmedinf.2024.105480

Yazdkhasti, M., Hashemi, S., Peyvandi, S., & Shafiee, M. (2024). The association between age of menopause and type 2 diabetes: A systematic review and meta-analysis. Nutrition & Metabolism, 21(1), 58. https://doi.org/10.1186/s12986-024-00858-0

Zahra R, et al., (2014, December 16). Advanced Predictive Modeling of Type 2 Diabetes Using XGBoost and Explainable AI, *Research Square.* Doi: https://doi.org/10.21203/rs.3.rs-5337562/v1

Zhao, Y., Liu, J., Zhang, N., & Wang, Y. (2022). Age at natural menopause and risk of Type 2 Diabetes: Findings from a 10-year prospective cohort study of 300,000 Chinese women. *Diabetologia, 65(9),* 1556–1568. https://doi.org/10.1007/s00125-022-05762-1

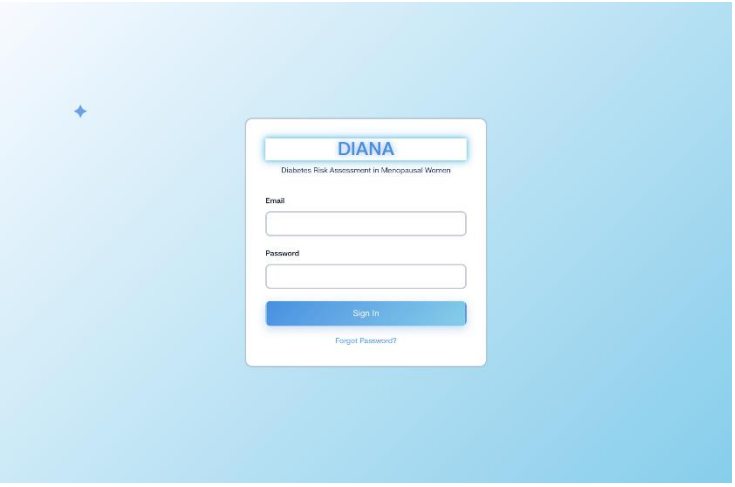
Zou, D., et. al. (2016, July 6). Analysis of risk factors and their interactions in type 2 diabetes mellitus: A cross-sectional survey in Guilin, China. *Journal of Diabetes Investigation. 8(2)*. 188-194. Doi: 10.1111/jdi.12549

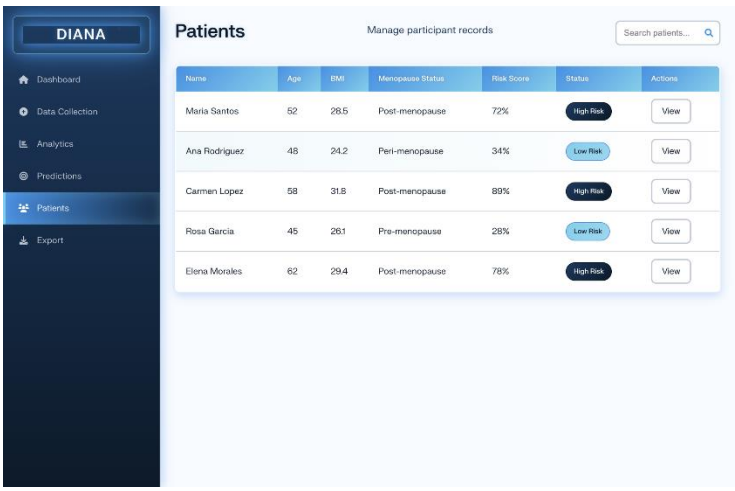
Zudilova-Seinstra, E., et. al., (2008). Trends in interactive Visualization. In *Advanced information and knowledge processing*. Doi: https://doi.org/10.1007/978-1-84800-269-2

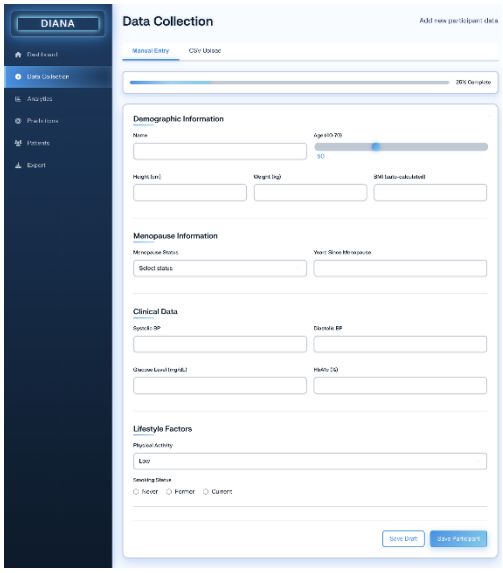
Appendices

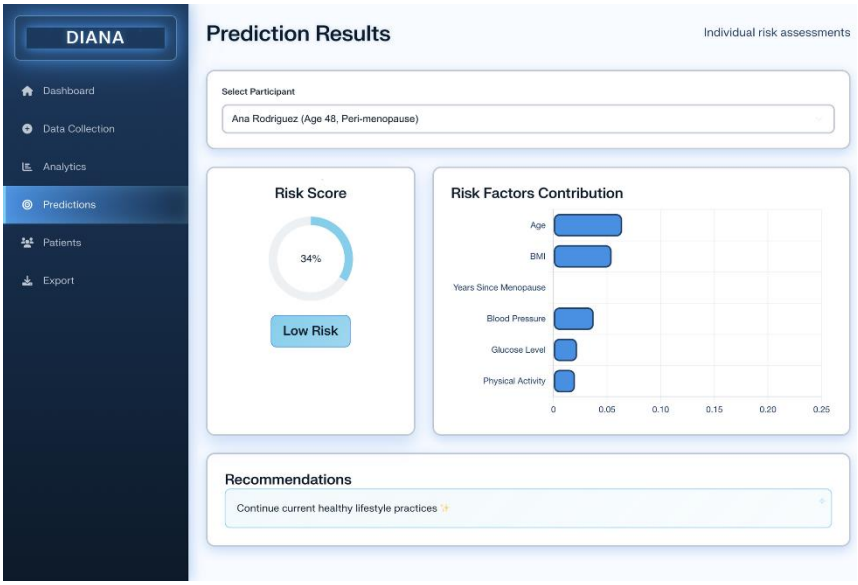
Appendix A

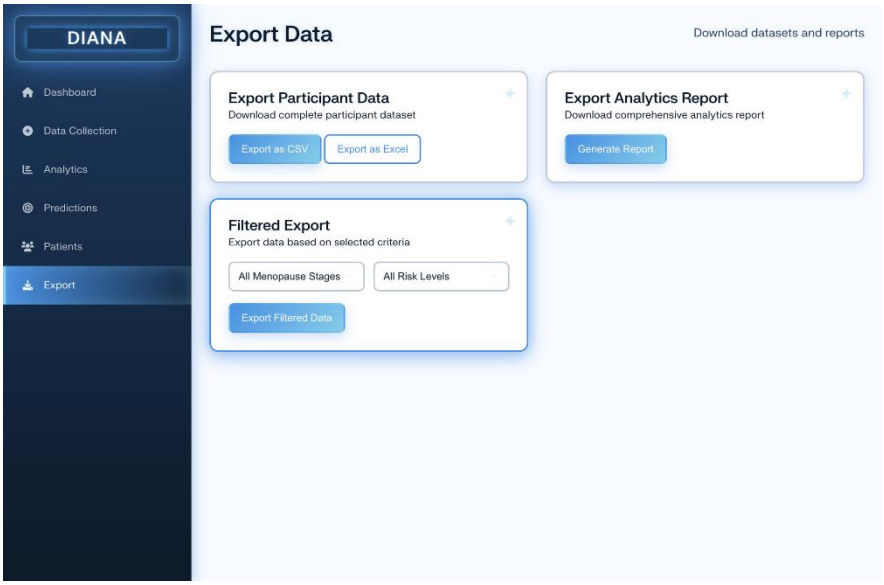
Prototype of the Web Application Features

