DESIGN OF AN EXPLAINABLE MOBILE HEALTH DECISION-SUPPORT SYSTEM FOR DIABETES RISK ASSESSMENT USING A HYBRID DEEP LEARNING APPROACH

BY

NNANNA SAMUEL OGBONNA

REG. NO: GOU/U22/CSC/884

DEPARTMENT OF COMPUTER SCIENCE AND MATHEMATICS, FACULTY OF COMPUTING INFORMATION TECHNOLOGY GODFREY OKOYE UNIVERSITY, ENUGU STATE.

IN PARTIAL FULFILMENT OF THE AWARD OF BACHELOR OF SCIENCE (B.SC), DEGREE IN COMPUTER SCIENCE.

SUPERVISOR: MR. CAMILUS NJOKU

JUNE, 2025

APPROVAL PAGE

This research, titled **"DESIGN OF AN EXPLAINABLE MOBILE HEALTH DECISION-SUPPORT SYSTEM FOR DIABETES RISK ASSESSMENT USING A HYBRID DEEP LEARNING APPROACH,"** has been assessed and approved by the Department of Computer Science and Mathematics Committees in the Faculty of Computing and Information Technology, Godfrey Okoye University, Enugu, Nigeria.

Supervisor ....

Signature Date

External Examiner ....

Signature Date

Dr. Frank Okebanama ....

HOD, Department of Computer Science Signature Date

Prof. I.A Ajah ....

Dean, Faculty of FACIT Signature Date

CERTIFICATION

I certify that I completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

NNANNA SAMUEL OGBONNA

GOU/U22/CSC/884

DEDICATION

This project is dedicated to Almighty God, whose grace and wisdom guided every step of this research, and to my family, whose unwavering love, support, and encouragement throughout my academic journey made this achievement possible.

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to my supervisor, Engr. Camillus Njoku, whose expert guidance, constructive feedback, and constant encouragement were invaluable to the successful completion of this research. I am equally grateful to the Head of Department, Dr. Frank Okebanama, and the entire faculty of the Department of Computer Science and Mathematics, Godfrey Okoye University, for their dedication to academic excellence and their contributions to my intellectual development throughout my undergraduate program.

Special appreciation goes to my parents and family members for their steadfast support, prayers, and motivation that carried me through every challenge encountered during this academic journey. I also thank my colleagues and friends who offered encouragement and assistance in various ways. Above all, I give thanks to Almighty God for the strength, wisdom, and perseverance to complete this work.

ABSTRACT

Diabetes mellitus remains a major global health challenge, with millions of cases undiagnosed, particularly in sub-Saharan Africa. Existing risk prediction systems are largely limited by isolated feature consideration, insufficient integration of genetic and lifestyle data, and a lack of explainability. This study designed and developed an explainable mobile health decision-support system for diabetes risk assessment using a hybrid deep learning approach. The methodology adopted a design science research framework, leveraging the CDC BRFSS Diabetes Health Indicators dataset containing over 70,692 records and 22 features. The proposed hybrid CNN-Transformer architecture was implemented and trained alongside baseline models including Logistic Regression, Random Forest, and XGBoost for comparative evaluation. The hybrid model integrated convolutional layers for local feature extraction with a Multi-Head Attention Transformer block to capture contextual relationships among healthcare indicators, and was optimised using residual connections, layer normalisation, dropout, and early stopping. SHAP explainability techniques were integrated to provide transparent and interpretable prediction outputs. Experimental results demonstrated competitive performance, with XGBoost achieving the highest testing accuracy of 75.52%, while the proposed CNN-Transformer model achieved 74.89% with superior explainability, scalability, and contextual feature learning. The SHAP analysis identified General Health, BMI, High Blood Pressure, Age, and High Cholesterol as the most influential predictors. A mobile-based decision-support interface was also designed and implemented to support real-time diabetes risk assessment. The study demonstrates the feasibility of integrating hybrid deep learning and explainable AI within a mobile health framework for early diabetes detection and preventive healthcare.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

1.1 Background of the Study

Diabetes mellitus (DM), which is a chronic health condition, is on the rise (Kim et al, 2019). There are many issues related to diabetes as a result of its complications, which ultimately lead to a considerable number of fatalities, either direct or indirect. About 8.3% of adults in the world are diabetic (Läll et al, 2017). DM is a common reason for vision loss, amputations, heart disease, and kidney problems. Life expectancy decreases by 5 to 10 years because of type 2 diabetes (T2D), which constitutes 80 to 90% of all diabetes in Europe (Läll et al, 2017).

If there is excessive glucose, often referred to as blood sugar, in the blood, then one has diabetes (Li et al, 2023). There are different forms of diabetes including type 1 diabetes, type 2 diabetes and gestational diabetes. Type 2 diabetes mellitus (T2DM) has been shown to be affected by both lifestyle and genetic factors. The evidence from many genetic studies indicates there is clear genetic predisposition to diabetes and its complications (Cole and Florez, 2020).With the aid of linkage analysis and genome-wide association studies (GWASs), scientists have been able to find more than 100 susceptibility genes and 200 susceptibility loci associated with the development, progression, and prognosis of type 2 diabetes over the last decade (Langenberg and Lotta, 2018). Using polygenic risk score from these genes, the possibility of having type 2 diabetes can be predicted (Talmud et al, 2020). In order to stress the importance of early detection and management, guidelines recommend early and regular screenings for individuals with risk factors for diabetes, including having family members with diabetes, being prediabetic, having gestational diabetes history, and insulin resistance (Kim et al, 2017).

Machine learning and data mining approaches were applied to detect important variables and assess their relationship. Such techniques have only been used recently in the healthcare sector to find hidden features and suggest computer-generated patterns for massive databases with diverse attributes (Arabasadi et al, 2017). There have been various models developed to predict diabetes under different conditions. Yet, as their performance has been inconsistent in different data sets and designs, among others, their application in clinical setting has been restricted. For example, currently available models for predicting risks of developing diabetes evaluate either genetic predisposition (polygenic risk scores) or lifestyle factors without considering the interaction between these variables. The goal of the study was to develop a model for predicting the risk of diabetes that takes into account the interaction between these two factors.

1.2 Statement of the Problem

However, there is inconsistency in feature selection despite the improvements made in machine learning for predicting diabetes. Most of the research pays more importance to accuracy than other features, which means that the features that are assumed to have high impact on the outcome are neglected. The current machine learning models do not use both genetic and lifestyle information. There are some issues with recent diabetes risk prediction systems. These include the following:

1. Isolated features: The current machine learning models for diabetes prediction consider one factor at a time such as either genetic or lifestyle.
2. Data integration: The current machine learning models are not based on the dataset that includes all the necessary genetic and lifestyle information.

3. Not Being Personalized in Interventions: Latest techniques fail to give personalized suggestions considering the individual risk factor profile and hence reduce their reliability and effectiveness.

1.3 Aim and Objectives of the Study

The purpose of this study is to propose and develop an interpretable mobile-based health decision support system for diabetic risk assessment with the help of hybrid deep learning approach, which helps in getting accurate and interpretable predictions. The specific objectives are:

1. Analyse existing machine learning and deep learning techniques for diabetes risk prediction and explainable AI approaches.
2. Collect and preprocess relevant healthcare datasets for diabetes risk assessment.
3. Train a hybrid deep learning model for accurate diabetes risk prediction.
4. Integrate explainable AI techniques to provide interpretable and transparent predictions.
5. Evaluate a mobile-based decision-support system for real-time diabetes risk assessment.

1.4 Significance of the Study

The current research study provides ample scope for improving personal health management due to the utilization of data analysis and predictive modeling. The completion of the current research study would be of great help for the patients in managing their health status on real-time basis. Personal health suggestions and efficient handling of chronic illnesses can be possible through the present research study. With the help of real-time monitoring and predictive analytics, the patients could identify the symptoms of health-related issues and act upon them in a proper way. In this case, the healthcare professionals and caregivers can gain access to the latest information about the patients in remote places and formulate effective treatment plans.

This research can provide the framework for future research regarding personal health tracking and predictive analysis. Future researchers may use this study in order to analyze the significance of predictive analysis in health care management.

1.5 Scope of the Study

In this study, the scope will be in the designing of a web based system for predicting the risk of diabetes which considers both genetic and lifestyle risk factors. The system will use comprehensive datasets on lifestyle risk factors, genetics, and clinical risk factors to recommend an individual of how he/she can improve his/her health status. This study will design a web based application which will run on web browser on Windows or any other platform. The application will have a user-friendly interface to collect data using web-based platforms.

1.6 Definition of Terms

Diabetes -Diabetes is a condition which results from insufficient insulin production by the pancreas or inability of the body to utilize insulin produced.

Type 2 diabetes - Type 2 diabetes is a situation where blood sugar levels increase because of poor body utilization of the insulin. It was referred to as adult onset diabetes in the past.

Lifestyle factors for diabetes - Lifestyle factors for diabetes involve activities which determine one’s likelihood of contracting type 2 diabetes. They include among others diet, exercise, sleeping patterns, etc.

Diabetes Genetic factors - Genetics have a strong role to play in determining one’s susceptibility to diabetes, specifically type 1, type 2, and monogenic diabetes. But, there are also interactions between genetics and lifestyle/factors in the environment.

Machine learning: This refers to a subfield within AI concerned with developing algorithms/models whereby a computer can learn from the data.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The current chapter in this research work, however, aims at evaluating the increasingly growing issue of diabetes risk prediction model which takes into account both genetic and lifestyle elements. What the method will be capable of is looking at datasets in the form of large-scale that include both genetic information and other medical records. In addition to this, the chapter will look at literature related to diabetes risk prediction, particularly the traditional approach and the application of machine learning to web based management systems. This chapter is divided into two main sections: theoretical background and related literature from various authors.

2.2 Theoretical Background

This section focuses on the theories, terminologies, and background of the disease diabetes and machine learning techniques used to diagnose it. We will delve further into the topic of machine learning, as well as the concept of diabetes and its types. Diabetes is a health condition associated with the presence of excess blood glucose or sugar in the body. According to Firdous et al, (2022), Diabetes Mellitus is a metabolic disorder caused by excess blood glucose resulting from insufficient secretion of insulin or insulin resistance.

2.2.1 Overview of Diabetes as a Global Health Concern

This is an irreversible metabolic disorder that affects various aspects of the patient’s life including their physical and psychological health. There is no way of reversing the process or achieving any outstanding improvements through any treatment procedure (Norris et al, 2020). According to the International Diabetes Federation (IDF), in 2023, Nigeria has a prevalence rate of 3.6% of diabetes among adults aged 20-79 years old with about 3.6 million people living with diabetes with 53.3%

being undiagnosed (IDF, 2024).The factors that contribute to diabetes include urbanization, aging, lack of physical activities, and overweightness and obesity (Olamoyegun et al, 2024). Symptoms of diabetes include blurred vision, feeling tired, loss of weight, increased thirst and hunger, confusion, need to urinate frequently, poor wound healing, infections, and inability to concentrate (Firdous et al, 2022). Diabetes is a condition with high blood sugar level. The problems arise when there is not enough production of the hormone insulin in the body (Wagai and Romshoo, 2020).

2.2.2 Types of Diabetes

Diabetes mellitus (DM) has many causes, which is a metabolic disease. Diabetes is characterized by constant hyperglycemia and metabolic changes in proteins, lipids, and carbohydrates because of inadequate insulin production or because of actions of insulin, or both. Diabetes is a chronic disease. If untreated, diabetes may lead to the damage of blood vessels and nerves in kidneys, heart, lower extremities, and eyes (Firdous, Wagai, & Sharma, 2022). Problems may occur due to prolonged high levels of blood glucose. Some of the problems related to oral health include gum diseases and cavities. Diabetic retinopathy may cause blindness and, in some cases, vision loss. There are heart and blood vessel diseases including peripheral artery disease and heart attacks (Firdous, Wagai, & Sharma, 2022). There are three types of diabetes: type 1, type 2, and gestational diabetes.

A. Type 1 diabetes (T1D)

Insulin production is hindered by type 1 diabetes. The absence of insulin prevents body cells from absorbing glucose from the bloodstream; hence, they have to use other sources of energy. Diabetes and its complications are caused by too much glucose in the bloodstream. The full medical term for such diabetes is insulin-dependent diabetes mellitus. It can affect people of all ages, although it tends to affect adolescents and teenagers. The balance between insulin injections (and sometimes taking pills), exercises, dietary management, and lifestyle changes is critical for treating such diabetes. Type 1 diabetes symptoms may consist of excessive urination, increased thirst, excessive hunger, rapid weight loss, weakness, exhaustion, nausea, and irritation (Firdous et al, 2022).

B. Type 2 diabetes (T2D)

Type 2 diabetes that can either exist independently or together with secretory dysfunction is a condition characterized by insulin resistance or the presence of an insufficient quantity of insulin required to keep the glucose levels in check.

Though type 2 diabetes is usually seen among individuals above 40 years of age, there are increased incidences of this disease among teens and children. Symptoms include feeling sleepy, having dry and itchy skin, experiencing unexplained weight gain or loss, blurred vision, tingling sensation and numbness, pain in legs, feeling easily tired, slow healing of cuts or bruises, and frequent infections like vaginal infections. Type 2 diabetes management includes diet, physical activity, lifestyle changes, and sometimes even insulin and oral drugs (Wagai and Romshoo, 2020).

C. Gestational diabetes

Gestational diabetes refers to high blood glucose in women who were not diabetic before their pregnancy. This temporary type of diabetes affects around 2% to 4% of pregnant women; it usually clears up after the delivery. Women who had gestational diabetes before are at risk of type 2 diabetes.

The reason for developing this type of diabetes is still unclear. While hormones produced by the placenta play an important role in development of a fetus, they interfere with proper function of the mother's insulin in her body, making it resistant to insulin. Gestational diabetes occurs when a woman's body cannot produce enough insulin or use it properly during pregnancy. Usually, women do not notice any signs of gestational diabetes. Two of its symptoms are excessive thirst and frequent urination.

The cause of this kind of diabetes is unknown. While placental hormones aid in the baby's development, they also hinder the mother's insulin from functioning correctly in her body, leading to insulin resistance. Gestational diabetes arises when a mother's body is unable to produce and utilize all of the insulin needed during pregnancy. Most women are not aware of any gestational diabetes symptoms or indicators. Two symptoms include increased thirst and increased frequency of urination.

Currently, many methods of medical diagnostics are in use for early diagnosis. For the prediction of early risk, ML is being utilized. As far as prediction of diabetes mellitus risk is concerned, some recent researches have given positive results. Machine learning basically consists of algorithms that help teach machines without any human intervention. They can be trained to do certain tasks without programming them and then apply their training on other similar tasks. Accuracy always remains an issue in medical sciences, and various algorithms result in the generation of different levels of accuracy in datasets. Comparison of algorithms or hybrid algorithm approach should be used for getting a better classifier. Today, machine learning is used in almost all industries. There are vast improvements that can be made in healthcare through machine learning in medical science.

2.2.3 Pathophysiology and Risk Factors

Metabolic abnormalities and diabetes are determined by genetic risk factors (Kong et al., 2015). Recently, the cause of type 2 diabetes has been elucidated via genome-wide association study, which revealed 143 genetic loci associated with the disease (Xue et al., 2018).

However, discovered genetic risk factors explain only partial phenotype variance (Xue et al., 2018), therefore, they are unreliable indicators of individual genetic (polygenic) risk of type 2 diabetes at early stages of life. A novel prospective technique to predict polygenic risk is a whole-genome approach that uses all common SNPs but not a few genome-wide significant SNPs (Fernando et al., 2017; Khera et al., 2019). Recently, it was shown that application of complex statistical approaches and designs may enhance polygenic risk predictions (Maier et al., 2018; Lloyd-Jones et al., 2019). In addition to the genetic factor, environmental factors, like unhealthy diet and a sedentary way of living, also increase the risk of type 2 diabetes (Dendup et al., 2018). Thus, reducing alcohol and tobacco consumption, as well as enhancing physical activities and nutrition are all included into T2DM treatment strategy. Despite effectiveness of these treatments, they are regularly recommended to people at high risk of metabolic syndrome.

2.2.4 Lifestyle-Related Risk Factors

Risk factors refer to characteristics of an individual, which may be behavioral, traits or environmental aspects that raise the probability of acquiring a certain illness. It is worth noting that lifestyle choices greatly contribute towards the development of the most common type of diabetes, type 2 diabetes. Lifestyle risk factors include:

Physical Inactivity: This is due to lack of physical exercise, hence reducing insulin sensitivity. Sitting idle in one place for extended periods may cause risk.

Poor diet: Taking a poor diet high in refined sugar, saturated fats, and canned food (processed food) can cause obesity and insulin resistance.

Obesity, being overweight, and smoking contribute to insulin resistance and inflammation most of the time limiting insulin action.