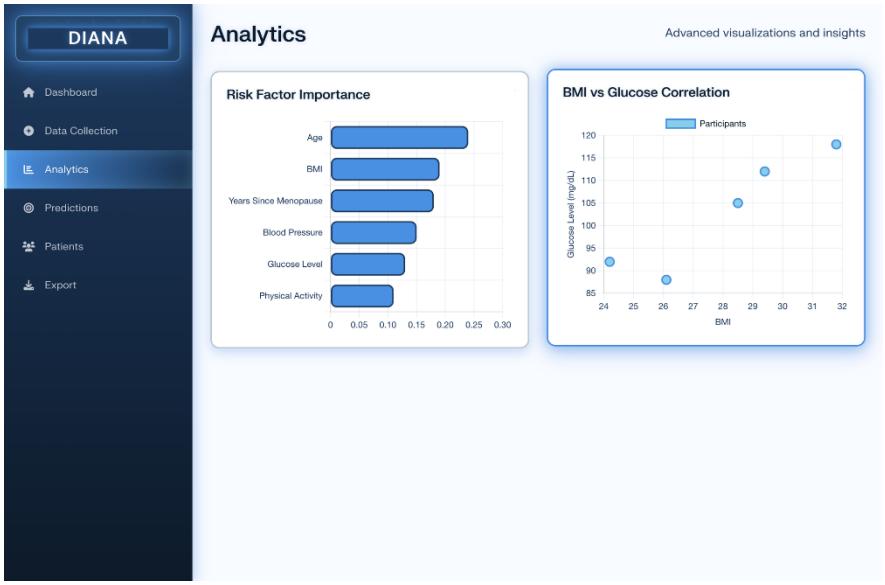







DIANA
Dashboard
Data Collection
Analytics
Predictions
Patients
Export
Prediction Results
Individual risk assessments
Select Participant
Ana Rodriguez (Age 48, Peri-menopause)
Risk Score
34%
Low Risk
Risk Factors Contribution
Age
BMI
Years Since Menopause
Blood Pressure
Glucose Level
Physical Activity
0
0.05
0.10
0.15
0.20
0.25
Recommendations
Continue current healthy lifestyle practices


DIANA
Dashboard
Data Collection
Analytics
Predictions
Patients
Export
Export Data
Download datasets and reports
Export Participant Data
Download complete participant dataset
Export as CSV
Export as Excel
Export Analytics Report
Download comprehensive analytics report
Generate Report
Filtered Export
Export data based on selected criteria
All Menopause Stages
All Risk Levels
Export Filtered Data


DIANA
Dashboard
Data Collection
Analytics
Predictions
Patients
Export
Analytics
Advanced visualizations and insights
Risk Factor Importance
Age
BMI
Years Since Menopause
Blood Pressure
Glucose Level
Physical Activity
0
0.05
0.10
0.15
0.20
0.25
0.30
BMI vs Glucose Correlation
Participants
Glucose Level (mg/dL)
120
115
110
105
100
95
90
85
24
25
26
27
28
29
30
31
32
BMI

Appendix B

Letter of Requests

Request to Conduct Interview for Dr. Pajanel

MAPÚA MALAYAN COLLEGES LAGUNA

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPÚA MALAYAN COLLEGES LAGUNA

October 27, 2025

Dr. Margarette Rose Pajanel
Internal Medicine – Endocrinology, Diabetes and Metabolism
The Medical City – South Luzon
Greenfield City, Brgy. Don Jose, Santa Rosa City

Greetings of peace!

We are 4th-year Computer Science students from Mapúa Malayan Colleges Laguna, currently taking up our Thesis course. As part of our academic requirements, we are conducting a study entitled:

"DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes."

In line with this, we respectfully seek your permission to conduct an interview with you regarding our study. The purpose of this interview is to gain professional insights and medical perspectives that will strengthen the accuracy and applicability of our research. Your expertise as a healthcare provider will be invaluable in validating our approach and ensuring that our work aligns with clinical practices.

Additionally, we would like to request access to de-identified patient data related to Type 2 Diabetes diagnostics and biomarker test results, limited only to clinical values (e.g., glucose levels, HbA1c, lipid profiles, relevant biomarkers) without including any personal identifiers such as names, addresses, or contact information. These data will be used strictly for academic purposes and handled with the highest confidentiality in compliance with ethical research standards.

We would be very grateful if you could let us know your availability at a time most convenient to you for the interview. Rest assured that all information you share, as well as the data provided, will remain confidential and used solely for academic purposes.

Thank you very much for your time and kind consideration. We truly appreciate your support in helping us accomplish this research.

Sincerely,

Marc Kenneth M. Angeles
Adrian Gabriell P. Francisco
Sophia Nicole D. Grefaldo
Neoron G. Lopez

Noted by:

Aurelia Sharlene O. Delos Santos
Thesis Adviser

Dr. Margarette Rose Pajanel

Address : Pulo Diezmo Road, Cabuyao City, Laguna 4025
Trunkline: +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 8520-8975
Email : mclinfo@mcl.edu.ph

Request to Conduct Interview for Dr. Dy

MAPÚA MALAYAN COLLEGES LAGUNA

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPÚA MALAYAN COLLEGES LAGUNA

October 22, 2025

Dr. Violeta Felipe-Dy
Obstetrics and Gynecology
FelipeDy ObGyn Clinic & OBGyn Ultrasound
P. Burgos St., Zone 3 Brgy. Sto Domingo, Binãn, Philippines

Greetings of peace!

We are 4th-year Computer Science students from Mapúa Malayan Colleges Laguna, currently taking up our Thesis course. As part of our academic requirements, we are conducting a study entitled:

"DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes."

In line with this, we respectfully seek your permission to conduct an interview with you regarding our study. The purpose of this interview is to gain professional insights and medical perspectives that will strengthen the accuracy and applicability of our research. Your expertise as a healthcare provider will be invaluable in validating our approach and ensuring that our work aligns with clinical practices.

Additionally, we would like to request access to de-identified patient data related to Type 2 Diabetes diagnostics and biomarker test results, limited only to clinical values (e.g., glucose levels, HbA1c, lipid profiles, relevant biomarkers) without including any personal identifiers such as names, addresses, or contact information. These data will be used strictly for academic purposes and handled with the highest confidentiality in compliance with ethical research standards.

We would be very grateful if you could let us know your availability at a time most convenient to you for the interview. Rest assured that all information you share, as well as the data provided, will remain confidential and used solely for academic purposes.

Thank you very much for your time and kind consideration. We truly appreciate your support in helping us accomplish this research.

Sincerely,

Marc Kennel M. Angeles
Adrian Gabriell P. Francisco
Sophia Nicole D. Grefaldo
Neoron G. Lopez

Noted by:

Aurelia Sharlene O. Delos Santos
Thesis Adviser

Dr. Violeta Felipe-Dy

Address : Pulo Diezmo Road, Cabuyao City, Laguna 4025
Trunkline: +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 8520-8975

mcl.edu.ph MapuaMCL mapuamcl

Request to Conduct Interview for Dr. Bilon

MAPÚA MALAYAN COLLEGES LAGUNA

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPÚA MALAYAN COLLEGES LAGUNA

October 22, 2025

Dr. Violeta Felipe-Dy
Obstetrics and Gynecology
FelipeDy ObGyn Clinic & OBGyn Ultrasound
P. Burgos St., Zone 3 Brgy. Sto Domingo, Binãn, Philippines

Greetings of peace!

We are 4th-year Computer Science students from Mapúa Malayan Colleges Laguna, currently taking up our Thesis course. As part of our academic requirements, we are conducting a study entitled:

"DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes."

In line with this, we respectfully seek your permission to conduct an interview with you regarding our study. The purpose of this interview is to gain professional insights and medical perspectives that will strengthen the accuracy and applicability of our research. Your expertise as a healthcare provider will be invaluable in validating our approach and ensuring that our work aligns with clinical practices.

Additionally, we would like to request access to de-identified patient data related to Type 2 Diabetes diagnostics and biomarker test results, limited only to clinical values (e.g., glucose levels, HbA1c, lipid profiles, relevant biomarkers) without including any personal identifiers such as names, addresses, or contact information. These data will be used strictly for academic purposes and handled with the highest confidentiality in compliance with ethical research standards.

We would be very grateful if you could let us know your availability at a time most convenient to you for the interview. Rest assured that all information you share, as well as the data provided, will remain confidential and used solely for academic purposes.

Thank you very much for your time and kind consideration. We truly appreciate your support in helping us accomplish this research.

Sincerely,

Marc Kennel M. Angeles
Adrian Gabriell P. Francisco
Sophia Nicole D. Grefaldo
Neoron G. Lopez

Noted by:

Aurelia Sharlene O. Delos Santos
Thesis Adviser

Dr. Violeta Felipe-Dy

Address : Pulo Diezmo Road, Cabuyao City, Laguna 4025
Trunkline: +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 8520-8975

Research Locale Request Letter for New Sanai MDI Hospital

MAPÚA MALAYAN COLLEGES LAGUNA

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPUA MALAYAN COLLEGES LAGUNA

New Sinai MDI Hospital

National Highway, City of Santa Rosa, Laguna

Subject: Request for Permission to Conduct Research and Collaborate as Local Research Site

Dear, Dr. Emmanuel Padilla and Dra. Marian Colasito

Good day!

We are Computer Science student, a undergraduate researchers from Mapua Malayan Colleges Laguna, currently conducting our thesis title" DIANA: A Predictive Model-based Application Using Selected Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes".

Our study aims to develop and evaluate a predictive model integrated into a digital application that utilizes selected blood biomarkers to identify menopausal women who may be at risk of developing type 2 diabetes, Specifically, the study seeks to identify relevant biomarkers such as Glycated hemoglobin, fasting plasma glucose, estradiol levels, among others, which are commonly assessed in routine laboratory tests.

In line with this, we would like to inquire if we may conduct our data collection within your institution and request your approval to collaborate with your medical professionals and laboratory staff as our local research site and validator. We also seek permission to interview selected medical practitioners who can provide professional insights related to diabetes risk assessment and biomarker evaluation. Additionally, we would like to request access to anonymized laboratory data (if permissible and in accordance with your data privacy and ethical guidelines) that may serve as parameters in developing our predictive model.

We assure you that all information gathered will be treated with the highest level of confidentiality and used solely for academic and research purposes. A formal research ethics clearance from our institution will also be presented upon approval of this request.

We would be very grateful for the opportunity to collaborate with your institution as a valuable partner in this study. Should you require any further information or documents, we will be happy to provide them at your convenience.

Thank you very much for your time and consideration. We look forward to your favorable response.

Sincerely,

Marc Kenriel M. Angeles

Sophia Nicole D. Grefaldo

Adrian Gabriell P. Francisco

Neoron G. Lopez

Noted by:

Aurelia Sharlene O. Delos Santos
Thesis Adviser

Address : Pulo Diezmo Road, Cabuyao City, Laguna 4025
Trunkline: +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 8520-8975

Research Locale Request Letter for The Medical City South Luzon

MAPÚA MALAYAN COLLEGES LAGUNA

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPUA MALAYAN COLLEGES LAGUNA

The Medical City South Luzon

L United Blvd, City of Santa Rosa, Laguna

Subject: Request for Permission to Conduct Research and Collaborate as Local Research Site

Dear, Robie Onciano

Good day!

We are Computer Science student, a undergraduate researchers from Mapua Malayan Colleges Laguna, currently conducting our thesis title" DIANA: A Predictive Model-based Application Using Selected Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes".

Our study aims to develop and evaluate a predictive model integrated into a digital application that utilizes selected blood biomarkers to identify menopausal women who may be at risk of developing type 2 diabetes, Specifically, the study seeks to identify relevant biomarkers such as Glycated hemoglobin, fasting plasma glucose, estradiol levels, among others, which are commonly assessed in routine laboratory tests.

In line with this, we would like to inquire if we may conduct our data collection within your institution and request your approval to collaborate with your medical professionals and laboratory staff as our local research site and validator. We also seek permission to interview selected medical practitioners who can provide professional insights related to diabetes risk assessment and biomarker evaluation. Additionally, we would like to request access to anonymized laboratory data (if permissible and in accordance with your data privacy and ethical guidelines) that may serve as parameters in developing our predictive model.

We assure you that all information gathered will be treated with the highest level of confidentiality and used solely for academic and research purposes. A formal research ethics clearance from our institution will also be presented upon approval of this request.

We would be very grateful for the opportunity to collaborate with your institution as a valuable partner in this study. Should you require any further information or documents, we will be happy to provide them at your convenience.

Thank you very much for your time and consideration. We look forward to your favorable response.

Sincerely,

Marc Kenley M. Angeles

Adrian Gabriell P. Francisco

Sophia Nicole D. Grefaldo

Neoron G. Lopez

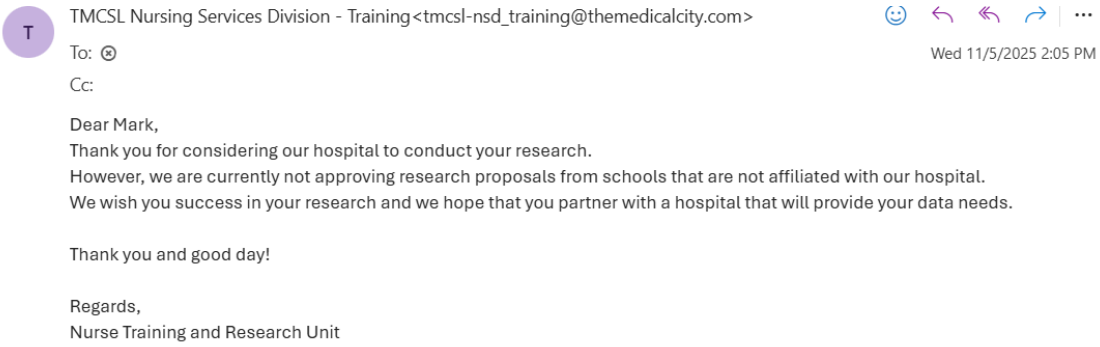
Noted by:

Aurelia Sharlene O. Delos Santos
Thesis Adviser

Address : Pulo Diezmo Road, Cabuyao City, Laguna 4025
Trunkline: +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 8520-8975

mcl.edu.ph MapuaMCL mapuamcl

Response of the Research Locale



T

TMCSL Nursing Services Division - Training<tmcsl-nsd_training@themedicalcity.com>

To:

Wed 11/5/2025 2:05 PM

Cc:

Dear Mark,
Thank you for considering our hospital to conduct your research.
However, we are currently not approving research proposals from schools that are not affiliated with our hospital.
We wish you success in your research and we hope that you partner with a hospital that will provide your data needs.

Thank you and good day!

Regards,
Nurse Training and Research Unit

Research Locale Request Letter for the Healthway QualiMed Hospital Sta. Rosa

MAPÚA MALAYAN COLLEGES LAGUNA

CCIS COLLEGE OF COMPUTER AND INFORMATION SCIENCE
MAPÚA MALAYAN COLLEGES LAGUNA

Healthway QualiMed Hospital Sta. Rosa

Nuvali North, West Nature Ave, City of Santa Rosa, Laguna

Subject: Request for Permission to Conduct Research and Collaborate as Local Research Site

Good day!

We are Computer Science student, a undergraduate researchers from Mapua Malayan Colleges Laguna, currently conducting our thesis title" DIANA: A Predictive Model-based Application Using Selected Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes".

Our study aims to develop and evaluate a predictive model integrated into a digital application that utilizes selected blood biomarkers to identify menopausal women who may be at risk of developing type 2 diabetes, Specifically, the study seeks to identify relevant biomarkers such as Glycated hemoglobin, fasting plasma glucose, estradiol levels, among others, which are commonly assessed in routine laboratory tests.

In line with this, we would like to inquire if we may conduct our data collection within your institution and request your approval to collaborate with your medical professionals and laboratory staff as our local research site and validator. We also seek permission to interview selected medical practitioners who can provide professional insights related to diabetes risk assessment and biomarker evaluation. Additionally, we would like to request access to anonymized laboratory data (if permissible and in accordance with your data privacy and ethical guidelines) that may serve as parameters in developing our predictive model.

We assure you that all information gathered will be treated with the highest level of confidentiality and used solely for academic and research purposes. A formal research ethics clearance from our institution will also be presented upon approval of this request.

We would be very grateful for the opportunity to collaborate with your institution as a valuable partner in this study. Should you require any further information or documents, we will be happy to provide them at your convenience.

Thank you very much for your time and consideration. We look forward to your favorable response.