Health behaviors have an influence on some biological risk factors. In particular, health behaviors and biological risk factors can be affected either directly or indirectly by social determinants of health (AIHW 2020). There is evidence that type 2 diabetes has been linked to socioeconomic factors including low annual household income and low education levels (Zhang et al. 2020). Factors that are modifiable and fixed can neither change risk factors. The fixed risk factors for type 2 diabetes include age, sex at birth, ethnicity, and genetic predisposition to diabetes.

Despite being uncertain about the exact cause of type 1 diabetes mellitus, a family history of the condition is considered a fixed risk factor for type 1 diabetes. Unlike other types of diabetes, there are no specific behavioral risk factors that increase the risk of having type 1 diabetes; however, living a healthy lifestyle is an essential aspect in the management of the symptoms and implications of the disease. Other non-traditional risk factors, such as mental disorders, are associated with an increased risk of developing all types of diabetes except type 1 (Lindekilde et al. 2021).

Type 2 and gestational diabetes are associated with multiple risk factors unrelated to the behavioral risk factors. However, there are some behavioral and biological factors connected to both diseases, increasing the probability of the condition and its consequences. Furthermore, the metabolic syndrome, which involves multiple biological risk factors sharing the same pathogenesis, increases the risk of type 2 diabetes (Harris 2013).

Diabetes mellitus is one of the many examples of chronic diseases which share common biological and behavioral risk factors. By reducing the risk of disease incidence and mortality, modification of behavioral risk factors reduces the probability of an individual developing type 2 diabetes early and will have considerable health benefits for him or her. According to research, individuals with recent diagnosis of type 2 diabetes can be in remission through behavior modifications such as dietary modifications and weight loss (Zoungas and Sumithran 2024; Hocking et al. 2024).

2.2.5 Genetic risk factors for type 2 diabetes

There are a variety of factors which influence the probability of having type 2 diabetes. One of such factors is hereditary one. There are many genes that cause the appearance of various metabolic problems. Nevertheless, the gene makes you susceptible to type 2 diabetes but the disease is not hereditary. Regardless of the absence of the genetic tendency, you might get type 2 diabetes at any point of your life.

As a rule, the choice of lifestyle and other external factors are those that determine if the susceptibility will be developed into the condition of type 2 diabetes. The key factors in this case are overweight, diet, and exercise. The genes responsible for the glucose metabolism can be affected by the way of living. Such factors are known as epigenetic ones.

The genetic information that makes up one's “blueprint” is encoded within the genes of that person. For example, a person's genes determine the color of his/her skin, hair, and eyes. The genes that read the genetic information and encode the proteins are also responsible for metabolic functions of the body and others. Epigenetics, a relatively young field of science, studies the connection between mutable external factors and immutable genetic information of a person. Your way of life (like what you eat or how much you exercise) can be considered an external factor. It can be affected by epigenetics, which is "above" genetics.

Epigenetics is the chemical modification of the activities of genes. The extent to which genetic information is accessed and utilized by the body is modified by such changes. Epigenetic influence on genes involved in metabolism can make the development of type 2 diabetes possible or impossible, depending on the nature of the epigenetic influence (Fritsche et al, 2021). The concept that epigenetic factors, apart from genes, could be passed on to offspring is one great achievement. Scientists used mice to show how epigenetic mechanisms could cause diet-induced type 2 diabetes to be passed on to the offspring via eggs and sperms.

Though the epigenetic factors can be intentionally altered, the genetic code of the gene itself cannot be modified. This can be done through adopting an appropriate lifestyle. Scientists say that epigenetic factors of many patients can be improved with the help of a healthy lifestyle involving a balanced diet and regular physical exercises. People who are affected can try to fight against the possibility of development of type 2 diabetes within their families or in case early signs of sugar metabolism problems appear. Type 2 diabetes can often be prevented or at least delayed through adopting a proper lifestyle. Here is what you can do to prevent type 2 diabetes:

If you are overweight, lose weight, especially the fat around your abdomen.

- Consume a diversified and well-balanced diet.

- Maintain an active lifestyle by engaging in enough and consistent physical activity.

2.3 Machine Learning Algorithms

Machine learning has continuously been applied over the years in the identification of what causes the development of diabetes as well as the probability of a person having diabetes. In addition to the task of predicting the risk of developing diabetes, machine learning is also employed in the prediction of the problems associated with diabetes, including retinopathy, nephropathy, and neuropathy.As time went on, researchers utilized psychological information for analysis and prediction of depression in diabetic patients (Khalil and Al-Jumaily, 2017). Many researchers have collected data through several secondary sources in the form of electronic medical records (EMR), rather than collecting data through survey.

Many medical applications have adopted machine learning technology. The majority of machine learning algorithms employ past data in their predictions. Latest developments in machine learning will simplify and reduce costs for the detection of diabetes. There are many datasets for diabetes. As a result, machine learning is required in medical diagnostics.

ML refers to several techniques used in statistics that enable computer systems to learn and develop without explicit programming (Abdelaziz et al., 2018). Machine learning involves making changes to the behavior of algorithms due to learning. It happens when ML is able to identify faces by analyzing millions of images of different people. The two types of machine learning are supervised learning and unsupervised learning. The former involves learning with guidance while the latter involves learning without any help. The healthcare industry is one of the largest industries in the world that can benefit from ML.These algorithms can be applied for making predictions and classification of the data. Remember that model and algorithm are not the same. In contrast, while a model is the output of the implementation of an algorithm on a dataset, an algorithm is a series of rules and processes designed for solving a certain problem or executing a certain task. You already have an algorithm before training your model. And only after the training process, you will get a model.

For example, machine learning algorithms are often used in the health care sector for such tasks as illness diagnosis, forecasting, and analysis of medical images. Analysis of various medical images such as MRI, CT, and x-ray scans can be conducted through the use of machine learning algorithms that allow finding hidden patterns in the image.

In addition to recommending tests for cancers, diabetes, and heart disease, machine learning algorithms can analyze information on genetics, symptoms, and other health-related factors. There are four main types of machine learning. Each type comes with its pros and cons, and therefore, choosing the right one for the particular task is crucial.

The first and the most common form is supervised machine learning. In supervised machine learning, the algorithm learns the rules based on labeled data. With supervised machine learning, the algorithm is taught to learn from a labeled data set, just like how a child recognizes fruit based on a picture book. An example of supervised machine learning would be spam email filters, in which the algorithm is taught using a labeled data set consisting of emails marked either as spam or not spam. The model can then classify new emails into spam and non-spam using the learned patterns. The data is labeled in supervised machine learning, and it is similar to how a pupil learns from a teacher.

On the other hand, unsupervised learning lacks the use of labels for the data. Unsupervised learning is more like self-learning through experience. The goal of the method is the prediction of the values of the variable. Data are described using the set of attributes and features. There is a pre-specified result in guided learning. Some of the commonly used techniques include decision trees (DT), random forest, logistic regression, linear regression, Naive Bayes classifier, k-Nearest Neighbors (k-NN), support vector machines (SVM), and artificial neural network (ANN). This type of supervised learning requires the work of an expert to label data for the correct result in the future. Through unsupervised machine learning, the computer is able to understand complex patterns and processes without having any form of input that has been labeled beforehand. In addition to utilizing unlabeled data as input, the unsupervised machine learning does not have any output, which includes knowing whether an email is a potential spam.

Clustering usually involves unsupervised machine learning algorithms that identify groups in the data set. After training, the algorithm can be able to recognize similarities and then classify the data in the correct cluster. Recommendation engine is an example of unsupervised machine learning algorithm that is used in customer applications to make suggestions of "customers who bought that also bought this." This algorithm can be able to identify patterns that differ from anomalies, making it useful in fraud detection. In unsupervised learning, the output is not predetermined and data comprises of values without labels. Prediction in this case happens through self-learning of the model. The primary goal of these models is to predict, classify, detect, segment and categorize data. Semi-supervised machine learning solves the problem of insufficiently labeled training data that is necessary to effectively train the model. You may have a large number of training data, for example, but you do not want to waste time and money on their labeling. Sometimes, in order to get a fully trained model, you need to use both supervised and unsupervised machine learning approaches. The training process starts in a similar way to supervised learning, where labeled data is used. As soon as there are no more labeled data sets to work with, unlabeled datasets are used to feed the semi-trained model. The unlabeled data is examined by the already trained model in order to help enrich the labeled dataset with additional data. If the model can determine an appropriate label for the sample with great accuracy, it will be added to the labeled dataset. The training process restarts with a larger set of labeled samples.

Reinforcement machine learning involves the use of unlabeled datasets whereby the algorithms analyze the data just as is the case with unsupervised machine learning. The difference between reinforcement learning and data exploration is that reinforcement learning aims at meeting the predetermined objective unlike the former which searches for patterns within the data. The algorithm learns in a trial and error way with the objective in mind. The algorithm uses the feedback from each action whether positive, negative or neutral to enhance its decision making ability. Reinforcement learning approach is able to meet the objective of the project on a large scale despite any negative consequences that may be experienced temporarily. Therefore, since reinforcement learning approach allows for the consideration of the objective of the project in the risk of choices, it is able to handle a more complicated situation compared to other approaches. For instance, training the computer to play chess; winning the game is the final objective but this may involve giving up some pieces during the game.

2.3.1 Convolutional Neural Networks (CNNs)

Convolutional neural networks are a type of deep neural network used to process grid-like structures of data, which has made them an important type of deep learning model (Mahmood and Qamar, 2025). The convolutional neural networks were initially developed for image recognition purposes by using convolutional layers to apply the learnable filters that detect the edges and textures of local patterns by weight sharing and local connectivity (Nerella et al., 2024). The process is further followed by the application of fully connected layer for high-level feature integration and classification and also pooling operations to reduce the spatial dimensionality. This will lead to efficient hierarchical feature extraction process by the CNN framework. Currently, the application of CNN in the time series analysis is increasing, which includes the forecasting of hydrology data, while CNN idea is proposed by Lecun et al. (1998) through the LeNet-5 model. Convolutional neural networks have brought about a revolution in computer vision and pharmacogenomics fields due to the outperformance of their traditional counterparts in terms of the image classification and biological sequence analysis, despite the fact that CNNs may require large amounts of data to avoid overfitting (Vaz & Balaji, 2021). With the passage of time, it was found that the same techniques that make CNNs able to learn hierarchical visual features could be used in the case of temporal data too. An experiment conducted by Bai et al. (2018) has provided evidence regarding the performance of CNNs over recurrent networks such as LSTM due to its efficient gradient flow. Figure 2.2 shows the architectural representation of CNN that is made up of input, convolutional layer, pooling layer, and output.

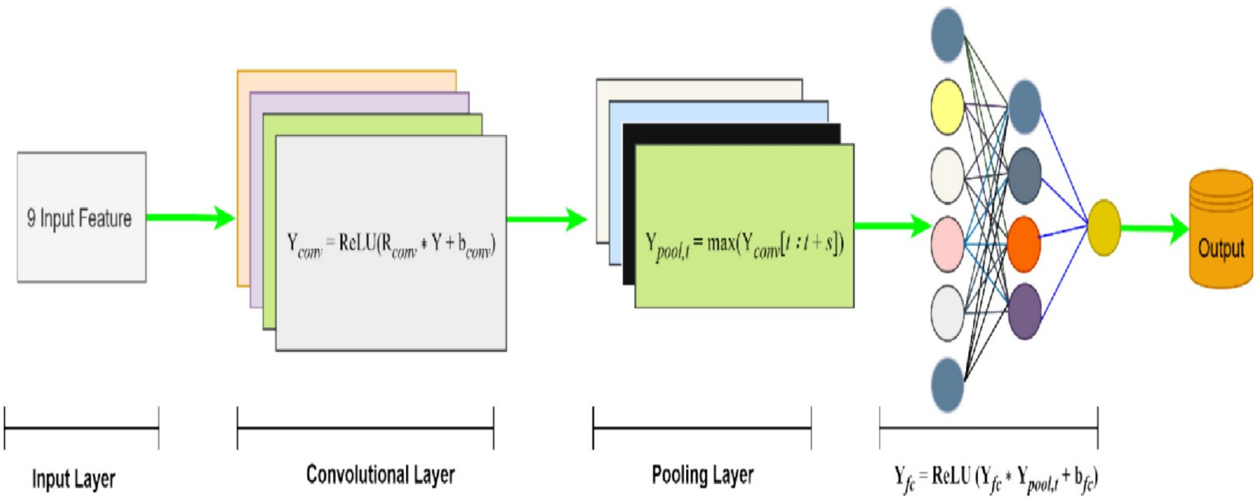


Figure 2.1 schematic diagram of the CNN architecture (Rahman et al., 2025)

2.3.2 Transformer model

Transformer Model by Vaswani et al. (2017)

Transformer architecture represents a shift of paradigm in sequence modeling since the approach does not employ recurrence and convolution but focuses on self-attention models (Rahman et al., 2025). This architecture includes stacked encoder and decoder blocks with multi-head attention, positional encodings that will be employed to maintain sequence order, and feed-forward networks, providing better scalability and training efficiency on tasks such as machine translation (Nerella et al., 2024). This approach makes it different from CNNs or RNNs, which deal with sequence modeling through sequential and local processing. It provides better opportunities for modeling long sequences when distant interactions are important. According to Vaswani et al. (2017), the possibility of simulating global context without being constrained by sequentiality has led to reaching state-of-the-art results for transformers, including 28.4 BLEU on English to German translation. Figure 2.3 shows the architectural design of the classic Transformer, which includes two blocks: an encoder and a decoder.

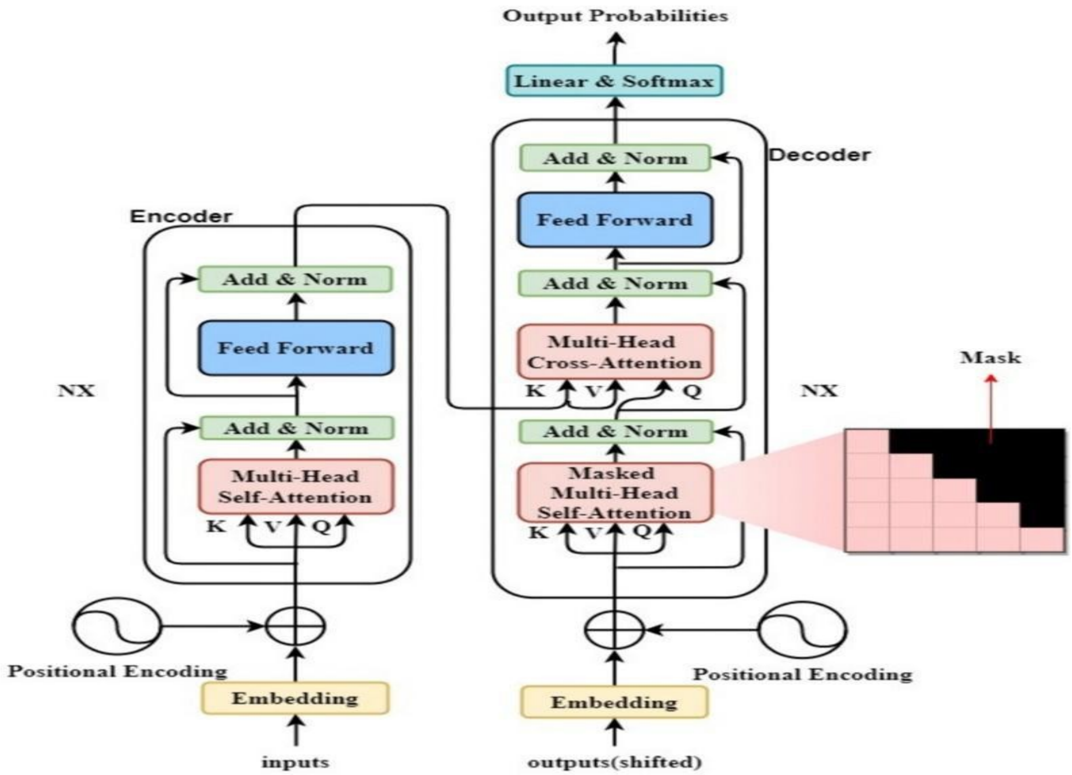


Figure 2.2 Architectural diagram of the Transformer model (Nerella et al., 2024)

The role of the encoder is to map the input sequence into latent representations by employing MHSA along with FFN with stacked layers, while the decoder takes up the responsibility to generate the output sequence by employing the mask self-attention mechanism over its inputs along with the encoder-decoder attention mechanism to leverage the encoder's representations (Rahman et al., 2025).