Sincerely,

Marc Kenneth M. Angeles

Sophia Nicole D. Grefaldo

Adrian Gabriell P. Francisco

Neoron G. Lopez

Noted by:

Aurelia Sharlene O. Delos Santos
Thesis Adviser

Address : Pulo Diezmo Road, Cabuyao City, Laguna 4025
Trunkline: +63 (49) 832-4000
Fax : +63 (49) 832-0017, +63 (2) 8520-8975

Appendix C

Example Test Results

Test Result Hemoglobin A1C (HBA1C) in Laguna Doctors Hospital – Santa Cruz, Laguna

AGE: 46
SEX: FEMALE

PHYSICIAN:
WARD: OPD

GLYCOSYLATED HEMOGLOBIN

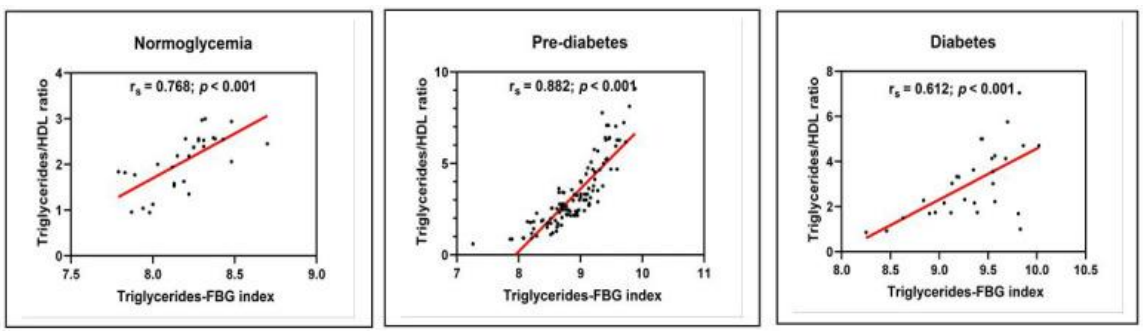
		Reference Range	*Interpretation*
RESULT:	**6.5%**	**4.0-5.7**	**NON-DIABETIC**
		5.8-6.4	**PRE-DIABETIC**
		> 6.5	**DIABETIC**
Method: Boronate Fluorescence Quenching Technology			
Remarks: Please correlate clinically.			

Example Test Results of Chem 12 in Laguna Doctors Hospitals – Santa Cruz, Laguna

SEX M	AGE 64	ROOM OPD	COMMENTS
REPORT TIME	09:53:49		ANALYZER VITROS 250
REPORT DATE	FEB 21 2022		

TEST NAME	RESULT	UNITS	REFERENCE VALUES
GLUCOSE	164. HIGH	mg/dL	76.-101.
BLOOD UREA NITROGEN	15.	mg/dL	9.-20.
CREATININE	1.	mg/dL	0.66-1.2
URIC ACID	7.1	mg/dL	3.5-8.5
CHOLESTEROL	159.	mg/dL	139.-201.
TRIGLYCERIDES	149.	mg/dL	0.-150.
DIRECT HDL	42.	mg/dL	40.-60.
LDL	88.	mg/dL	0.-130.
SGPT (ALT)	33.	U/L	21.-72.
ASPARTATE AMINOTRANSFERASE	28.	U/L	17.-59.

Correlation of Lipid Profile Parameters with Pre-Diabetes (Jasim, et. al., 2025)



Appendix D

Interview Pictures

Interview with Dr. Bilon – Internal Medicine, Diabetes Educator

Interview with Dr. Felipe-Dy – Obstetrics and Gynecologist (OBGYN)

Interview with Dr. Pajanel – Internal Medicine, Endocrinologist

Appendix E

Transcript of Interviews

Transcript of Interview with Dr. Bilon

Interviewer: Marc Kennel Angeles, Sophia Nicole Grefaldo, and Neoron Lopez

Interviewee: Dr. Adonis Bilon

Date of Interview: October 17, 2025

Location: Sta. Cruz, Laguna Polymedic Hospital Inc.

Marc: Good day, Dr. Bilon. We are 4th year Computer Science students from Mapua Malayan Colleges Laguna, currently conducting our thesis entitled:

DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes.

The purpose of this interview is to gain professional and medical knowledge from experts like you to support our study, particularly the relationship between menopause and Type 2 Diabetes, as well as your perspective on the possible use of predictive tools in healthcare.

Before we begin, may we have your permission to record this interview for documentation purposes only. The Information you will provide will be used strictly for academic purposes and kept confidential.

Dr. Bilon: Yes, once na sinabing type 2 diabetes, type 2 diabetes ka na. Hindi yan type 1 diabetes maging type 2 or type 2 maging type 1.

Marc: assisting tool po siya to help patient understand paano po nangyayari ang risk assessment sa kanila bakit sila nagiging type 2 diabetes.

Dr. Bilon: sige, anong parameters ba ang gusto ninyo malaman?

Marc: Hba1c po

Dr. Bilon: and Fasting Blood Sugar?

Marc: ang mga gusto po namin itanong is yung Fasting Plasma Glucose and Tryglicerides. Paano po namin siya mabibigyan ng correlation in stages paano po nagdedevelop and diabetes sa kanila

Dr. Bilon: Actually magkaiba yon, ang diabetes is pag increase ng sugar mo so yun ang diabetes. Kumaba para masabi kang diabetes may certain level yon. Hindi porket high and nakalagay sa laboratory result mo diabetic ka na. May requirements yon. So sa diabetes is kailangan ang sugar mo is 126 and above sa fasting blood sugar, sa random blood sugar ay 210 above tapos aside from that is yung pangatlong requirement ko dyan is dapat may sign and symptoms ka ng diabetes Polydipsia, polyuria, and polyphagia, weight loss. Pero kung ikaw ay nasa 126 wala ka pang symptoms, uulitin ka ulit.

Marc: Paano po yung ganoon doc? 126 and still not diabetic po and high, ano po ang ginagawa ninyo as a doctor?

Dr. Bilon: Diet and exercise tapos ulitin mo after 2 weeks yes parang lifestyle. Lifestyle modification yon.

Marc: Regarding, How do you currently decide which meonoapusal patient needs diabetes screening and can you walk through on your decision making? Paano po sila nabibigyan ng screening sa mga ganon po?

Dr. Bilon: Menopausal? Post menopausal?

Marc: Menopausal po

Dr. Bilon: kapag kasi menopausal, nag memenopause na. Hindi lang kasi menopausal women ang nagkaka diabetes. Pero ang target nyo ba is menopausal women ba?

Marc: Yes po menopausal women po

Dr. Bilon: wala naman kasi pinagkakaiba sa lalaki sa babae, sa bata pa, or sa may edad na ang pagkaroon ng diabetes, pag detect. Pare pareho lang sila. Tulad ng fasting blood sugar 126 and above random blood sugar 210 and above tsaka yung signs and symptoms. Sa lahat iyon pa rin and standard. Kumbaga walang direct menopausal siya, peri menopausal siya or post menopausal siya. Or kahit lalaki same pa rin.

Marc: sa fasting blood sugar, gaano po katagal ang fasting?

Dr. Bilon: sa diabetic 8 to 10 hours iyon. Yung iba hanggang 12. 8 to 12 hours

Marc: ano po yun nakadipende po sa age po nila? Or still the same po?

Dr. Bilon: same lang

Marc: Ano po, what blood test do you routinely order po in assessing diabetes risk po?

Dr. Bilon: Usually lipid profile, creatinine, uric acid, yung sa liver sgot/sgpt sa liver. Kasi doon sa creatinine laging may risk and diabetic na pwede magkaroon ng renal problem yun ang most common niya na side effect ng diabetic. kaya mas marami ang nag dialysis na diabetic. Before marami yun ang sakit. Basta mga infection sa kidneys, sa urine or nephrotic syndrome. Lately 2000s marami na ang dialysis na diabetes. Complication ng diabetes. Usually, kasama ang hypertension sa diabetic, usually lang hindi lahat. Kapag sinabi mong diabetic tinatawag na cardiac

patient na yan. Common ang side effect ng cardiac problem sa diabetic. Usually yan ang inaatake sa puso dahil mataas ang cholestorol niya.

Marc: At what stage are most of your menopausal patients diagnosed with Type 2 Diabetes during prediabetes, at early onset, or after complications have developed? What prevents earlier detection in your experience?

Dr. Bilon: kapag diabetic na, diagnose na siya ng diabetes. Advise ko lang is regular checkup, sugar nila, tsaka regular exercise kasama ang diet. Kaya lang kadalasan failed lagi sa diet and exercise. Ang management mo is gamot pa rin pero individualize pa rin ang pasyente. Iassess mo pa rin ang pasyente. Hindi lahat pare parehas yung iba nangangailangan ng insulin, yung iba oral agent. Pero kapag prediabetic wala kang ibang gagawin dyan kundi monitor lang sila, advise mo lang sila every 3 months 6 months magpacheckup ng sugar, diet and exercise and avoid ang sigarilyo and alak.

Marc: What accuracy level would you need to trust a tool like DIANA 80%, 90%, 95%? What features would make you actually use it in practice risk visualization, trend tracking, patient reports? What would make you NOT use it?

Dr. Bilon: actually first time ko marinig ang diana. Yan ay study nyo diba? First time ko marinig kaya di ko pa alam ano ang outcome niyo. Medyo excited din ako kung ano yan para makatulong itong interview natin na madevelop niyo. Looking forward ako sa gawa niyo. Ang tanong mo is kung ano ang prevention?

Marc: accuracy level to predict to help the doctors in actual using of the web application po, kung ano po ang gusto nyong nakikita feature sa web application like risk evaluation, risk dashboard, risk visualization. Para pag may pinapakita po tayo sa patient, naiitindihan po nila, tracking po.

Dr. Bilon: kumabaga may graph yon?

Marc: Yes po nakikita nila yung changes and mabibigay niyo po yung evaluation niyo na ganto ang nangyayari, within years paano naging ganto, bakit di nagbabago yung result mo.

Dr. Bilon: tinatanong mo ako kung ano ang magiging itsura? Kung para sa akin maganda yung may graph from, yung patient is yung hemoglobin a1c is 1st year andito, then after 10 years or 2 years andito na tumataas na. Makikita ng patient ang graph. Tumataas habang tumatagal ibig sabihin hindi mo naccontrol. Atleast nakikita ng pasyente "oo nga no hindi ko nagagawa" kaya mas maganda naka graph.

Marc: pati po yung mga patient report po? Mas mabibilis po mattrack.

Dr. Bilon: yes, mas better mas lenient para sa akin kasi input ko lang andon na siya. Input ko lang makiktia na yung graph.

Marc: ano naman po features na ayaw niyo po makita?

Dr.Bilon: mas maganda pag bukas mo andoon na yung graph, naka graph na siya yung mga parameters niya, yung hemoglobin a1c niya, cholestorol niya, yung creatinine niya. Yung one picture andoon na siya.

Marc: ang sa amin po kasi is pagkaopen niyo po is mag login lang po kung sino yung doctor, makikita niyo po yung dashboard ilan po yung patient niya, ilan yung namomonitor, patient history, track record po, kapag clinick mo po makikita mo doon yung record niya. Magkakainput lang po is if mag add po kayo ng another patient po.

Dr. Bilon: yes maganda yon.

Marc: What advice would you give to Computer Science students like us to ensure this predictive model is clinically useful and not just technically impressive? What mistakes should we avoid?

Dr. Bilon: mahirap makita ang mistake dahil wala ka pa na model, as long as andoon lahat ng information dapat mayroon clinical analysis example is ganto yung kidney niyo within years. Pero yung mali hindi ko pa nakikita iyon. Pero kung ang plano niyo is kunyare 2020 ito ang graph niyo, ilang taon na lang kaya ang kidney niyo by 2030 ganto na lang mag function. Baka kailanganin mo na ng diabetes that time. Mayroon na Framingham Study na kapag mataaas ang cholesterol by that time na ganto na ang value mo may chance na magkakaroon ka na ng atake. Kapag high risk ka in the future ilang percent na ang chance mo na aatakihin ka sa puso andoon yon sa Framingham Study iyon. Maganda sa study niyo na may ganoon. Ilan ang magiging survial rate mo, kung kaya niyo mapredict iyon para maanalyze.

Marc: tingin niyo po ba if magagawa po siya na correct and tama po magiging useful po ba siya sa mga doctors and patient po in the long run po?

Dr. Bilon: sa tingin ko oo, same with the Framingham Study, useful siya. Mapapakita mo sa patient mo na ang chance mo ang taas ng cholesterol mo. Ganto ang chances mo na aatakihin ka sa puso. Naanalyze to. Ang chances nito is ganito, makikita mo to sa Framingham.

Neoron: additional question po, gaano po kaimportante ang insulin sa pagdetect po ng diabetes po?

Dr. Bilon: insulin? Hindi siya nagddetect, gamot siya ang insulin hindi parameter. Kapag ikaw ay mataas ang sugar mo. May mga insulin na tinuturok pero may mga tao naman na hindi naman kailangan to ang ginagamit nila is gamot.

Neoron: may chances po ba na ang insulin is nagccause ng diabetes po?

Dr. Bilon: No, siya ang gamot sa diabetes. Actually ang katawan natin ang pancreas iyon ang nagpproduce ng insulin. Kunyare kumain tayo ng matamis, magsisignal yung sa katawan natin na may kinain tayong matamis. Magsisignal yon sa pancreas na magproduce ng insulin. Kapag may sugar ka na maactivate ang beta cell mo, magpproduce ito ng insulin, baba ang sugar kapag maayos ang beta cells mo.

Marc: tanong ko lang din po, paano naman po nadedevelop yung insulin resistance?

Dr. Bilon: actually mga medyo matataba pero di naman lahat e, yung insulin resistance na yon. Tinamad na yung insulin, di na nakakapagpababa na nag sugar. Kumabaga natulog lang kaya tinatawag na insulin resistance, kasi di siya bumababa sa insulin na naproduce niya. Common yun sa matataba. Kaya kung minsan kailangan mo lang ng exercise para maging active pero konti lang yon, yung ibang matataba mag insulin man sila dahil insulin resistance di mo mapababa yung sugar kaya kailangan mo mag exercise or metformin para sa insulin resistance.

Sophia: Additional question lang po, kapag menopausal po ba wala po siya connection sa diabetes?

Dr. Bilon: ang diabetes ay di lang yan common sa menopausal, bakit marami din na lalaking diabetic?

Sophia: may nabasa po kami na study which is yung isa sa symptoms ng hormonal changes po is doon nagkakaroon ng insulin resistance:

Dr. Bilon: sa insulin resistance, ay pwede. Pero siyempre ang ano pa din doon ay pagtaba diba kapag bata bata ka pa ay usually active ka, kapag may edad ka na medyo mabagal na ang kilos so mahina na ang metabolism mo kaya tumataba ka ngayon. Kaya very common kapag tumataba ang isang babae kapag menopause na. Pero walang kinalaman ang diabetes sa menopausal maaring sa

insulin resistance pero meron naman nasa younger stage na may insulin resistant na sila kasi mataba sila. Usually obese ang insulin resistance. Pero kung mga payat may mga insulin resistance pero di ganoon sa mga obese. Tsaka suggest ko sa inyo, ilan ba ang iinterviewhin niyo?

Marc: currently tatlo po

Dr. Bilon: siyempre mas maganda kung madami, dahil maaring di ko makita ang nakita niyo or makita nila ang di ko nakita.

Marc: Thank you very much for your time and valuable insights, Dr. Bilon. Your expertise will greatly help us ensure that DIANA addresses real clinical needs and is designed with practical healthcare applications in mind.

If we have any follow-up questions during our research, would it be possible to reach out to you again?

Once again, all information shared today will remain confidential and will be used solely for academic purposes. Thank you.

Dr. Bilon: Looking forward.

Transcript of Interview with Dr. Dy

Interviewer: Marc Kennel Angeles, Adrian Gabriell Francisco, Sophia Nicole Grefaldo, and Neoron Lopez

Interviewee: Dr. Violeta Felipe-Dy

Date of Interview: October 22, 2025

Location: Felipe-Dy ObGyn Clinic & OBGyn Ultrasound, Binãn Laguna.

Adrian: Good day, Dr. Dy, we are 4th-year Computer Science students from Mapúa Malayan Colleges Laguna, currently conducting our thesis entitled: DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes.

The purpose of this interview is to gain professional insights from an OBGYN perspective regarding the clinical relationship between menopause and Type 2 Diabetes risk in your practice. This interview should take approximately 15 minutes.

Before we begin, may we have your permission to record this interview for documentation purposes only? The information you provide will be used strictly for academic purposes and kept confidential.

Dr. Dy: Type 2 Diabetes in? Menopausal? When you say menopausal, ano ang inyong age na papasok sa inyong study?