2.3.3 Hybrid models

In machine learning and statistics, hybrid models are those that integrate several algorithmic paradigms or model components to utilize their unique capabilities, thus improving prediction, interpretability, and robustness as compared to standalone techniques. In predictive modeling, hybrid models will blend interpretable models such as linear models and association rules with black-box models such as neural networks or ensemble methods. The interpretable components will be designed to process data subsets where they work best, replacing the black box for balancing accuracy and interpretability through objective functions that discourage misclassification errors, poor coverage, and model complexity (Tong & Qihang, 2021). An example of hybrid models is CNN and RNN, convolutional neural networks (CNNs) to extract spatial features and at the same time recurrent neural networks (RNNs) including LSTMs and GRUs to do temporal modeling, where it is particularly effective in modeling spatiotemporal data such as videos, time series or sequences of images (Hu et al., 2018; Munguia-Siu et al., 2024).

According to Vadhnani and Singh (2022), the use of SVM variants in MRI image analysis was surveyed with very accurate results. The use of SVMs was used to integrate curvelet transformation and ant colony optimization to improve the quality of the images and increase the accuracy of classification in a research by Hussain et al. (2020). This is also in line with previous studies that concluded that the combination of CNN and RNN models in offline handwritten text recognition performed better than individual CNN and RNN models. According to Ma (2023), this model resulted in a decrease in CER by 5.13% and WER by 12.04% in comparison with baseline CNN.

2.4 Review of Related Literature

The authors Birjais et al (2019) performed experiments using the Pima Indians Diabetes (PID) dataset. The machine learning data from the UCI repository consists of 768 observations and 8 attributes. According to the statement made by the World Health Organization (WHO) in 2014, diabetes is one of the fastest-growing chronic conditions in the world; therefore, they wanted to focus more on the diagnosis of the disease. Three types of classifiers were applied to diagnose whether a person has diabetes or not; these are gradient boosting, logistic regression, and naive Bayes. The accuracy of gradient boosting was 86%, of logistic regression – 79%, and of naive Bayes – 77%.

Experimentations were carried out using the diabetes data set in the UCI repository by Sadhu and Jadli (2021). A total of 16 features and 520 instances were used. An attempt was made by them to concentrate on early detection of diabetes. Classification techniques such as k-NN, logistic regression, SVM, naive Bayes, decision trees, random forests, and multilayer perceptrons were employed to validate the data set under consideration. From the findings obtained after training different machine learning models, the random forests classifier emerged to be the best algorithm for the considered data set, achieving an accuracy score of 98%. Others were logistic regression (93%), SVM (94%), naive Bayes (91%), decision tree (94%), random forests (98%), and multilayer perceptron (98%).

In their study, Xue et al (2020) used a dataset for diabetics available in the UCI repository, which had 17 attributes and 520 patients. They tried to concentrate on early detection of diabetes. Their training was done on real data on 520 patients suffering from diabetes and probable patients suffering from diabetes aged between 16 and 90 years old by using supervised machine learning approaches such as SVM, LightGBM, and naive Bayes classifiers. SVM has the best accuracy in classification and recognition. Naive Bayes Classifier has the best accuracy of 93.27% and hence it is the most used classification algorithm. It has the highest accuracy rate of 96.54%. The accuracy rate for LightGBM is only 88.46%.

Diabetes risk prediction during the early stages was assessed by Le et al. (2020). The data set used in this study was obtained from the UCI repository and comprised of 16 attributes and 520 patients. They introduced a new method for the forecasting of the early occurrence of diabetes among patients. In their work, grey wolf optimizer (GWO) and adaptive particle swarm optimization (APSO) algorithms have been applied in this innovative wrapper-based feature selection algorithm, to optimize MLP and reduce the number of input attributes needed. Moreover, they compared the performance of the developed algorithms with those of other traditional machine learning techniques, such as logistic regression (LR), naive Bayes classifier (NBC), random forest classifier (RFC), SVM, DT, and k-NN. As seen in the computational results of the suggested algorithms, it is possible to attain high prediction accuracy (96% for GWO-MLP and 97% for APSO-MLP) while reducing the number of required features. This research can serve as a clinical instrument to assist physicians.

In this experiment, Julius et al. (2020) evaluated a data set from the UCI repository through the use of the Waikato Environment for Knowledge Analysis (Weka) application platform. This data set consisted of 520 samples, each having a total of 17 features. This study aimed at predicting diabetes in advance through machine learning classification based on observable sample features. The classifiers used were the k-NN, SVM, FT, and RFCs. The accuracy attained was 98% for k-NN, 94% for SVM, 93% for FT, and 97% for RF.

The prediction of diabetic illness was conducted by Khanam et al. (2021). As diabetes is presently incurable, it is vital to diagnose the disease early. In the research mentioned above, diabetes was predicted through NN methods, data mining, and machine learning. It should be noted that the researchers developed an accurate approach to predicting diabetes. The authors analyzed the data of PID data set from the UCI repository. The dataset contained information about 768 people and their nine attributes. Seven machine learning algorithms were used to predict diabetes: DT, k-NN, RFC, NBC, AB, LR, and SVM. To preprocess the data, the researchers used the Weka tool. It was discovered that diabetes could be predicted accurately by using a model with a combination of LR and SVM. After several experiments with various epochs and creating a NN with two hidden layers, it was found out that the accuracy of the NN with two hidden layers equals 88.6%. The values for ANN, LR, NBC, and RFC equalled 88.57%, 78.85%, 78.28%, and 77.34%, respectively.

An online application was suggested by Nai-Arun and Moungmai (2021), involving disease classifiers and the use of real data set. From 2012 to 2013, data was obtained from 30,122 patients who visited 26 primary care units of Sawanpracharak Regional Hospital. Before developing an online application, 13 categorization models were tested in order to identify a suitable prediction model. These models included bagging and boosting procedures that incorporated the DT, NN, LR, NBC, and RFC algorithms, with the only exception being the RFC approach. Robustness of each model was determined based on the accuracy and ROC curves calculated and compared to other models' ones. It was found that RFC demonstrated the highest level of accuracy and ROC curve. There might be many reasons for this. Important variables were taken into account while using RFC technique, besides the data and inputs chosen randomly. Precision values became higher. Therefore, this technique was applied in the development of the application.

2.5 Summary of Literature

S/N	AUTHOR(S)	CONTRIBUTION	TECHNIQUE	LIMITATIONS
1.	Birjais et al. (2019)	Compared Gradient Boosting, Logistic Regression, and Naïve Bayes using the PIMA Indian Diabetes dataset, achieving 86%, 79%, and 77% accuracy, respectively.	Machine Learning and comparison technique	Limited to three classifiers; no exploration of ensemble or hybrid models.
2.	Sadhu and Jadli (2021)	Tested seven classifiers (k-NN, LR, SVM, Naïve Bayes, DT, RF, MLP) on a UCI dataset, with RF	Hybrid Machine Learning	Focused on a single dataset; lacks real-world

		achieving 98% accuracy.	Approach	validation.
3.	Xue et al. (2020)	Used SVM, LightGBM, and Naïve Bayes on a UCI dataset; SVM achieved the highest accuracy (96.54%).	ML technique and OOAD methodology	Limited to supervised learning; no feature selection techniques explored.
4.	Le et al. (2020)	Proposed GWO and APSO for feature selection, improving MLP accuracy to 96–97%.	Enhanced Machine Learning and Agile method	Complex implementation; may not be scalable to larger datasets.
5.	Julius et al. (2020)	Tested k-NN, SVM, FT, and RF on a UCI dataset; k-NN achieved 98% accuracy.	ML approach	Limited to Weka platform; no comparison with advanced ML models.
6.	Khanam et al. (2021)	Combined LR and SVM for diabetes prediction; NN with two hidden layers achieved 88.6% accuracy.	Hybrid ML approach	limited dataset.
7.	Nai-Arun	Developed a web application	ML and OOAD	Limited to a

	and Moungmai (2021)	using RFC, achieving high accuracy and ROC values.	method	specific hospital dataset; lacks generalization.
8.	Mujumdar and Vaidehi (2019)	Proposed AdaBoost for diabetes prediction, achieving 98.8% accuracy.	ML and AdaBoost	Limited to specific extrinsic factors; no comparison with recent ML models.

2.6 Research Gap

Given that many other studies have considered machine learning algorithms in many ways, including comparative and hybrid ones, to predict diabetes diseases while taking into account both lifestyle and genetic factors, very few have considered a model of hybrid deep learning algorithms. Apart from that, many of them have ignored incorporating Explainable Artificial Intelligence (XAI) techniques into the prediction process. Also, there are many studies that ignore using localized data sets according to specific populations along with mobile health decision-supporting interface. This study will, however, fill the gaps created by ignoring these aspects and aims to develop a diabetes disease predicting system using hybrid deep learning algorithms along with a mobile health decision-supporting interface.

In this way, we can obtain more interpretable and accessible predictions of the patients' health conditions.

CHAPTER THREE

SYSTEM ANALYSIS AND DESIGN

3.0 Introduction

The chapter gives a detailed insight into the method adopted in developing a diabetes risk prediction model incorporating both genetic and lifestyle factors in this study. The chapter exploits the power of machine learning algorithms in developing a robust diabetes prediction model. The chapter seeks to enhance accuracy, minimize false positive or negatives and give a better early detection system.

3.1 System Analysis

The process of system analysis is defined by analyzing the proposed system and the elements that make up the system and their interaction. The part of the research work that will handle system analysis will focus on carrying out the analysis of the proposed system.

3.1.1 Analysis of Existing System

Most of the current systems used for predicting diabetes in Africa rely on the conventional manual system. The conventional manual system involves the prediction of whether a person will develop diabetes using his/her genes, past medical records, and possibly lifestyle. The manual process does not conduct any kind of analysis and therefore it cannot analyze large volumes of data to make an accurate prediction since intelligent systems can do this. Several predictive models have been developed for making the predictions of diabetes under various circumstances, but they cannot be used clinically due to inconsistent outcomes. They also lack machine learning techniques such as Random Forest, Artificial Neural Networks, XGBoost, and SVM.

3.1.2 Weaknesses of the Existing System

There are various flaws in the current prediction systems for diabetes that make them ineffective and inaccurate. They include:

Single Algorithm Approach: Current methods depend on a single algorithm for predicting diabetes. A single model is inadequate for complex problems because it may not suit the type of input or the parameters of the method.

i. Compatibility of Dataset: The chosen dataset may not be compatible with the algorithm. Such problems create less accuracy in the predictions of the output.
ii. Availability of Dataset: The current models rely on a limited dataset.
iii. User Interface Problem: There is usually no friendly interface available in the current prediction system
iv. User Interface Challenges: A user-friendly interface with easy navigation is often lacking in the existing systems.

3.1.3 Analysis of the Proposed System

AI techniques like machine learning and data mining algorithms have been extensively applied to uncover useful variables and determine their correlation with each other. These techniques have been applied recently in the medical domain to uncover underlying hidden variables and provide computing patterns that can be based on big databases with multiple parameters. The proposed system exploits the predictive ability of DL models by developing and comparing the performance of hybrid deep learning model against benchmark baseline models like Random Forest and XGBoost. The hybrid system will apply the CNN algorithm to find hidden feature patterns, find important health care factors, simplify the feature structure, and learn from new features. While the transformer algorithm will be employed for robust attention mechanism to learn feature relations, capture long-range dependencies among variables, and learn complex diabetic risks relationships.The hybrid DL system will undergo training with a high amount of data about lifestyle and genetics in order to improve the prediction and efficiency of the new system. Restating the objective of the study, which is to create an explainable mobile health decision support system for assessing the risks of developing diabetes via a hybrid deep learning system, so as to produce accurate and easily understandable predictions for preventive healthcare.

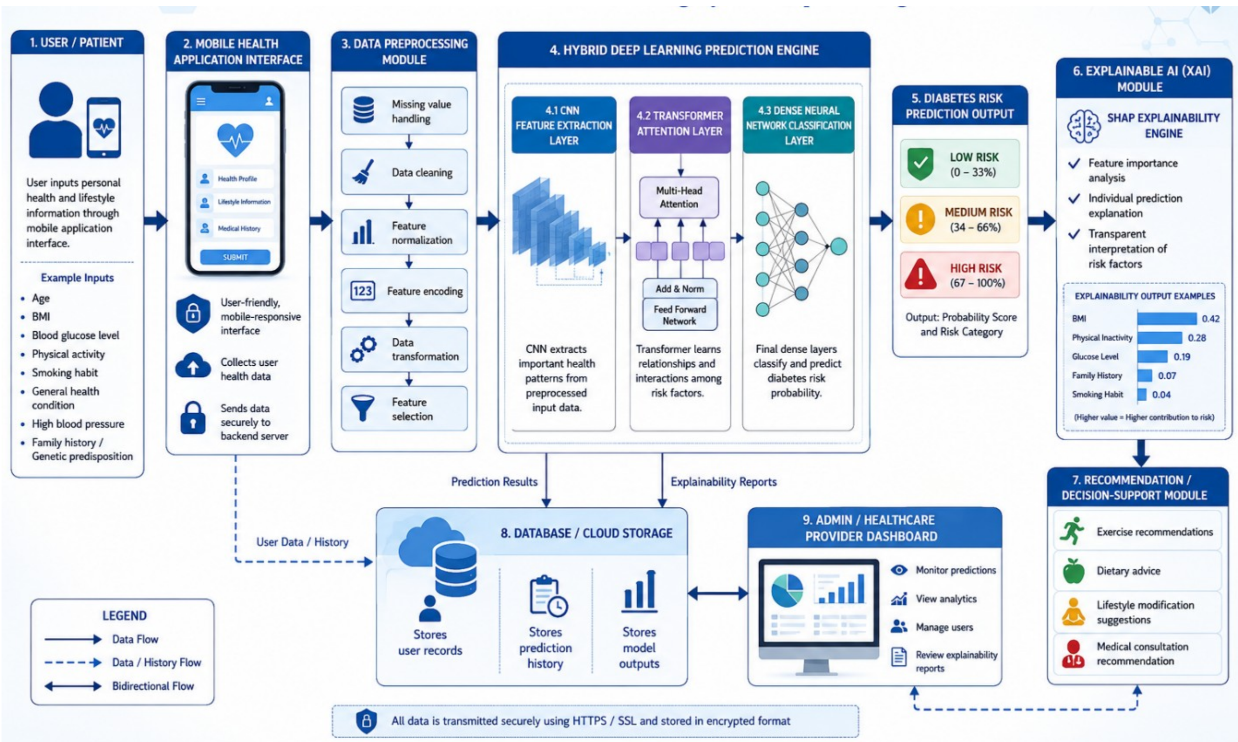


Figure 3.1 Proposed Model Data Flow Diagram

The proposed model flow diagram can be seen in Figure 3.1. In this figure, the user should enter his personal information and health information by means of the mobile application. Patient's information is transmitted in a secure way to the back-end of the system where pre-processing and normalisation of the patient's data takes place before it is being analyzed. The system applies feature selection on the patient's information in order to filter out unnecessary attributes to increase prediction accuracy. Hidden health-related patterns are extracted from the dataset by the CNN layer, whereas the transformer learns about relationship between various factors which affect the chances of having diabetes. Dense neural networks classify the overall risk of having diabetes. Predictions would be classified according to their risk levels which include low-risk, medium-risk and high-risk. On the other hand, XAI module implemented in the model which makes use of the SHAP engine explains reasons behind each prediction in order to increase transparency of the model. Moreover, the recommendation module included in the model will come up with recommendations for health improvements, exercise and diet.

3.1.4 System Requirements

System requirement is defined as the science of coming up with the critical processes or describing what the system must be capable of doing or the assumed qualities of the system. There are two classes of system requirements: Functional Requirements and Non-functional Requirements.

A. Functional Requirements:

These requirements state the functionalities that the system will have. The functional requirements for the Diabetes Risk Prediction System are:

1. Software application component of the system should be easy-to-use by end users.
2. Roles of users:
3. Admin: Administrator role would manage users' accounts and administer the system.
4. Patients: In the proposed system, patients would be able to feed in their health parameters (glucose levels, BMI, lifestyle, etc.) and get their risk predictions done.
5. Data collection and management: Users can enter their data, as well as save and fetch if required. System will analyze the entered data and make a risk prediction out of it.
6. Recommendations: Health recommendations should come as a result of risk scores shown.

B. Non-Functional Requirements:

i. Performance: The system needs to have high accuracy in prediction of diabetes risks.
ii. Speed: The system needs to have a fast processing time that allows real-time operation.
iii. Usability: It is important to have an intuitive user interface that allows corrections to be made manually.
iv. Responsiveness: The system needs to be responsive enough to be adaptable on different sizes of electronic screens used by mobile phones.
v. Reliability: The system is going to be reliable irrespective of any variation in the data or environment.
vi. Robustness: The system needs to be robust to any variation in data characteristic such as noise, outliers, and missing values.
vii. It is important for the system to provide authentication to users including the patients, medical staff, and the administrator.

3.2 Research Design

The research process in this study will involve use of design science research methodology and an experimental approach for developing the artificial intelligence system to successfully develop the hybrid system. The experimental approach will be used in developing the intelligent health care artifact that will include the artificial intelligence prediction as well as mobile health care interface. This will be through the use of predictive modeling as well as software engineering.

3.3 Dataset Collection and Description

The primary source of data that will be used for this experiment is the CDC Diabetes Health Indicators Dataset, which can be accessed publicly. There are over 250,000 records of CDC

BRFSS available in the database, which have attributes such as demographics, lifestyle, and health indicators that can be used to conduct diabetes risk prediction experiments. The type and nature of the data available in the dataset makes it a relevant source that could provide generalization during the training phase of the AI model, apart from having binary diabetes classification labels and attributes such as BMI, smoking, physical activity, and health status. The second dataset under consideration is the PIMA Indians dataset, which is a benchmark dataset for conducting diabetes predictions.

3.3.1 Data Preprocessing

The steps in the data preprocessing stage include loading the dataset in the python environment in order to understand the shape of the dataset and names of columns. This process will also help determine the number of rows and feature dimensions of the dataset, identifying and handling missing values for all variables, as well as discarding duplicates in order to balance the classes of the dataset. This step is crucial for detecting any abnormal values or outliers in the data, and the treatment of outliers based on the nature of the data domain.

The collected dataset was carefully prepared prior to training the model in order to guarantee consistency and reliability of the learning process. Then, dates and times were transformed to machine readable format, and categorical variables were transformed into numerical form. Environmental variables in numerical form were normalized using Min-Max normalization and standardization in order to enhance the performance of the deep learning model. Finally, the dataset was separated into input and output variables, and then separated into train, validation, and test sets.

Table 3.1 Sample of Dataset

1	HighBP	HighChol	CholCheck	BMI	Smoker	Stroke	HeartDisea	PhysActivit	Fruits	Veggies	HvyAlcohol	AnyHealthc	NoDocbcCc	GenHlth	MentHlth	PhysHlth	DiffWalk	Sex
2	0.880716	-1.05078	0.160486	-0.68737	-0.95191	-0.25673	-0.41492	0.647525	-1.26086	0.516538	-0.2133	0.218636	-0.32033	1.048676	-0.45926	-0.27672	-0.57903	1.086887
3	0.880716	0.951673	0.160486	-0.40578	1.05052	-0.25673	2.410123	-1.54434	-1.26086	-1.93597	-0.2133	0.218636	-0.32033	1.948107	-0.45926	2.412148	1.727022	-0.92006
4	0.880716	-1.05078	0.160486	-0.40578	1.05052	-0.25673	-0.41492	0.647525	-1.26086	-1.93597	-0.2133	0.218636	3.121761	1.048676	0.768253	2.113385	1.727022	-0.92006
5	-1.13544	0.951673	0.160486	-0.68737	-0.95191	-0.25673	-0.41492	0.647525	-1.26086	-1.93597	-0.2133	0.218636	-0.32033	0.149245	-0.45926	-0.57548	-0.57903	-0.92006
6	0.880716	0.951673	0.160486	-0.54657	1.05052	-0.25673	-0.41492	0.647525	-1.26086	-1.93597	4.688249	-4.57381	3.121761	0.149245	-0.45926	-0.3763	1.727022	-0.92006
7	0.880716	-1.05078	0.160486	-0.96897	-0.95191	-0.25673	-0.41492	0.647525	0.793107	0.516538	-0.2133	0.218636	-0.32033	0.149245	-0.45926	-0.57548	-0.57903	1.086887
8	0.880716	-1.05078	0.160486	-1.39136	-0.95191	-0.25673	-0.41492	-1.54434	-1.26086	-1.93597	-0.2133	0.218636	-0.32033	0.149245	-0.45926	-0.57548	-0.57903	-0.92006
9	-1.13544	0.951673	0.160486	-0.40578	-0.95191	-0.25673	-0.41492	0.647525	-1.26086	0.516538	-0.2133	0.218636	-0.32033	1.048676	-0.45926	-0.07754	-0.57903	-0.92006
10	0.880716	0.951673	0.160486	1.42459	1.05052	-0.25673	2.410123	-1.54434	0.793107	0.516538	-0.2133	0.218636	-0.32033	0.149245	-0.45926	-0.57548	-0.57903	-0.92006
11	0.880716	-1.05078	0.160486	-0.12418	-0.95191	-0.25673	-0.41492	-1.54434	0.793107	0.516538	-0.2133	0.218636	-0.32033	-0.75019	-0.45926	-0.57548	-0.57903	-0.92006

3.4 System Architecture

The architecture of the hybrid DL model was structured in a way that would allow for the efficient, secure, and timely prediction of diabetes risk via a mobile health decision-support system. The architecture is made up of a mobile frontend interface that will allow the user to input his/her health and lifestyle information within an interactive environment, which is provided via the mobile app's design. The information collected will be transferred securely to a backend prediction API, which will pre-process the data and communicate with the hybrid deep learning inference engine. The inference engine will be constructed via a CNN–Transformer model, which will predict the risk of diabetes.

In order to enhance the transparency and credibility of the system, an explainability engine is introduced, which will use explainable AI approaches to offer interpretations of the prediction results. Another recommendation component included in the system will be used to offer personalized health recommendations and preventive measures according to the level of risk identified. Another component included in the system is that of database management, which will be used to store user data, prediction results and generated reports. Moreover, an authentication and security component is included in the system in order to guarantee secure access control, privacy and protection of sensitive user health data. Overall, the architecture allows for smooth and real-time prediction process to enable accurate and easy diabetes risk prediction. Figure 3.2 below depicts the block diagram of the entire system architecture.

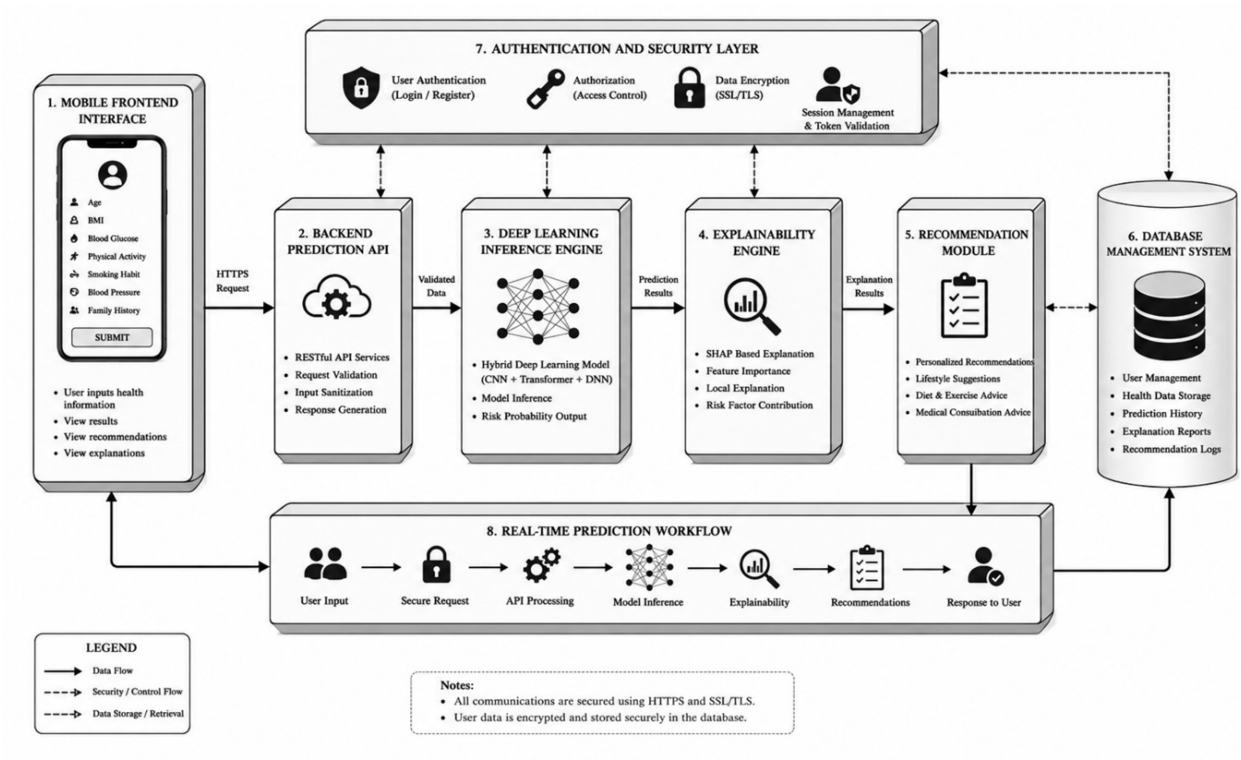


Figure 3.2 Block diagram of system interconnected modules

The integration of several inter-connected modules is presented in Figure 3.2, showing how the proposed architecture can help with diabetes risk estimation and decision making. In particular, the health data offered by the user will be essential for this process through the mobile frontend interface, which would be safely transferred to the backend prediction API. It should be noted that the deep learning inference engine will conduct analysis of the processed data in order to develop diabetes risk estimations, while the explainability engine will interpret the risk factors in order to maintain the transparency of the whole system. Moreover, the recommendation module will offer personalized health recommendations based on the results of the prediction made by the model, while the database management and authentication layer will guarantee safety of the data.

3.4.1 Proposed Hybrid Deep Learning Architecture

The designed hybrid DL algorithm is made up of CNN and Transformer models. The CNN part of the hybrid DL architecture will help in extracting hidden feature patterns, important healthcare indicators, simplify the complexity of features, and facilitate feature learning ability. Similar to the above, the Transformer model will use an attention mechanism to effectively learn feature relationships, learn long-range dependencies of variables, better understand the context of health indicators, and facilitate the learning of complex diabetes risk relationships.

The DNN layer of the hybrid model will be used for performing classification processing, nonlinear transformation of features, and generation of probability of diabetes risk. In order to improve the performance and robustness of the model, activation functions and regularisation approaches such as ReLU, Sigmoid, Dropout, Batch Normalisation, and Early Stopping are used in the hybrid model architecture. Figure 3.2 shows the diagram of the hybrid DL architecture designed for diabetes risk prediction.

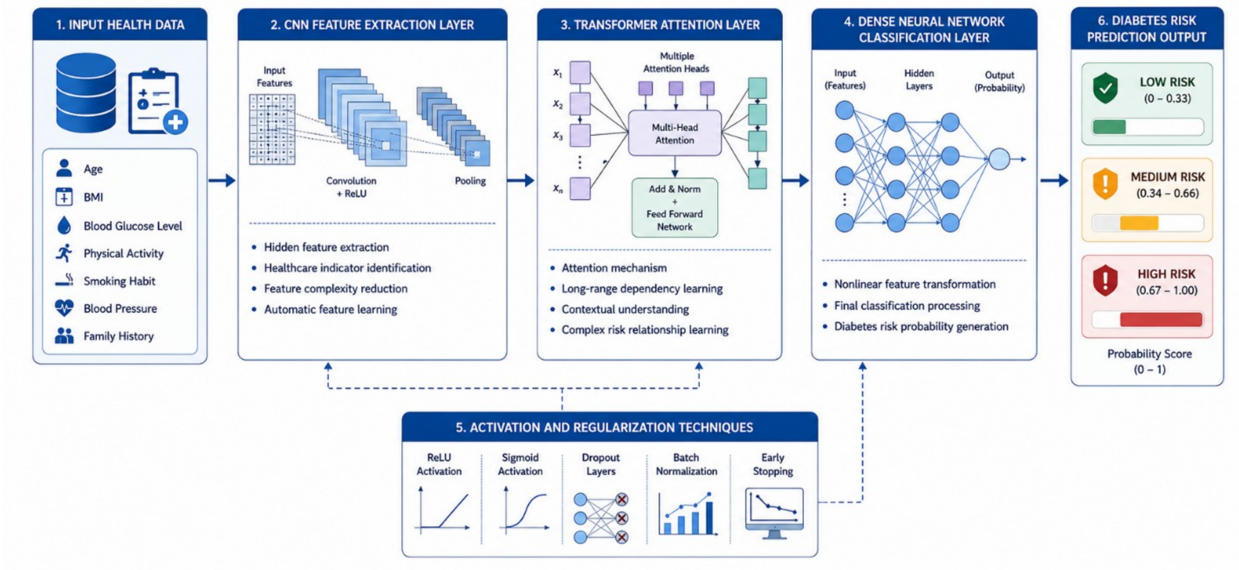


Figure 3.3 Proposed hybrid deep learning architecture

3.4.2 Explainability Framework

The use of this system is expected to incorporate the use of the SHAP (Shapley Additive Explanations) technique with the aim of enhancing the clarity and interpretability of diabetes risk predictions that have been created by the hybrid deep learning algorithm. The explainability approach offers not only global features importance but also local explanation on the individual predictions which will help ensure that the people using this system fully understand the contributing factors to the prediction results of the system. In addition, this explainable AI framework will illustrate important diabetes risk factors including BMI, glucose level, and sedentary lifestyle.

3.5 Proposed System Design Methodology

Object-oriented analysis and design, which can be abbreviated to OOAD, is a methodology that is widely used in software engineering, taking advantage of the concepts of object-oriented programming in the design of complex software systems. The key point in OOAD is that the researcher should first understand the requirements for the system, then identify objects and figure out how they will interact. This helps to improve maintainability, reusability, and scalability of software. The steps involved in OOAD are listed below:

a. System Requirements: Numerous research studies attribute the thorough analysis of the requirements needed in designing the proposed system to ensure that the final product aligns with the needs and expectations of its end users, as a core process of OOAD.

b. Classes and System Relationship: This step is critical in creating a system that is both efficient and effective; it succeeds the system requirement phase. After the system requirements have been defined, the next step in the OOAD process involves identifying the classes and their relationships within the system, with each class representing a distinct component of the system, and the relationships between these classes dictate how they interact with one another.

3.5.2 Justification for Methodology

The choice of OOAD is because it offers numerous benefits compared to traditional procedural approaches to software development. In OOAD, objects and classes can be reused across different projects, thereby reducing time and effort in development. Also, the modular design of object-oriented systems allows us to properly organise our code, making it easier to manage and extend, minimise errors, speed up development, and improve time-to-market.

In OOAD, changes to one part of the system can be localised, minimising the impact on other components, thus enhancing maintainability. They possess scalability features that enable systems to easily accommodate changing requirements and scale to meet evolving needs. Despite the numerous advantages, there are a few bottlenecks, such as complexity. Object-oriented systems can become complex, especially as the number of classes and their interactions increases.

Managing the complexity of large-scale object-oriented projects requires careful design and planning.

3.5.3 Unified Modelling Language

UML (Unified Modeling Language) diagrams are essential tools used to create a visual representation of the software system, its components, and their interactions. These diagrams help in designing and understanding the architecture, functionality, and flow of the system. In this project, we utilize Use Case Diagrams**,** Activity Diagrams**,** and Class Diagrams to depict various aspects of the Diabetes Risk Prediction System**.**

A. Use Case Diagram of Proposed System

There are four classifications of the use case diagram and these include the actors that act as the external entities of the system, use cases that act as the functional parts of the system, association between the actors and use cases, and finally the system boundary that acts as the scope of the system that the actor uses. In regards to the suggested system, there are three major actors who include the user, clinician, and admin. Figure 3.4 below shows the use cases and actors.