Marc: We are following the standard Word Health Organization age gap po 45

Sophia: 45 to 60

Dr. Dy: so, yun ang inyong inclusion?

Marc, Sophia, Neoron: Yes po

Dr. Dy: okay, ano ang gusto ninyo malaman?

Adrian: Could you please tell us about your specialization and how many years you have been in practice?

Dr. Dy: I'm an Obstetrics and Gynecologist, 22 years in practice. I graduated my Residency training in Fabella Memorial Hospital.

Adrian: Approximately what percentage of your patient population consists of perimenopausal and postmenopausal women?

Dr. Dy: Ilang percent ang patient ko ang?

Sophia: Kahit yung estimated nyo lang po

Adrian: Perimenopausal and postmenopausal women po

Dr. Dy: Kapag Perimenopausal those are the ages before menopause, it depends. Mayroon kasing as early as 42, 43 kapag may symptoms na pero nag mmenstruate pa perimenopausal na. Nag start na yan ng ages 45. Ilang percent ang nakikita ko? Siguro mga 20 to 25 percent.

Adrian: In your practice, have you noticed any relationship between menopause and Type 2 Diabetes risk?

Dr. Dy: I don't really, ano. Hindi ko sila ni-nonote, your question is again?

Adrian: Have you noticed any relationship between menopause and Type 2 Diabetes Risk po?

Dr. Dy: Menopause and Diabetes Risk. Actually, Diabetes is nasa lahi, nasa genes. Kapag diabetic ka minsan as early as reproductive age, nag iincrease na talaga ang sugar mo dipende sa lifestyle and diet. But the correlation kung may relation ba ang age sa diabetes? Definitely yes. Kasi as we age doon lumalabas ang sakit. Kung may family history the parents, the lolo's and the aunties are diabetic, the chance of you being diabetic also at the ages of 40s and above more likely lalabas. Nasagot ko ba is there are correlation between perimenopausal and Diabetes? Yes

Adrian: Yes po, During the menopausal transition, what metabolic or physical changes do you routinely observe in your patients?

Dr. Dy: Weight Gain, definite yan, ang 40s hirap na mag diet yan, hirap ng mag exercise yan, pag nag exercise na yan para bang isang- tatlong araw na kain feeling nila may bloatedness na uli mataba na sila uli. Diba? Problem ng mga parents nyo yan, mga mommy's nyo yan.

Adrian: In your experience, do women who undergo surgical menopause show different metabolic patterns or diabetes risk compared to women experiencing natural menopause?

Dr. Dy: It's the same, as long as you are menopause or yung inyong ovaries ay natanggal na kasi yan ang magccause ng menopause. Alam nyo ba kung bakit nagmmenopause? Nag mmenopause kasi yung egg cell namin mga babae we are born with aroung 3500 egg cells, when we reach puberty isa isa yang narrelease bago mag menstruate, nag oovulate kami, kapag di na fertilize ng sperm hindi mabubuntis, mag rrelease ulit kami monthly hanggang maubos siya. Kapag naubos na ang egg cell, wala ng production ang ovaries mag menopause na kami. Wala na kami menstruation. Pareho lang. Kapag surgical menopause wala ka ng ovaries, natural menopause ubos na ang reserve mo sa ovaries mo, so pareho na silang menopause di na sila nag mmenstruate kulang na ang hormones sa katawan. Kaya pareho lang.

Adrian: As an OBGYN managing menopausal women, do you routinely screen for diabetes or metabolic syndrome as part of your gynecological care? If yes, what blood tests do you typically order?

Dr. Dy: Yes, we usually ano, I usually request for Hemoglobin A1C and the magic 12 or the chem 12 or chem 15 it includes fasting blood sugar, uric acid, cholesterol, triglycerides, SGPT/SGOT and sometimes the thyroid hormones just to complete the chemical feature ng patient.

Adrian: At what point do you typically refer a menopausal patient to an endocrinologist or internal medicine specialist for diabetes evaluation?

Dr. Dy: if there is already an increase in the value of their Hemoglobin A1C and their Fasting Blood Sugar, also kapag ka mayroon ng symptoms ng frequent candidacy or yeast infection sa urine and sa discharge. Yan dalawang yan, kapag ka doctora palagi akong may discharge kulay puti. How is your sugar? Dahil common sa diabetic patient ang mayroon vaginal discharge.

Adrian: Is the relationship between menopause and diabetes risk something you routinely discuss with your patients during menopausal counseling?

Dr. Dy: Not necessarily, kasi usually patients come na "doctora I'm already on maintenance, I'm already on oral hypoglycemic or insulin. Is more of internist patients kasi ang diabetes but the correlation between the OBGYN, menopause and diabetes parang not necessarily na menopause causes diabetes. Meron ng diabetes kahit menopause, perimenopausal o hindi. Parang different entity sila.

Adrian: What information or insights would help you identify which menopausal patients should be prioritized for earlier diabetes screening or intervention?

Dr. Dy: Actually lahat, all women age 40 above should be screen for diabetes. With the lifestyle and diet that we have lahat prone. Kahit walang family history, bigla kapag mataas ang sugar why? Because of the diet, lifestyle.

Adrian: Do you currently use any risk assessment tools, scoring systems, or guidelines to identify which menopausal patients are at higher risk for developing Type 2 Diabetes?

Dr. Dy: I'm not using any, siguro yung iba. When we attend some convention our diebetologist friends actually request us to screen our patients but since small lang ang patients ko na nakikita na menopausal, perimenopausal, hindi masyado. Kasi we have seeing so many patients it is time consuming, yung kanilang page, dalawang page ata o isang page na gaganunin tapos if they cannot understand the secretary will be assisting them so I need to educate pa my secretaries. Parang medyo hindi ko pa masyado napag tuunan ng pansin. Anyway, clinically kasi kapag nedyo oops mababa, 40s, automatic yan. Medyo ano na yan, candidate na yan. And I do screen naman ang ating laboratories nag rerequest na ako ng mga labs.

Adrian: From an OBGYN perspective, what are the biggest challenges you face in identifying which menopausal patients need earlier diabetes screening?

Dr. Dy: Financial, kasi yung pasyente di naman lahat may kaya. Yung area ko wala ako sa alabang, nasa palengke ako. So 50 percent of my patient are indigent kapag ka ganoon nirrefer ko na sa ating local government hospital so financial minsan ang screening. Kasi kapag nag request ka hindi nila nagagawa kasi wala sila pang bayad ng laboratory, kaya nag iintay sila "doctora iintayin ko lang po mayroong free sa center" tapos kahit may free babalik sa akin magbabayad sila parang dinidiresto ko na sila sa diabetologist para makatipid sila sa time and budget.

Marc: Follow up question ko lang po about sa correlational din po, sa type 2 diabetes sabi nyo po is may correlation sila based on age po. Paano po natin masasabi yung development ng diabetes sa isang menopausal is mabilis po o mabagal po?

Dr. Dy: Wala silang mabilis or mabagal, kasi kapag andyan, andyan. Majority of the patient are asymptomatic from the start. You don't know you have it. Polydipsia, polyuria, and polyphagia. Those are the 3 symptoms of diabetes in pregnancy. Pala ihi, pala inom, pala kain. Noong araw, madaling ma identify kasi kapag nag ihi ka lalanggamin ihi mo. May toilet na tayo diba? Madaling sabihin sa mga guys, kasi minsan yung mga lalaki kung saan umihi sa kanto, sa likod ng bahay nilalanggam. Alam mo na, pero sa babae we don't do that anymore. So, in proper education sa children ng mga parents with diabetes. Dahil sila na mismo ang mag iingat sa anak nila. Kami, wala akong lahing diabetes but my husband mayroon. Tinetrain ko na ang mga anak ko na don't masyado sa sweets, kapag nagiincrease ng weight when they reach already the age of 25 tapos parang nag ooverweight na sila, "anak magpa chem 10 na kayo, mag palaboratory na" so babantayan mo na. So proper education na kapag may family history. Kasi alam mo yung family history na napuputulan ng paa, yung mga diabetic foot. Pero hindi ibig sabihin kapag nagkasugat ang diabetic automatic putol. Hindi totoo yon. May pasyenteng diabetis ayaw mag pa cesarian kasi baka di gumaling ang sugat. Mali iyon, gumagaling ang sugat ng mga adult diabetis. Kasi maccontrol na, alam na natin paano ito magagamot. So alin ang napuputol? Ang mga micro circulation, kung saan maraming maliliit na ugat? Sa daliri, marami dyan. Kapag na-infect, unti unti na, namamtay yung ugat. Kapag namatay yung ugat kasama ang nerve, kapag nakasama ang nerve walang sensory. May sugat na di pa nararamdaman. So nabubulok yung parte ng daliri na hindi nararamdaman ng pasyente. Ang nakakaalam ay kamag-anak bakit? Ay mabaho. Kung kaya't minsan ang pasyente ba't mabaho, pagkita sa ano, nabubulok na. Nilalagyan naman ng

bandaid, ng sapatos, hindi nakikita. Kaya kapag nakita, itim na hanggang bandang paa na. So, ganoon ang diabetes. Pero kapag cesarian section, microcirculation yan, malaki ang blood vessel dyan, nakikita kaya nattreat. Ang sa paa hindi.