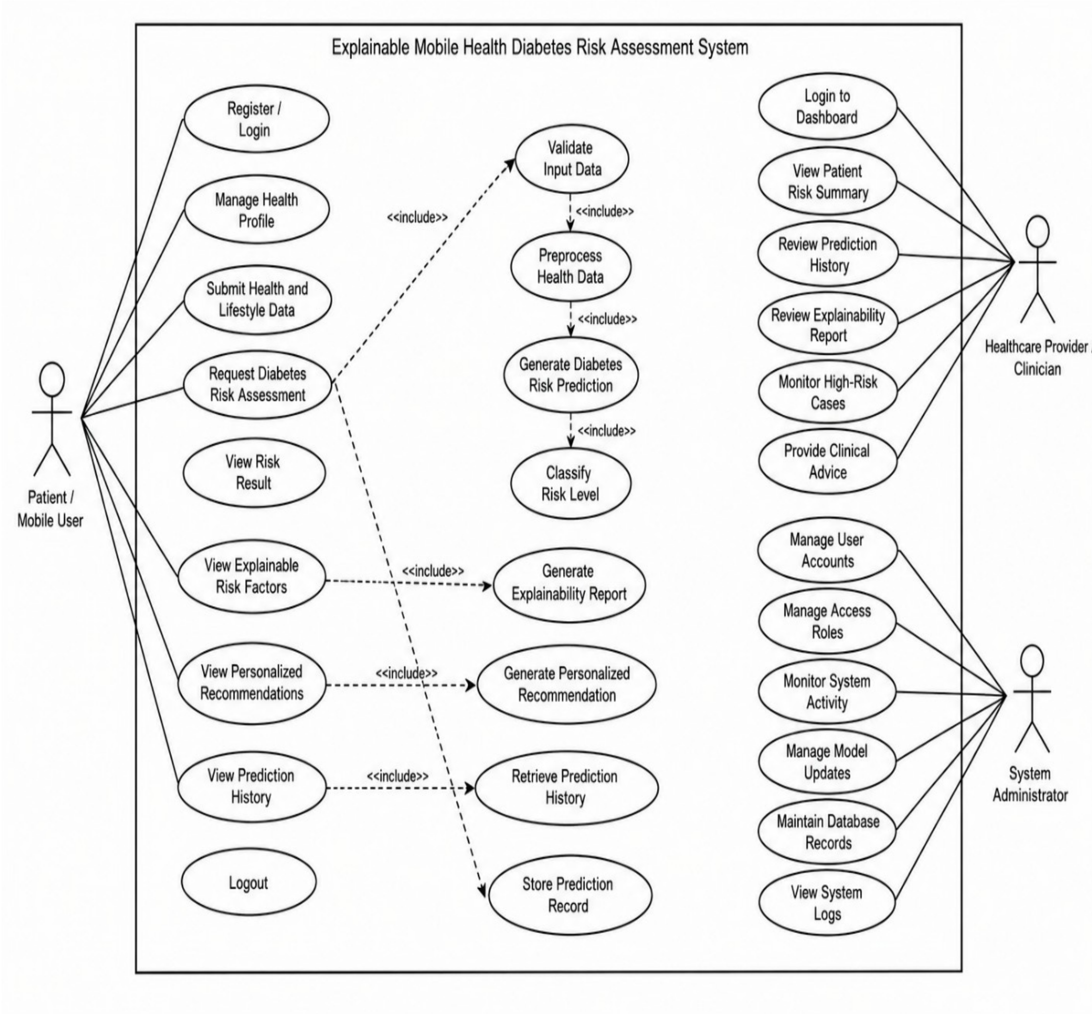


Figure 3.4: Use case diagram of the proposed system

As per Figure 3.4, the user inputs different types of health and lifestyle information into the system. The model will analyse these data in order to make predictions and give reports, and the explainable part will give reasons for the predictions. The users get recommendations on their actions on the basis of the predictive analysis done by the system. The clinician would be able to login into the dashboard to view risk summaries and prediction history of the patients along with the explainability reports, and giving his/her clinical advice.

B. Activity Diagram of the Proposed System

Another key diagram in the Unified Modelling Language (UML) approach is the activity diagram. An activity diagram is a diagram which represents the dynamic behavior of the software system. An activity diagram, on the other hand, is a flow chart describing the flow of activities systematically in the system. The flow of activities of the proposed system is illustrated in figure 3.5 below:

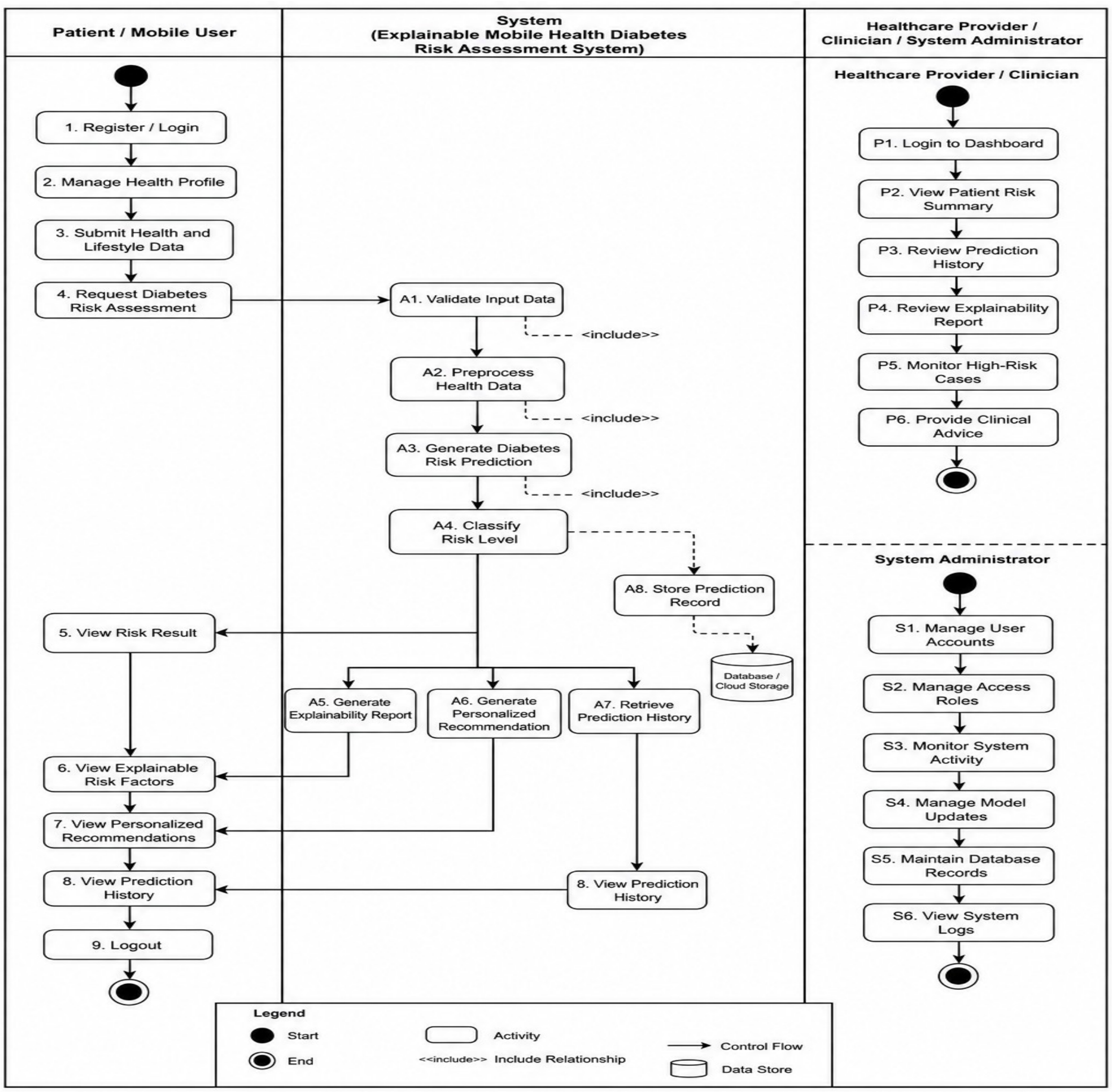


Figure 3.5 Activity diagram of the proposed system

The process starts with the user choosing to use the application. After signing in or creating an account, users feed in various types of data regarding their lifestyles and genetics. This data is analyzed in order to generate predictions and make recommendations using the data visualization techniques such as graphs and charts. Using the machine learning algorithm, analysis of the data is done by the computer in order to predict possible health outcomes based on past data.

C. Class Diagram of the Proposed System

The following section contains the class diagram of the proposed system. The class diagram in the Unified Modelling Language (UML) is one of the UML diagrams used to describe the architectural structure of a software system through defining and arranging the classes within it, their attributes, the methods, and the relationships between those objects in the system. The following figure illustrates the class diagram of the proposed system.

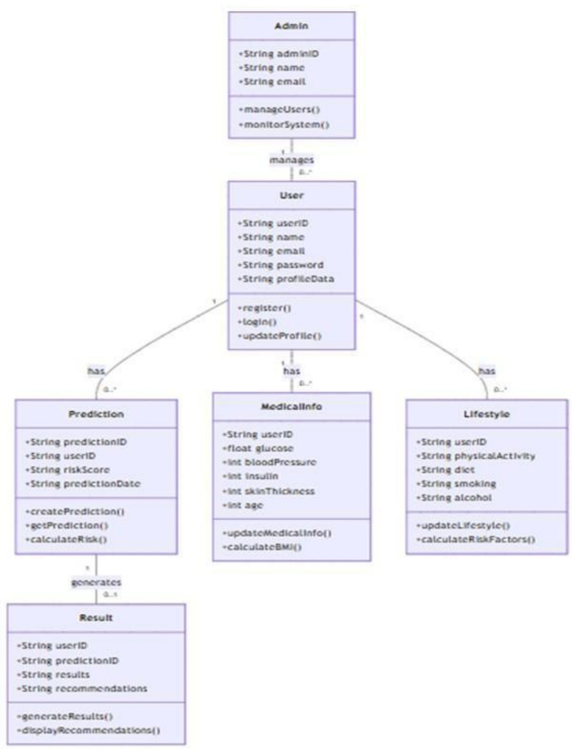


Figure 3.4 Class Diagram of the Proposed System

These describe the structure of a system software by mapping out its classes, attributes, methods, and object relationships. Figure 3.4 presents an illustration of the class diagram for the system

D. Sequence Diagram of Proposed System

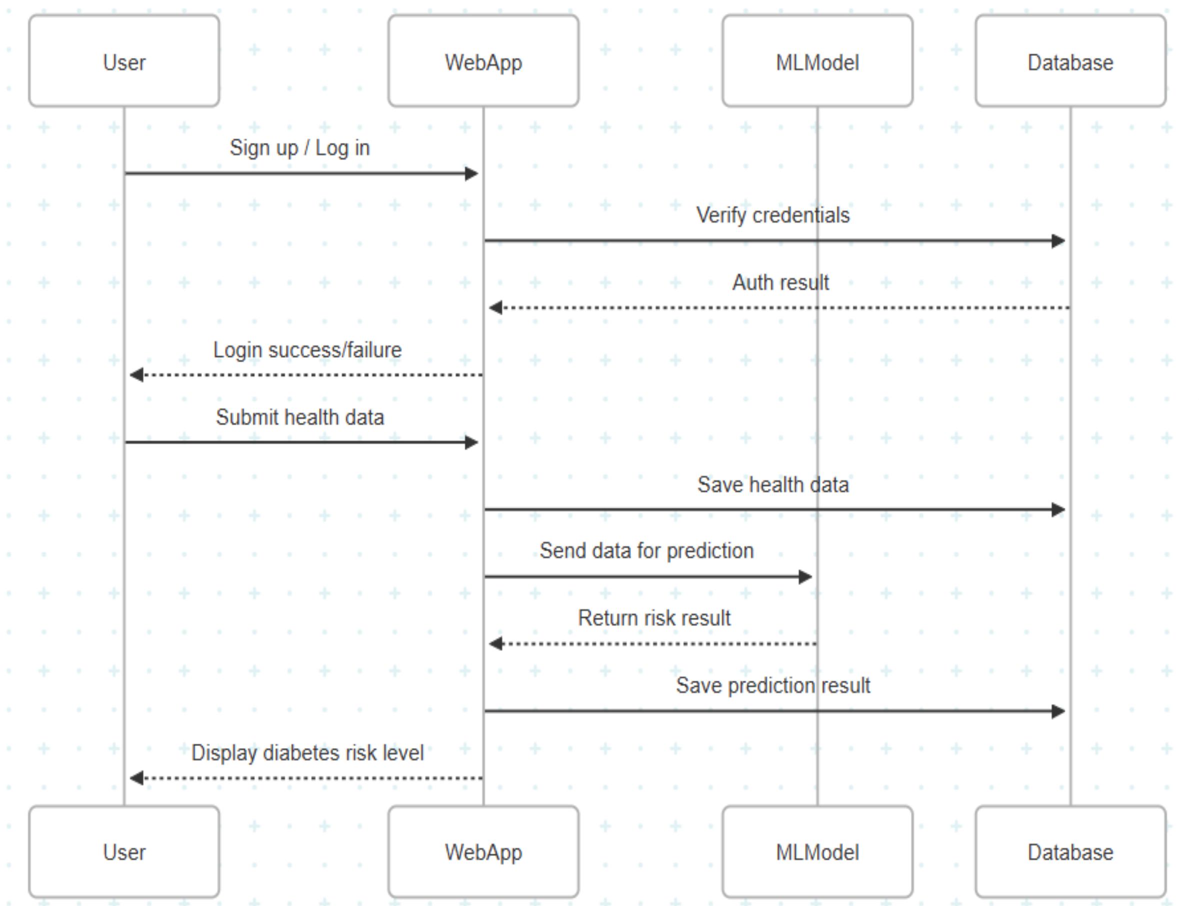


Figure 3.5 Sequence diagram of the proposed system

Figure 4.8 shows the Sequence Diagram of the Diabetes Risk Prediction System. This diagram illustrates the interaction between the user, the system, and the machine learning model throughout the diabetes risk prediction process. The sequence starts with the user initiating the assessment by providing the necessary medical information (e.g., blood glucose levels, BMI, age) and lifestyle data (e.g., physical activity, diet, smoking habits). The system receives this input and sends it to the backend for data preprocessing. The machine learning model processes the preprocessed data, extracting the most relevant features such as glucose levels, BMI, and genetic risk, to calculate the diabetes risk score. After prediction, the system generates the risk score and personalized recommendations for the user. The results are then displayed to the user on the result page**,** and the system can also store these results in the user's history for future reference.

3.5.4 Database Design of the Proposed System

Table 3.2: User

Attribute	**Data Type**	**Description**
UserID	INT	Unique identifier for each user.
FullName	VARCHAR	Stores the user's full name.
Email	VARCHAR	Stores the user's email address.
PasswordHash	VARCHAR	Stores the encrypted user password.
Gender	VARCHAR	Stores the gender of the user.
DateOfBirth	DATE	Stores the user's date of birth.
Role	VARCHAR	Defines user roles such as patient, clinician, or administrator.
CreatedAt	TIMESTAMP	Stores account creation date and time.

Table 3.3: Health Assessment

Attribute	**Data Type**	**Description**
AssessmentID	INT	Unique identifier for each health assessment.
UserID	INT	References the related user account.
GlucoseLevel	DECIMAL	Stores blood glucose measurement.
BMI	DECIMAL	Stores body mass index value.

BloodPressure	VARCHAR	Stores blood pressure reading.
PhysicalActivity	VARCHAR	Stores physical activity status.
FamilyHistory	VARCHAR	Indicates family history of diabetes.
AssessmentDate	TIMESTAMP	Stores date and time of assessment.

Table 3.4: Risk Prediction

Attribute	**Data Type**	**Description**
PredictionID	INT	Unique identifier for each prediction result.
AssessmentID	INT	References the related assessment record.
RiskProbability	DECIMAL	Stores predicted diabetes risk probability.
RiskLevel	VARCHAR	Stores classified risk level.
ModelVersion	VARCHAR	Stores the hybrid deep learning model version used.
PredictionDate	TIMESTAMP	Stores date and time of prediction generation.

Table 3.5: Explainability Report

Attribute	**Data Type**	**Description**
ReportID	INT	Unique identifier for each explainability report.
PredictionID	INT	References the related prediction result.
KeyRiskFactors	TEXT	Stores important features influencing prediction.
SHAPSummary	TEXT	Stores SHAP-based explanation summary.

Recommendation	TEXT	Stores personalised health recommendations.
GeneratedAt	TIMESTAMP	Stores report the generation date and time.

Table 3.6: System Log

Attribute	**Data Type**	**Description**
LogID	INT	Unique identifier for each system log entry.
UserID	INT	References the related user account.
Activity	VARCHAR	Store activity performed within the system.
LogStatus	VARCHAR	Stores the activity execution status.
LogDate	TIMESTAMP	Stores the date and time of logged activity.

CHAPTER FOUR

SYSTEM IMPLEMENTATION, RESULTS AND DISCUSSION

4.1 Introduction

This section presents the implementation, evaluation, and discussion of the explainable mobile health decision-support system developed for diabetes risk assessment by detailing the procedures followed during dataset preparation, model training, optimisation, evaluation, explainability integration, and deployment preparation. The study implemented different machine learning, deep learning, and hybrid deep learning models, which were comparatively evaluated using structured performance metrics in order to adopt the best in the context of diabetes prediction and future scalability.

The study implemented different machine learning, deep learning, and hybrid deep learning models, which were comparatively evaluated using structured performance metrics in order to adopt the best in the context of diabetes prediction and future scalability. Explainable AI techniques were integrated into the proposed system in order to improve prediction transparency and interpretability. Furthermore, the study discusses the generated results, system performance, and practical implications of the developed mobile health framework.

4.2 Development Environment and Tools

4.2.1 Programming Languages and Libraries

The Python programming language was used in this study for the entire implementation of the proposed system due to its flexibility and extensive support for machine learning and deep learning applications. Jupyter Notebook integrated within Visual Studio Code was used in the process of model development, experimentation, testing, and result visualisation. The pandas library was used for dataset loading, manipulation, preprocessing, and tabular data analysis, while NumPy was used for numerical computations, matrix operations, and tensor reshaping. The process also involved the use of the Scikit-learn library for preprocessing, train-test splitting, evaluation metrics generation, and baseline machine learning model implementation. TensorFlow and Keras frameworks were used for implementing the ANN, CNN, and Hybrid CNN–Transformer deep learning architectures, while for the regression model, the XGBoost library was utilised for implementing the gradient boosting baseline model used for comparative evaluation.

Furthermore, the Matplotlib library was used for plotting confusion matrices, training curves, SHAP explainability plots, and result visualisations. On the other hand, the SHAP library was integrated into the system in order to provide explainable AI interpretations and feature contribution analysis. We also used the Joblib library for saving preprocessing objects, feature names, and deployment metadata for future deployment integration.

4.2.2 Development Platform and Hardware Requirements

The system was implemented on a computer system with sufficient RAM and processing capability to support deep learning operations, and the GPU acceleration was utilised where available to improve CNN training speed. However, during the course of the training and implementation, adequate storage capacity was maintained for image dataset handling and model storage. The development platform or integrated development environment employed for this study includes the Visual Studio Code (VSCode), which was used for coding and debugging, the Jupyter, which was extensively used for model training, and finally, the Git for version control, collaboration and deployment.

The hardware requirements for the system implementation are:

i. Minimum 8GB RAM required for smooth processing
ii. Recommended 16GB RAM for deep learning tasks
iii. Intel i5 or higher processor recommended
iv. NVIDIA GPU used for faster model training
v. CUDA support required for GPU acceleration
vi. SSD storage improves data loading speed
vii. High VRAM GPU preferred for diffusion models
viii. Stable internet required for cloud-based training

4.3 System Implementation

The system implementation started with data collection, which prompted us to access the BRFSS Diabetes Health Indicators dataset for the implementation of the proposed system. The dataset contained 70,692 records and 22 attributes, which consist of diabetes risk indicators and patient health conditions. And it was presented in a binary classification problem consisting of diabetic and non-diabetic classes. The dataset was balanced so as to contain equal distributions of diabetic and non-diabetic patient records before the exploratory analysis was conducted to inspect dataset dimensions, feature structure, class distribution, and missing values.

At the conclusion of the exploratory analysis, no missing values were detected during the data validation stage. Immediately, 21 predictor variables were extracted as input features, and the diabetes status column served as the target variable. Some of the important healthcare indicators included BMI, blood pressure, cholesterol, smoking status, physical activity, age, general health, physical health, and income level. The dataset was then divided into training and testing subsets using an 80:20 ratio.

Feature standardisation was performed on the dataset before training using StandardScaler in order to improve training convergence and optimisation stability, in which standardisation transformed the feature values into a uniform scale suitable for deep learning processing. Finally, the preprocessed dataset sample was saved for reproducibility, transparency, and thesis documentation purposes.

For the model implementation, three baseline machine learning models were implemented for comparative performance evaluation, which include Logistic Regression, Random Forest and XGBoost. Logistic Regression was implemented as a traditional statistical classification benchmark model, while the Random Forest was implemented as an ensemble learning model, which is capable of handling nonlinear relationships among healthcare variables. And the XGBoost was implemented as an advanced gradient boosting classifier, which is widely recognised for high predictive performance on structured healthcare datasets. After training, confusion matrices were generated and saved for all baseline models in order to support result visualisation and error analysis, and feature importance analysis was also performed using XGBoost to identify the most influential diabetes risk indicators.

For the proposed Hybrid approach which consist of CNN–Transformer architecture was implemented as the flagship deep learning model of the study. The architecture combined CNN-based feature extraction with Transformer attention-based learning mechanisms, where the CNN layers were initially used to extract meaningful local feature representations from the healthcare indicators and a Multi-Head Attention Transformer block was integrated in order to learn contextual relationships among the healthcare variables. The residual connections were incorporated into the Transformer block of the hybrid architecture in order to improve gradient flow and training stability.

Subsequently, layer normalisation was implemented in order to stabilise attention outputs and improve optimisation consistency. The study used global average pooling in order to reduce parameter explosion and improve generalisation performance. The dense layers and sigmoid output activation were used for final diabetes risk prediction, while the dropout regularisation and EarlyStopping optimisation techniques were integrated during the training epochs in order to reduce overfitting. The EarlyStopping automatically terminated training when validation performance stopped improving while training was going on. The optimised CNN–Transformer model achieved a significantly higher average testing accuracy, which demonstrates stable convergence and reduced validation loss during training.

The architecture of the proposed hybrid CNN-Transformer model showed competitive performance compared with classical machine learning models. Although XGBoost slightly outperformed the hybrid architecture in raw accuracy, the proposed model demonstrated stronger deep representation learning and attention-based feature interaction modelling. The proposed hybrid model architecture further provided better alignment with explainable AI-driven mobile health systems and future multimodal healthcare expansion.

4.3.1 Explainable AI Integration Using SHAP

The study opted for SHAP explainability techniques, which were integrated into the developed system in order to improve prediction transparency and trustworthiness. SHAP analysis was performed on the optimized CNN–Transformer model, and SHAP summary plots were generated in order to identify which and how the features contribute toward diabetes risk predictions. The local explanation plots were further generated in order to explain individual patient prediction outcomes, with the positive and negative feature contributions toward prediction decisions visualised using SHAP force plots.

The BMI, physical health, age, blood pressure, and general health indicators contributed significantly to diabetes risk predictions. Overall, the explainable AI integration improved the interpretability of the proposed diabetes risk prediction system, which will be deployed to the healthcare facilities. The explainability engine supported informed clinical interpretation and transparent decision-making.

4.3.2 Deployment Preparation and Model Serialisation

The deployment preparation procedures were implemented in order to support future mobile health integration, which requires trained models to be serialised and saved in H5 format for reproducibility and deployment consistency. The StandardScaler preprocessing object was saved using Joblib serialisation, while feature names and deployment metadata were preserved for future API integration and prediction consistency. The deployment structure was organised into models and pipeline directories, and a deployment metadata object was created in order to preserve model configuration and preprocessing information. Finally, a sample inference test was performed using the saved optimised CNN–Transformer model, and the deployment pipeline successfully generated diabetes risk predictions from new patient records, confirming the feasibility of integrating the model into a mobile health decision-support environment.

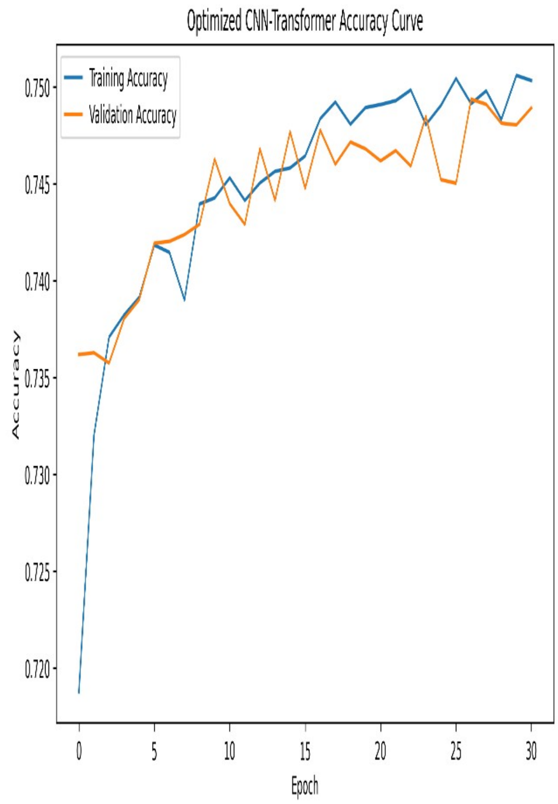
4.4 Result of the Model and Discussion

The comparative evaluation demonstrated that all implemented models achieved competitive diabetes prediction performance, with the XGBoost model having the highest testing accuracy among all evaluated models. Table 4.1 presents the results of the implemented models, while Figures 4.1, 4.2, 4.3, and 4.4 present the confusion matrix, accuracy plot, loss graph, and SHAP plot of the adopted model.

4.4.1 Results of Implemented Models

Table 4.1: Comparative Results of Implemented Models

Model	Accuracy	Precision	Recall	F1-score
XGBoost	75.52	73.23	80.24	76.57
Optimized CNN-Transformer	74.89	72.97	78.82	75.78
Logistic Regression	74.84	73.72	76.98	75.31
Random Forest	73.71	71.72	78.04	74.75



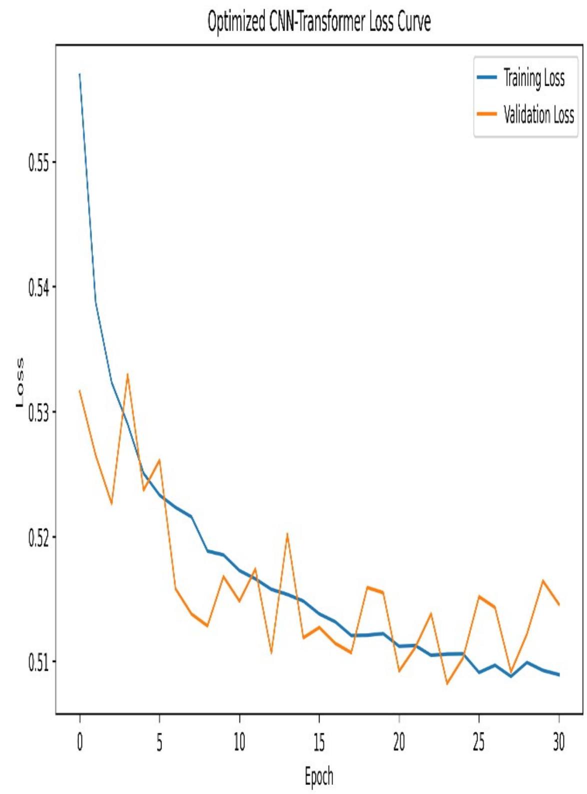


Figure 4.1: Plot of Accuracy **Fig 4.2 Plot of Loss**

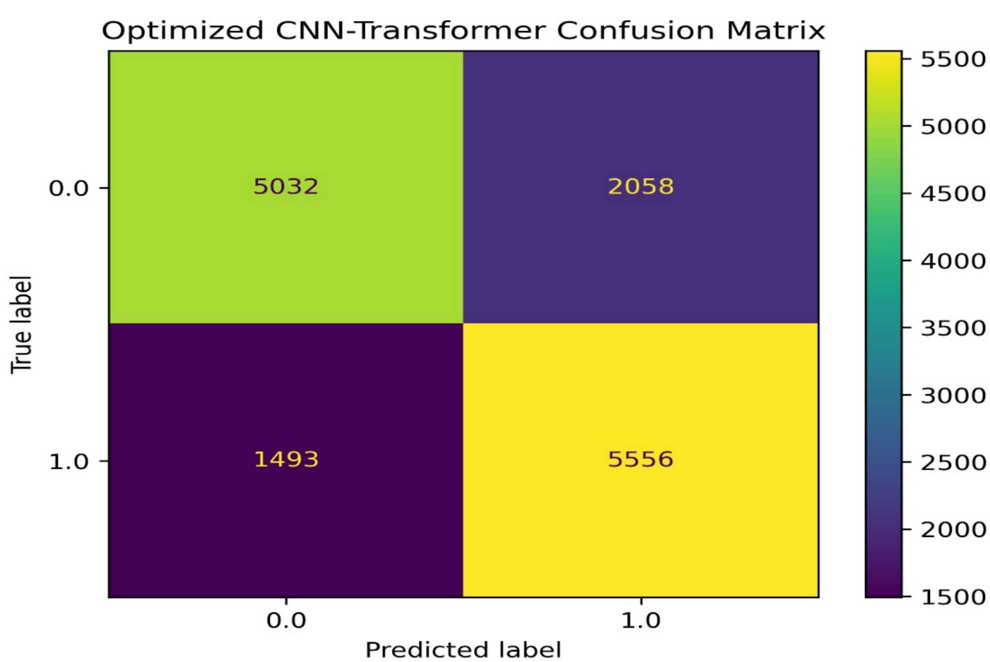


Figure 4.3: Confusion matrix of the Hybrid Model

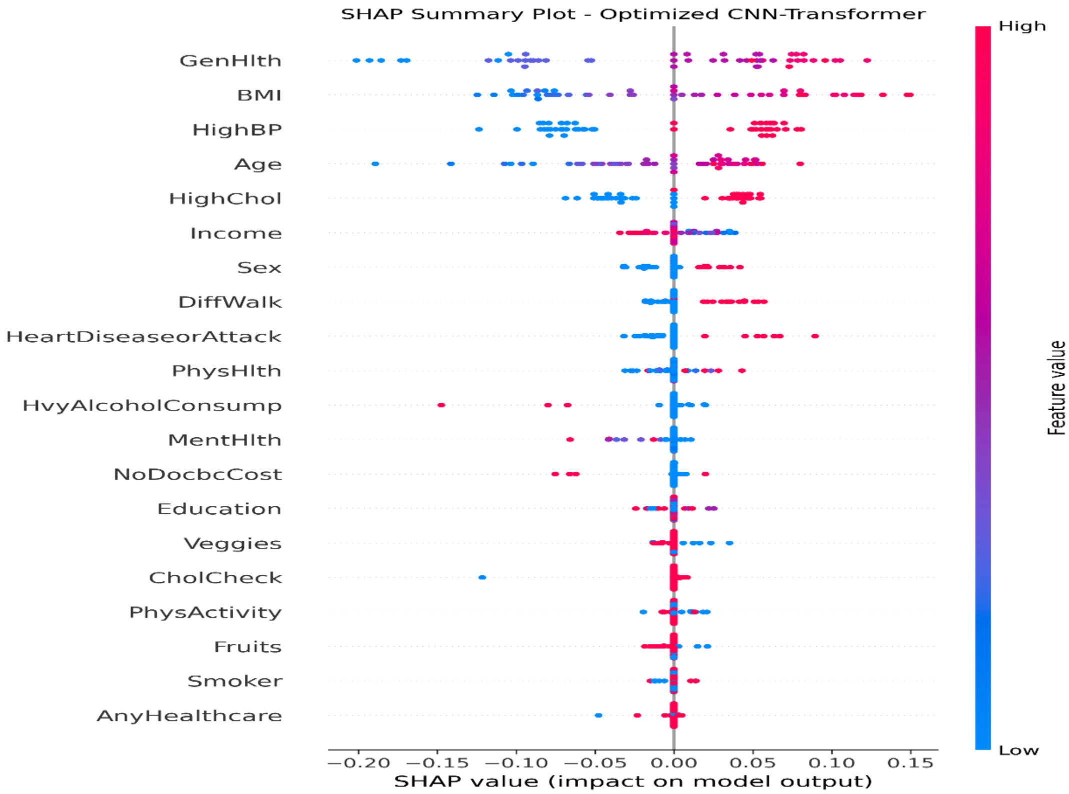


Figure 4.4: SHAP Summary Plot

The accuracy curve in Figure 4.1 shows that both the training and validation accuracies improved steadily across the epochs, which enabled it to attain an approximate accuracy of75%, which indicates that the optimised CNN-Transformer model learned the dataset effectively without unstable fluctuations. Figure 4.2 presents the training and validation loss curves, and observations from the plot show that both decreased gradually and remained very close throughout the training process, which suggests that the model achieved stable learning performance with minimal signs of overfitting or underfitting during its training process. The confusion matrix presented in Figure 4.3 demonstrates that the model correctly classified a large number of samples in both classes, with 5032 and 5556 correctly predicted instances, which indicates that the adopted model is very strong and robust in predicting diabetes risk occurrence and has a balanced classification performance, however, the relatively lower number of misclassified samples (2058 and 1493) compared to the correctly classified cases shows that the model maintained good generalisation ability when tested on unseen data. Finally, Figure 4.4 demonstrates the SHAP summary plot, which describes features such as General Health (GenHlth), BMI, High Blood Pressure (HighBP), Age, and High Cholesterol (HighChol) as features which had the strongest influence on the prediction outcomes of the adopted optimised hybrid model, confirming that the model's decisions were interpretable and aligned with known diabetes risk factors

4.4.2 Discussion of Findings

The proposed optimised CNN–Transformer model demonstrated competitive classification performance, with improved deep feature learning capability, as the Transformer attention mechanism enabled contextual learning among healthcare indicators, indicating that the hybrid architecture has stronger scalability potential for future multimodal healthcare applications.

The explainable AI integration significantly improved model transparency and interpretability, with SHAP explainability outputs revealing the dominant healthcare indicators influencing diabetes risk predictions. The confusion matrices showed that the implemented models achieved balanced sensitivity and specificity performance, while the adopted optimized CNN–Transformer model achieved high recall performance, indicating a strong ability to correctly identify high-risk diabetic patients. The overall results validated the effectiveness of combining deep learning and explainable AI within a mobile health decision-support framework. Overall, the proposed architecture supports future expansion into wearable devices, real-time monitoring systems, and intelligent healthcare recommendation platforms.