Sophia: Follow up question ko lang po, may possibility kapag nasa manopausal stage na yung babae is may risk po talaga ng type 2 diabetes? Nagkakaroon na po talaga ng risk?

Dr. Dy: May risk dahil more of ano tayo, genetics pero manopause and diabetes parang incidental finding lang dahil menopause nag weight gain ka, kapag nag weight gain ka iccheck mo na if may diabetes ka.

Marc: Mostly po hormonal changes in the body po?

Dr. Dy: yung hormonal changes metabollic? Pwede rin kaya lang more on genetics, more of genes talaga. Kapag prone ka sa diabetic, kahit payat ka, kahit di ka pa menopause. When you are still in your 30s pwede ka na magkadiabetes.

Sophia: yung hormonal changes po ng manopausal, hindi po ba siya big factor for diabetes po?

Dr. Dy: nagiging factor siya kasi like polysitic ovaries, nagkakaroon ka ng increase weight gain. Pagka increase ka ng weight gain, definitely yung insulin mo nagkakaroon na ng imbalance, nagkakaroon ka na ng diabetes.

Neoron: Additional question ko lang po, tungkol lang po sa insulin, like may experience po kayo in insulin injection po?

Dr. Dy: Ano ang tanong mo?

Neoron: since sa menopausal women po, like kapag ininject nyo po ng insulin po, is there a possibility po na insulin is nagccause ng diabetes?

Dr. Dy: No, insulin is the treatment for diabetes

Marc: To correct his question po, nagkakaroon po ba ng build-up resistance po ba kapag yung insulin is injection po?

Dr. Dy: Hindi, ganito yan, ang diabetes ay sakit ng pancreas, nagkukulang sila ng insulin kaya ang sugar level mo ay tumataas. Kailangan mo ng insulin para ang sugar level mo bababa and insulin is the cure for diabetes. Nagkaroon ng study na first line na ngayon ng diabetes ang insulin para gumaling ang pancreas mo, di ko lang alam which patients belong to that category. Parang mas maaga ka nag insulin, mas nassave ang pancreas mo, parang napoprolong nya yung life ng iyong pancreas. So insulin resistance, may tinatawag tayong insulin resistance so may ibang metabollic problem iyong pasyente. Medyo mahirap maexplain ng insulin resistance kaya biochemical reaction ng katawan na may kulang kang hormone para magkaroon ng increase uptake sa insulin mo matanggap ng katawan mo. May tinatawag tayong insulin resistance e, kung iyon ang tanong mo. Dahil mayroon na hindi tinatablan ng insulin pero hindi naman. Pero may term tayong na insulin resistance patient pero I'm not sure if yun yung iniinject na insulin, hindi yata. Parang ang hirap no? Na gets mo ba? Kasi pysological speaking, ang insulin resistance ay hindi galing sa insulin na gamot, kung hindi doon sa katawan ng tao mismo, hindi yung artificial insulin na iniinject. So yung insulin resistance nasa sistema ng tao kung bakit tumataas ang sugar mo kung baga sa loob ng katawan mo di ka nagpproduce ng insulin yan. Insulin resistant hindi yung insulin na yung gamot na iniinject ay di natalab. Iba yon. Magkaiba yon. Akala niyo yung insulin ay iniinject lang? Hindi, mayroon tayong- katawan natin nagpproduce ng insulin maaring iba ang tawag pero nagpproduce tayo para ma ano natin ang sugar na kinakain ng katawan, maconvert natin yung sugar into energy.

Adrian: Thank you very much for your time and valuable insights, Dr. Dy. Your expertise from the OBGYN perspective will greatly help us ensure that DIANA addresses the real clinical needs of menopausal women in Philippine healthcare.

Transcript of Interview with Dr. Pajanel

Interviewer: Marc Kennel Angeles, Adrian Gabriell Francisco, and Neoron Lopez

Interviewee: Dr. Rose Margarett Pajanel

Date of Interview: October 27, 2025

Location: The Medical City – South Luzon, Sta. Rosa, Laguna

Adrian: Good day, Dr. Pajanel. We are 4th-year Computer Science students from Mapúa Malayan Colleges Laguna, currently conducting our thesis entitled: DIANA: A Predictive Model-Based Application Using Selected Blood Biomarkers for Identifying Menopausal Women at Risk of Type 2 Diabetes.

The purpose of this interview is to gain professional insights from an endocrinology and internal medicine perspective regarding the clinical relationship between menopause, metabolic changes, and Type 2 Diabetes risk in your practice. This interview should take approximately 15-20 minutes.

Before we begin, may we have your permission to record this interview for documentation purposes only? The information you provide will be used strictly for academic purposes and kept confidential.

Dr. Pajanel: Okay

Adrian: Could you please tell us about your specialization as an Internal Medicine-Endocrinologist and how many years you have been in practice?

Dr. Pajanel: Okay, Internal Medicine, propoer practice is 2 years. I started practicing after my boards last year. So 2 years. Internal Medicine is deal with adult diseases, generally for internal organs, hypertension, diabetes, mostly like that. Lifestyle diseases, Dyslipidemia, everything I am related so all your organs we can deal with that. Syempre iba naman sa subspecialty. So, I subspecialized with endocrinology that deals with diabetes, thyroid, obesity, dyslipidemia, bone and calcium disorder, secretory disorders, adrenal disorder, retro for your hormones we work with your retro endo it is a sub specialty of your ob and for gender hormone therapy for those in transition.

Adrian: Okay po, approximately what percentage of your patient population consists of perimenopausal and postmenopausal women?

Dr. Pajanel: Perimenopausal pretty smalle right now, your main concern is your perimenopause or just by the patient population?

Adrian: Population po

Dr. Pajanel: like a lot, but those naman they don't perimenopause issue. Menopause concern or age nila na menopause? Kasi magkaiba yon e, I get so many patient na menopause na ang concern nila is diabetes they don't worry about being menopause. Magkaiba yon, alin doon ang gusto nyo malaman?

Marc: yung menopause po

Neoron: Menopause po

Dr. Pajanel: nasa 80% of my patient population. Madami talaga.

Adrian: In your practice, have you noticed any relationship between menopause and Type 2 Diabetes risk?

Dr. Pajanel: No.

Adrian: As an endocrinologist managing menopausal women at risk for diabetes, what blood tests do you typically order?

Dr. Pajanel: FBS, HBA1C, Type 2 diabetes diba? Lahat ba?

Marc: hindi po, yung standard lang po

Dr. Pajanel: okay sige, CBC, BUN, creatinine, FBS, lipid profile, HBA1C and urinalysis

Adrian: How significant is insulin resistance as a predictor of Type 2 Diabetes progression?

Dr. Pajanel: Very significant. The most significant factor.

Adrian: Do you routinely test for insulin resistance in your at-risk menopausal patients?

Dr. Pajanel: No, HOMA – IR yon, mahal yon. Wala naman dito, pwede naman fasting insulin pero with no added benefits.

Adrian: Considering the Philippine healthcare context, which of these biomarkers are easily accessible and affordable for most patients?

Dr. Pajanel: Which one? Those that I've mentioned? All of them except the HOMA – IR.

Adrian: Are any covered by PhilHealth or available for free at government hospitals?

Dr. Pajanel: Oo, maccover naman sa PhilHealth but PhilHealth has also own algorithm you can look it up. Parang mayroon silang cinocover talaga.

Adrian: When multiple test results are abnormal like high HbA1c –

Dr. Pajanel: But normal FBS? Yun ba yung next question?