4.5 System Interface Description

The study designed the system architecture to support mobile-based diabetes risk assessment and decision support.

A. Landing Page: The home page or the main user interface that displays a button to log in or sign up, as well as information about the application.

Figure 4.5 Home Page

B. Login Page: Where the authorised user has to key in their details in order to fully access the application.

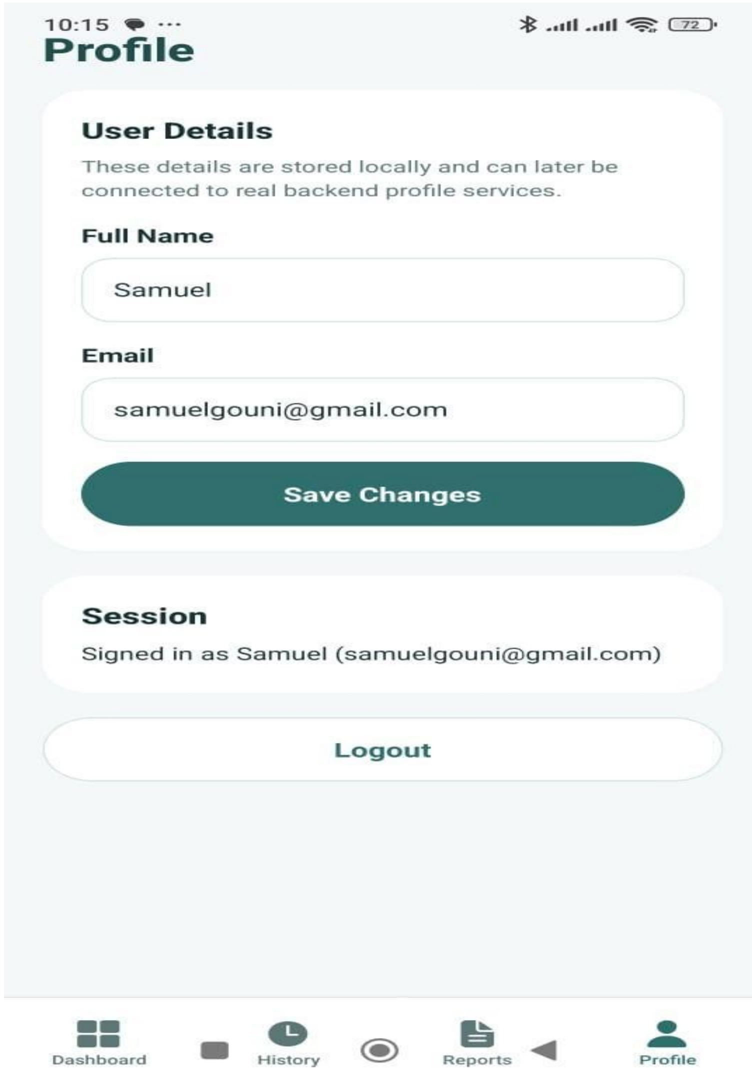


Figure 4.6: Login Page

C. Signup Page: An unauthorised user signs up to be authenticated and granted the credentials to log in.

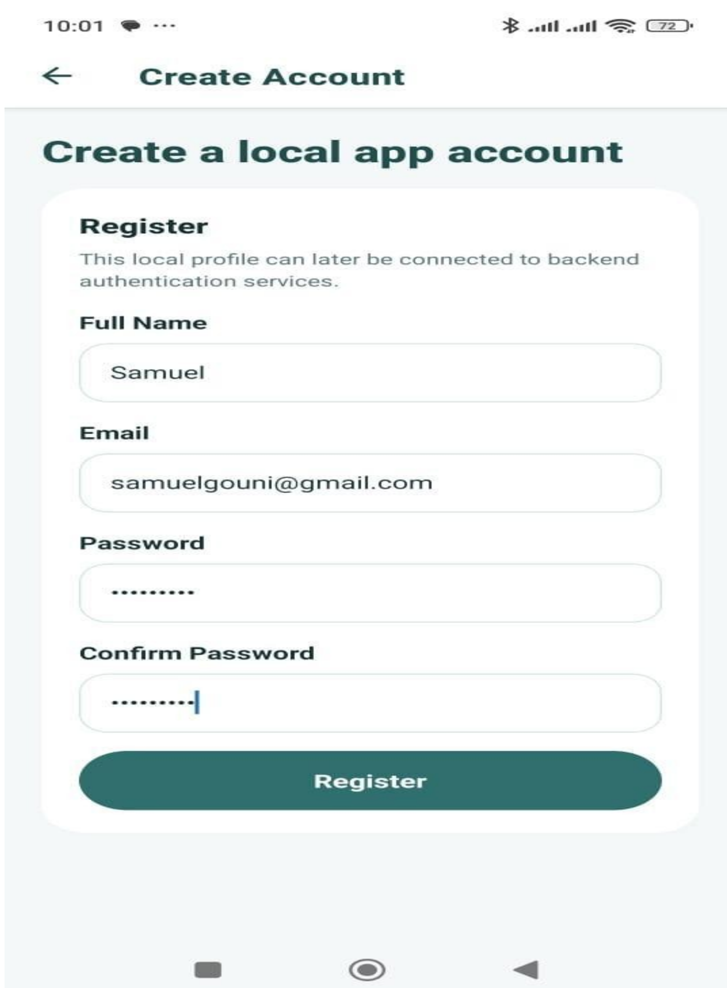


Figure 4.7 Signup page

D. Prediction Input: This is where the user is prompted to key in the data for the system to make its prediction.

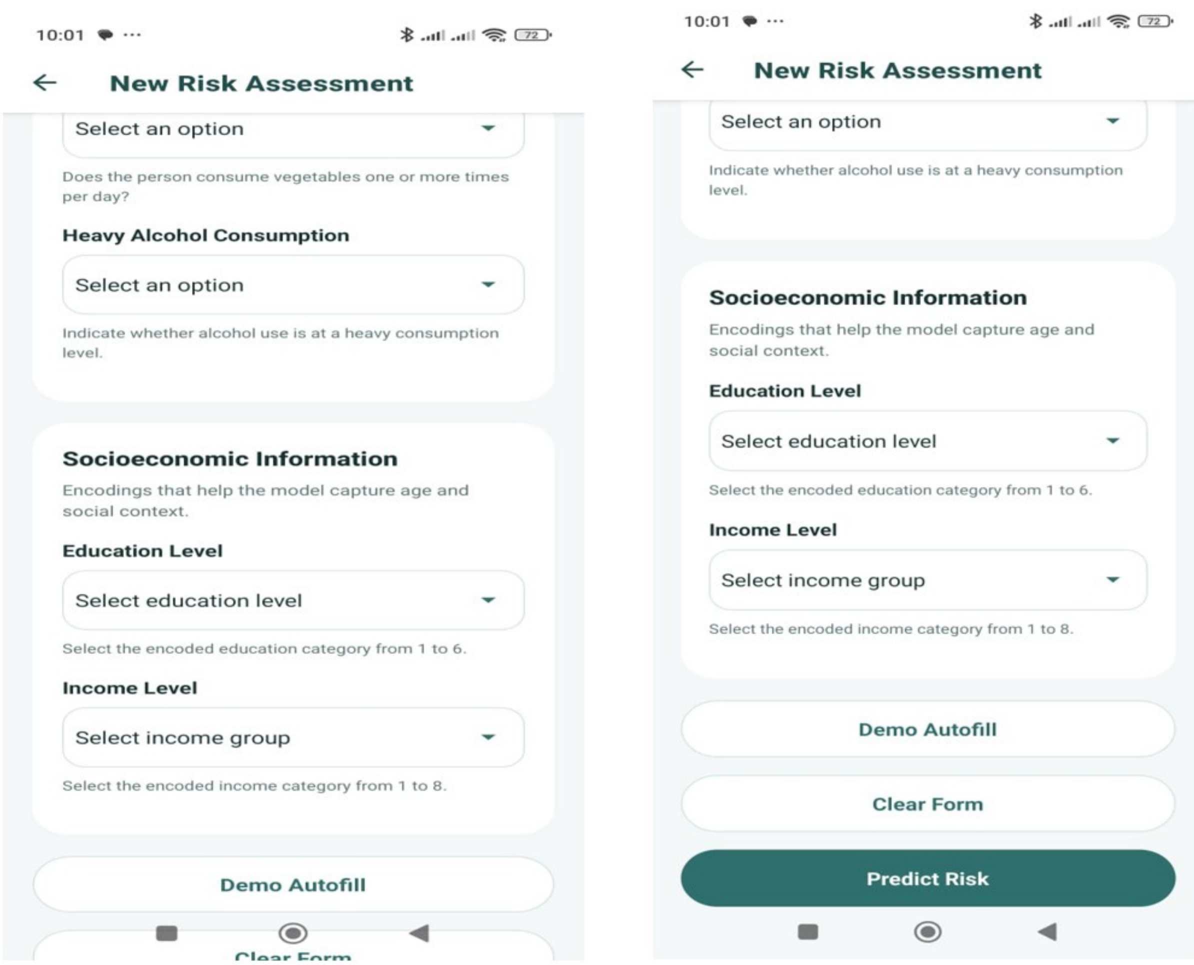


Figure 4.8: Prediction Input Screen

E. Prediction Output: It displays the result of the prediction

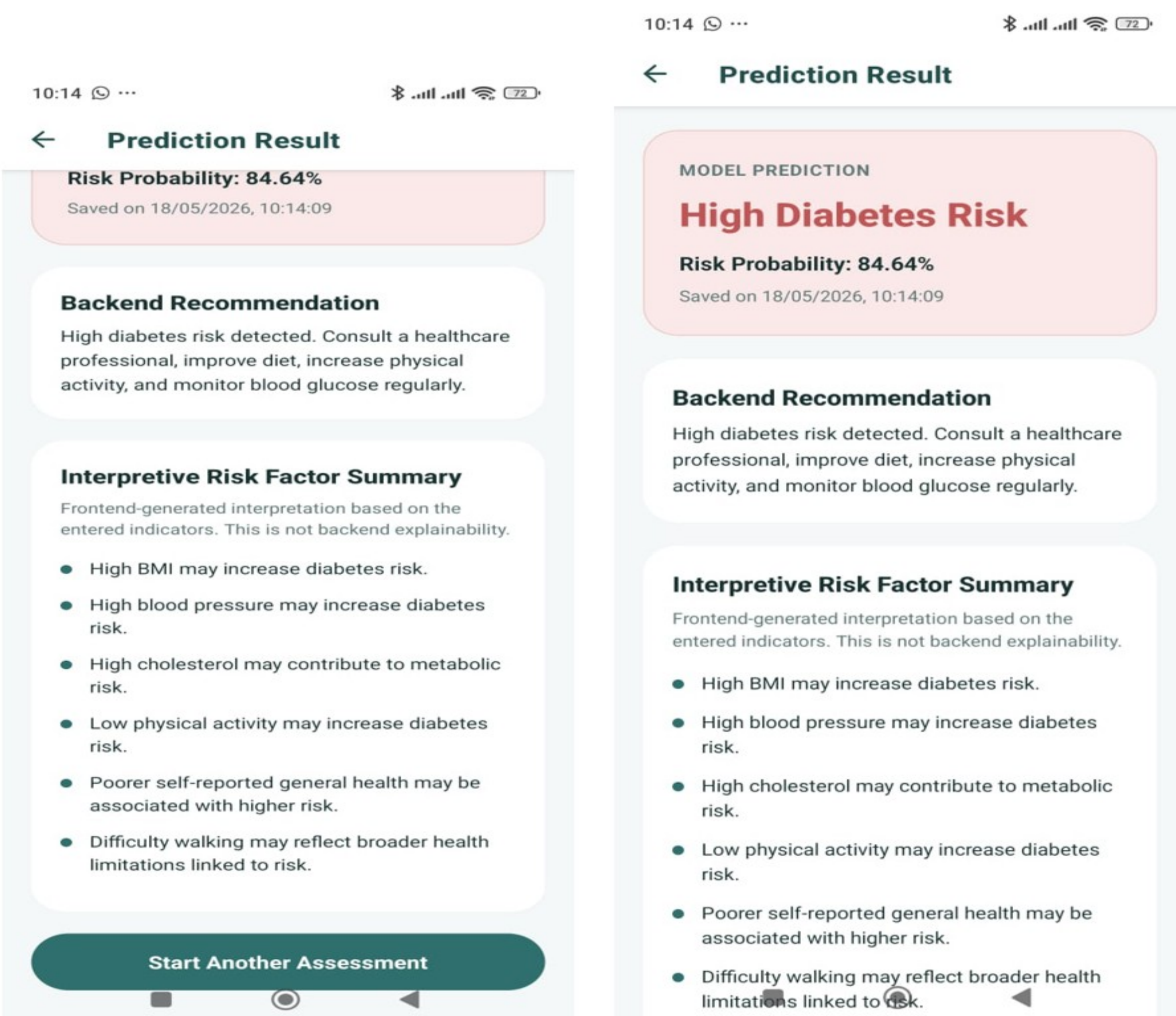


Figure 4.9: Prediction Result

F. Prediction History: This displays the past predictions performed by the system.

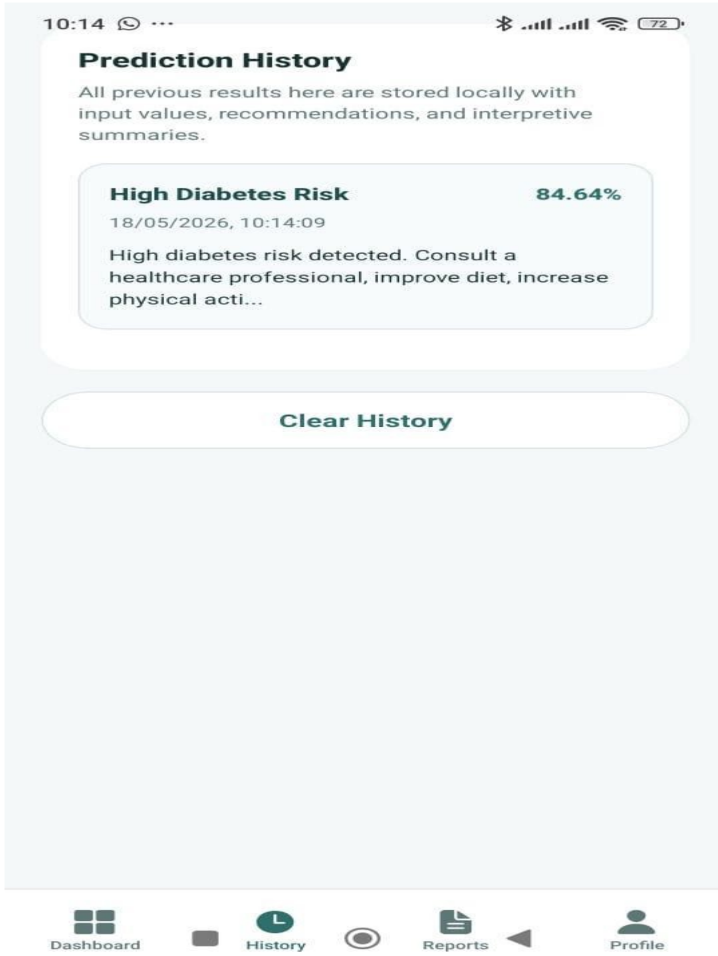


Figure 4.10 Prediction History

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

This study designed and implemented an explainable mobile health decision-support system for diabetes risk assessment using hybrid deep learning techniques, and also employed the BRFSS Diabetes Health Indicators dataset from the CDC for the model training and evaluation. The data preprocessing operations performed pre-model training, including feature scaling, train-test splitting, and standardisation, were successfully implemented. We selected multiple baseline machine learning models, including Logistic Regression, Random Forest, and XGBoost, which were implemented and comparatively evaluated against the proposed Hybrid CNN–Transformer models, which were further developed and tested. The proposed hybrid model was optimised to integrate attention mechanisms, residual connections, normalisation, and explainable AI capabilities.

The SHAP explainability analysis was vital as the study, being a health-related topic, demands full prediction transparency and feature interpretation, which the integration of the XAI module SHAP helped improve. The experimental evaluation demonstrated competitive predictive performance across all implemented models, with XGBoost achieving the highest testing accuracy, while the proposed hybrid architecture exhibited stronger deep representation learning and explainability, making it the final choice and then, the deployment preparation and model serialisation procedures followed up, which were successfully completed for future mobile health integration. The study demonstrated the feasibility of integrating hybrid deep learning and explainable AI within a real-time mobile healthcare framework.