Adrian: Yes

Dr. Panajel: yes, because multi therapy na tayo dito e, so usually diabetes kasama na yung medication for cholesterol e. not really, pero it will change what medication will I give. Do you see pattern when several? Not really there's no pattern. So usually there's a specific lipid profiles for diabetics e that mataas ang LDL which your bad cholesterol then low HDL so we can see that lipid profile elevation for diabetics. Progression from insulin resistance actually prediabetic pa lang makikita na agad ang insulin resistance. So if ang gusto nyo malaman insulin resistance have a significant impact on research don't check the type 2 dahil ang type 2 ay may insulin resistance na talaga yan kung yung insulin resistance per se you might want to change your sample to prediabetics so yun lang elevated insulin level is a risk yes. It means mayroon siyang marker for insulin resistance pero HOMA-IR dapat not really fasting insulin but its not easly accessible. Eventually, ang marker pancreatic data cell exertion we can check C-peptide so that's your main insulin excretion you can check C-peptide yun na yung pinaka maganda if you are looking for academic value for it not clinically value.

Adrian: Our tool DIANA is designed to predict diabetes risk in menopausal women using these biomarkers and machine learning. In your opinion, should this type of tool be accessible directly to patients for self-assessment, or should it be restricted to healthcare provider use only?

Dr. Pajanel: pwede naman [for patient self-assessment] they can do it to check themselves pero I guess not really for restriction, but they should still correlate with their healthcare provider kasi its similar to people Googling or Chat GPT their symptoms.

Marc: Explain lang po namin yung gagawin po. Samin po kasi DIANA is a predictive model base application using web browser you can input your patient data for checking and comparison-

Dr. Pajanel: if they are at risk diba? For developing DM?

Marc: yes po mappredict nya po within years po baka ayun na yung chance na maging diabetes na po sila and then within years mabibigyan po natin sila ng assessment based on the blood biomarkers kung doon tumataas po ba and bumababa and because of that mabibigay natin po sa patient o mapapakita po natin patient bakit ganoon kataas yung risk assessment nila based sa blood biomarkers na iniinput po natin and for the past checkup po nila makikita nila doon naka graph po yung per past po nila and with that po coverage of DIANA yun po yung natatanong po namin diyan.

Dr. Pajanel: I think it should be para macheck din nila sarili nila they have their own record but it should be correlated with their healthcare provider to make proper sense of the data.

Adrian: Do you currently use any risk assessment tools or scoring systems to identify which menopausal patients are at higher risk for Type 2 Diabetes?

Dr. Pajanel: Not really for type 2 DM, but there is a grading system for menopause or frax if you are familiar it is a osteoporosis risk yon. You can look at that para at least macheck nyo. Pero ayon wala naman talaga risk assessment or scoring system. Just by being a Filipino you are already at risk. If not well. Wala naman talaga risk assessment tool or scoring system for type 2 DM.

Adrian: From an endocrinologist's perspective, what are the biggest challenges you face in identifying which menopausal patients need earlier diabetes screening and intervention? How do time constraints, financial barriers, or lack of tools affect your ability to provide proactive care?

Dr. Pajanel: Challenges? Wala naman, siguro patient preference patients lang na ayaw magpacheck ng sarili nila. We try to deal with what we can lalo na sa financial side. Well if you notice most of here are HMO so atleast may caution na matetest naman lahat basta macover ng HMO nila.

Adrian: What is the typical socioeconomic profile of your menopausal patients? What percentage can afford comprehensive lab testing without financial hardship?

Dr. Pajanel: typical socoeconomic profile? Middle to upper. What percentage can afford? 100% or they try to make time for it, they really trying to make budget for it. Ginagawan nila ng paraan, they try to find a way to help themselves.

Adrian: What features would make a predictive tool like DIANA most useful in your daily practice? For example: risk categorization (low/moderate/high), trend visualization over time, or automated alerts for high-risk patients? Which would you prioritize?

Dr. Pajanel: features?

Marc: Doon po sa idedevelop po namin na tool ano po yung gusto niyong makita makita sa web application po namin.

Dr. Pajanel: Sample ng interface? You can put risk categorization and it would be nice if okay naman yung may trends ng lab nila over time. Automated alerts? Paano yung? Maalert ba yung patient na high risk sila?

Marc: We're asking po if you think it should be added the automated alert for high risk

Dr. Pajanel: sino yung maalert? Yung patient?

Marc: yung doctor po

Dr. Pajanel: siguro you can put na lang I think for me it won't change the management naman if high risk siya. Gagamutin pa rin naman. Pwede may risk assessment lang pero yung risk marker or alert siguro may tag na lang siguro kung anong risk nila.

Adrian: When you receive referrals of menopausal women from OBGYNs or general practitioners, what clinical indicators or concerns are most commonly cited?

Dr. Pajanel: The menopausal symptom, the fatigue, the heat flushes, the lack of energy yung yung concern nila kung menopause ang concern nila, pero iba yung manopause na diabetic ang concern nila iba yon. In that question, parang yung question is ano yung concern ng manopausal? The fatigue, the energy, the low of libido, lack of motivation, they find hard to lose weight even they do whatever they really workout, they can't keep the weight down.

Adrian: How do you counsel menopausal women who are at high risk for Type 2 Diabetes but have not yet developed the disease?

Dr. Pajanel: Oo definitely we need to do screening same with menopausal screening naman atleast they have to be aware of their current values, their current health status, by screening if there is nothing to be done, I advocate healthy lifestyles.

Adrian: Have you observed any differences in diabetes-related biomarker patterns or metabolic profiles in Filipino menopausal women compared to international reference ranges or Western populations?

Dr. Pajanel: normally, but that doesn't mean translate to the levels. Siguro same lang din naman although cocation counterparts they have bigger fats or mass so but with us we are generally also sensitive to medications

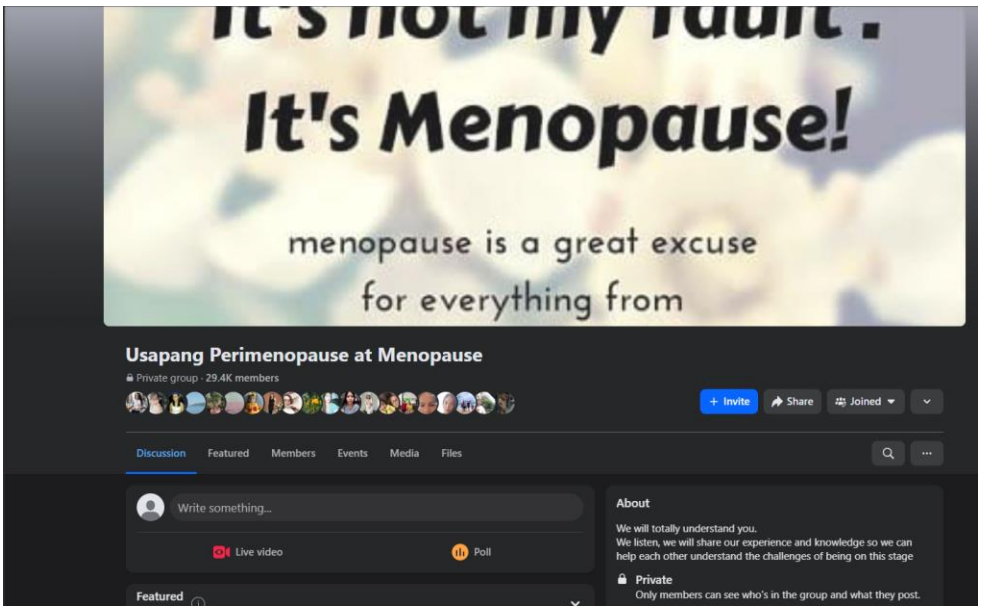
Marc: Additional question lang po, if ever po okay lang po ba na lumapit po sa inyo to test or to verify yung magagawa po namin yung application.

Dr. Pajanel: pwede naman, message niyo lang si ma'am weng para di tayo tumapat sa toxic na araw. Para masched natin yung schedule. Message niyo lang siya.

Adrian: Thank you very much for your time and valuable insights, Dr. Pajanel. Your expertise from the endocrinology and internal medicine perspective will greatly help us ensure that DIANA addresses the real clinical needs of menopausal women in Philippine healthcare and supports early, evidence-based diabetes risk assessment.

Appendix F

Facebook Group: Usapang Perimenopause and Menopause



Approval Message for the admin through Facebook Messenger