5.2 Conclusion

The study has successfully developed an explainable mobile health decision-support system, which has the capability of predicting diabetes risk using healthcare indicators and hybrid deep learning techniques, through the integration of CNN and Transformer architectures, which enabled effective feature learning and contextual relationship modelling among healthcare variables. An explainable AI was integrated in order to improve the system's interpretability and transparency, thereby increasing the trustworthiness of prediction outputs.

A comparative conducted earlier during the implementation showed that classical machine learning and deep learning approaches can both achieve strong diabetes prediction performance on structured healthcare datasets after careful analysis. The proposed hybrid architecture demonstrated strong scalability potential for future intelligent healthcare applications, as it confirms that mobile health decision-support systems can support early diabetes detection and preventive healthcare intervention. Overall, the research achieved its aim and objectives by designing an explainable hybrid deep learning system suitable for real-time diabetes risk assessment and future mobile health deployment.

5.3 Recommendations

The study recommends that future studies should explore the use of larger and more diverse healthcare datasets in order to improve model generalisation capability, and additional healthcare modalities, such as wearable sensor data, ECG signals, and glucose monitoring streams, should be incorporated into future system extensions.

i. Future studies should employ or explore the use of advanced Transformer architectures and hyperparameter optimisation techniques to further improve predictive performance.

ii. Federated learning and privacy-preserving AI techniques should be explored in future healthcare deployment scenarios.

iii. The study recommends the Integration of cloud-based healthcare infrastructure, which may improve scalability and accessibility of the system.

iv. Future systems should incorporate stronger authentication and healthcare data protection mechanisms in order to secure patient information management.

References

Abbas, M., Alhasan, A., & Hamza, U. M. (2015). Perceived ease of use and utilisation of e-learning technologies by academic staff in Federal College of Education, Zaria.

Abisoye, O. A., Boboye, F., & Abisoye, B. O. (2013). Design of a computer-based inventory system for supermarkets. *International Journal of Computing and Technology*, 2(2), 45–50. https://www.ijctjournal.org/Volume2/Issue2/IJCT-V2I2P6.pdf

Ahmed, S. K., Naji, Z. H., Hatif, Y. N., & Hussam, M. (2020). Design and implementation of a drug inventory system. *Diyala Journal of Engineering Sciences*, 13(2), 80–90. https://www.iasj.net/iasj/article/192207

Althabatah, A., Yaqot, M., Menezes, B., & Kerbache, L. (2023). Transformative Procurement Trends: Integrating Industry 4.0 Technologies for Enhanced Procurement Processes. *Logistics*, 7(3), 63. https://doi.org/10.3390/logistics7030063

Buseri, T. F., Aremu, T. O., & Suleiman, I. A. (2023). Knowledge and practice of supply chain management of pharmaceuticals by health professionals in four states of South-South, Nigeria. *African Journal of Research and Pharmaceutical Development*, 15(2), 91–100.

Chen, J. (2025, May 31). Industrial Revolution: Definition, history, pros, and cons. *Investopedia*. https://www.investopedia.com/terms/i/industrial-revolution.asp

Chopra, S., & Meindl, P. (2019). *Supply chain management: Strategy, planning, and operation* (7th ed.). Pearson. https://www.pearson.com/store/p/supply-chain-management-strategy-planning-and-operation/P100001039313

Christopher, M. (2016). *Logistics & supply chain management* (5th ed.). Pearson Education. https://www.pearson.com/store/p/logistics-and-supply-chain-management/P100001006753

Coelho, L. C., & Laporte, G. (2015). An overview of vendor-managed inventory: Coordination and inventory management. *International Journal of Production Research*, 53(17), 5223–5246. https://doi.org/10.1080/00207543.2015.1033032

`Du, M., Luo, J., Wang, S., & Liu, S. (2020). Inventory control optimization with genetic algorithm and BP neural network. *Neural Computing and Applications*, 32, 1981–1994. https://doi.org/10.1007/s00521-018-3687-1

Felea, M., & Albăstroiu, I. (2013). Defining the concept of supply chain management and its relevance to Romanian academics and practitioners. *Amfiteatru Economic Journal*, 15(33), 74–88. https://www.amfiteatrueconomic.ro/RevistaDetalii_EN.aspx?Cod=240

Holm, M. R., Rudis, M. I., & Wilson, J. W. (2015). Medication supply chain management through the implementation of a hospital pharmacy computerized inventory program in Haiti. *American Journal of Health-System Pharmacy*, 72(17), 1450–1453. https://doi.org/10.2146/ajhp140671

Kritchanchai, D., & Meesamut, W. (2015). Inventory management in a hospital using demand categorization and resource allocation. *International Journal of Supply Chain Management*, 4(2), 11–19. https://ojs.excelingtech.co.uk/index.php/IJSCM/article/view/1355

Linda, D. M., Iheanyi, N. R., Chika, O., & Yaya, A. (2016). Development of standalone inventory management system for consumer goods. Unpublished manuscript, Federal College of Forestry, Jos, Plateau State, Nigeria.

Mao, Y., Zhang, Y., & Zhai, S. (2008). Mobile text messaging for pharmaceutical care in a hospital in China. *Telemedicine and e-Health*, 14(4), 412–417. https://doi.org/10.1089/tmj.2007.0081

Medical Packaging. (n.d.). Pharmaceutical supply chain management: Key steps & challenges. Retrieved June 12, 2025, from https://medpak.com/pharmaceutical-supply-chain-management/

Mu, Y., Boussemart, J.-P., Shen, Z., & Vardanyan, M. (2025). Sustainable supply chains with socially undesirable and intermediate outputs: Evidence from Chinese agricultural cities. *Annals of Operations Research*. Advance online publication. https://doi.org/10.1007/s10479-025-06658-w

Okeke, C. O., Uzochukwu, I. C., & Okoye, M. I. (2016). Online pharmaceutical management system. *European Scientific Journal*, 12(12), 139–150. https://doi.org/10.19044/esj.2016.v12n12p139

Onuiri, E. E., Oyebanji, I. G., Fayehun, S. A., & Chukwujioke, S.-D. (2016). Online pharmaceutical management system. *European Scientific Journal*, 12(12), 139–150. https://doi.org/10.19044/esj.2016.v12n12p139

Salih, H. S., Ghazi, M., & Aljanabi, M. (2023). An automated inventory management system for SMEs. *Iraqi Journal for Computer Science and Mathematics*, 4(2), 238–244. https://journal.ijcsm.org/index.php/ijcsm/article/view/219

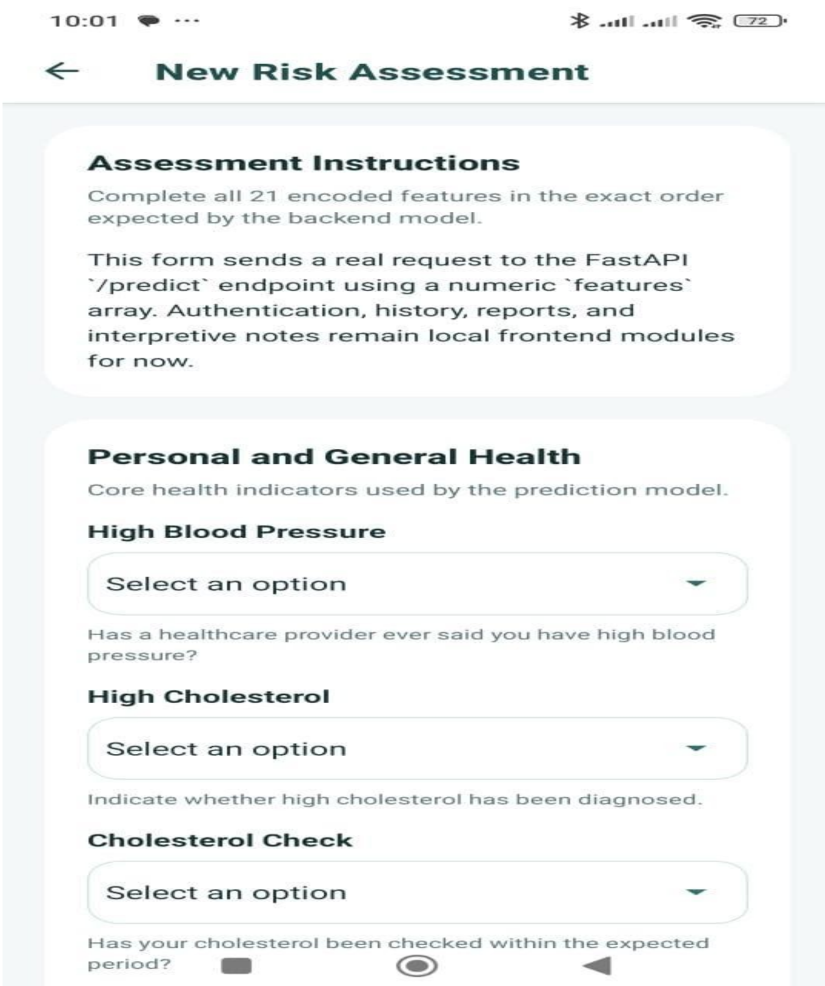
Sukasih, E., Apriyanto, G., & Firdiansjah, A. (2020). Drug inventory management from a financial perspective in the ABC hospital pharmacy. *IOSR Journal of Business and Management*, 22(8), 54–61. https://www.iosrjournals.org/iosr-jbm/papers/Vol22-issue8/Series-5/H2208055461.pdf

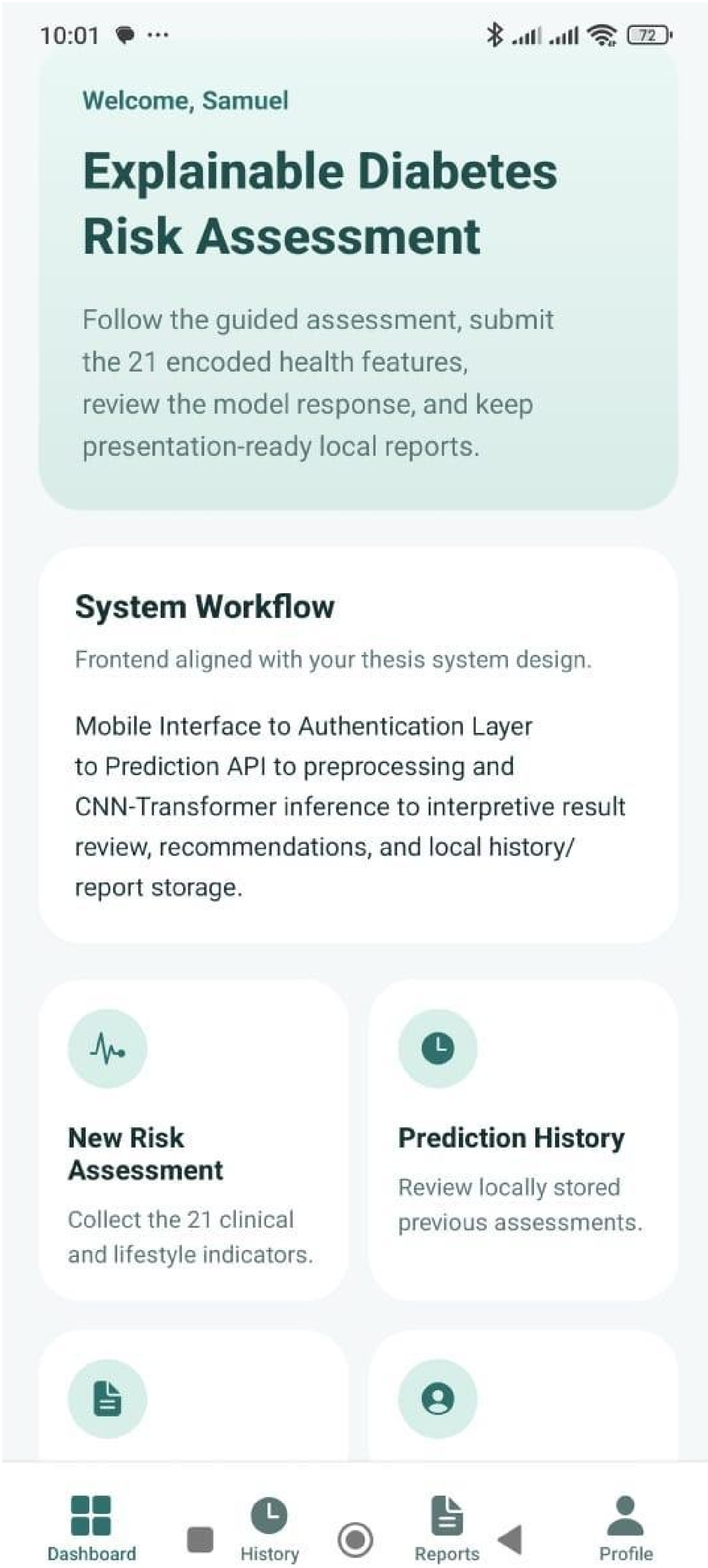
Toyota Motor Corporation. (2020). Toyota production system overview. *Toyota Global*. https://global.toyota/en/company/vision-and-philosophy/production-system/

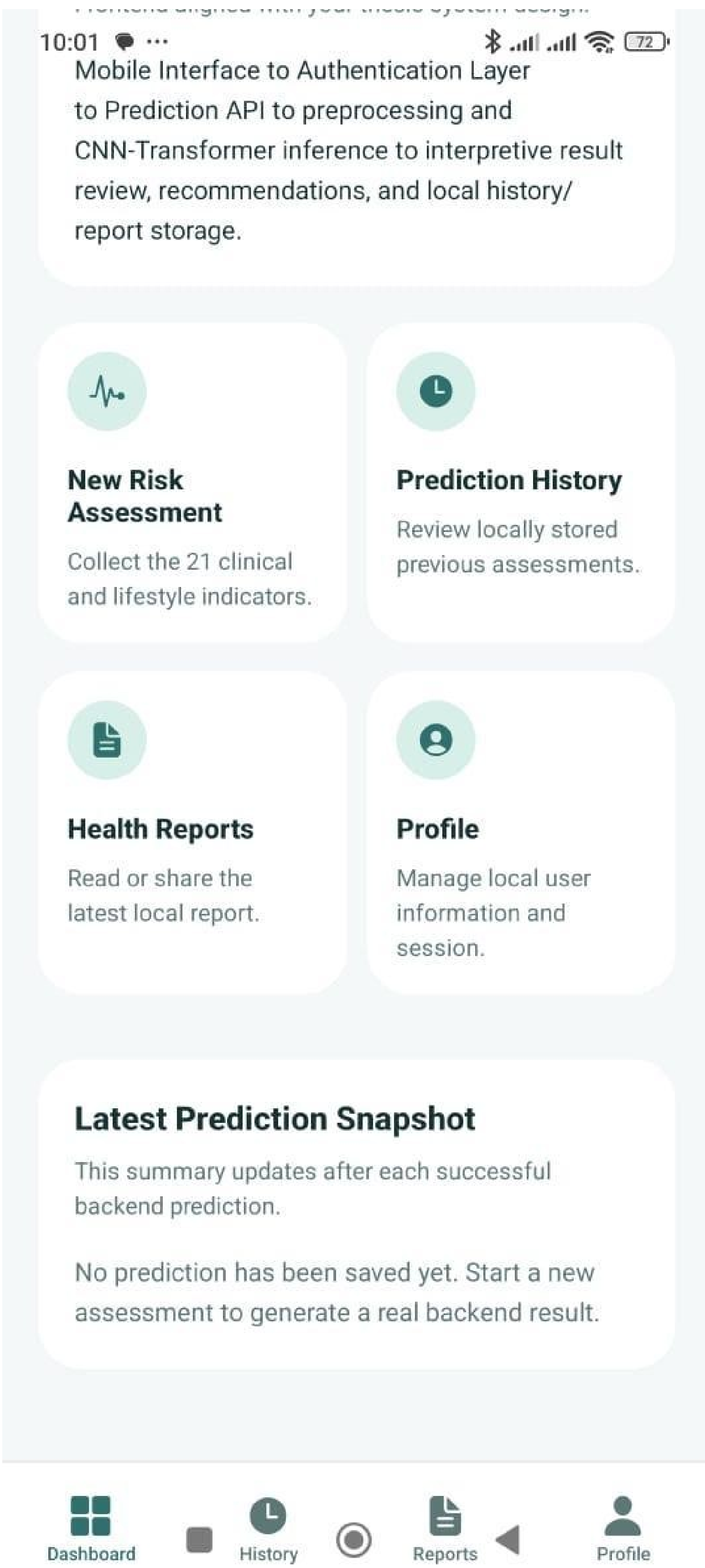
Uğrak, U., Çınar, Ş., Seyhan, F., & Kavak, D. (2025). Evaluation of the supply process in public hospitals: A qualitative study. *Verimlilik Dergisi*, 59(1), 101–114. https://doi.org/10.51551/verimlilik.1495743

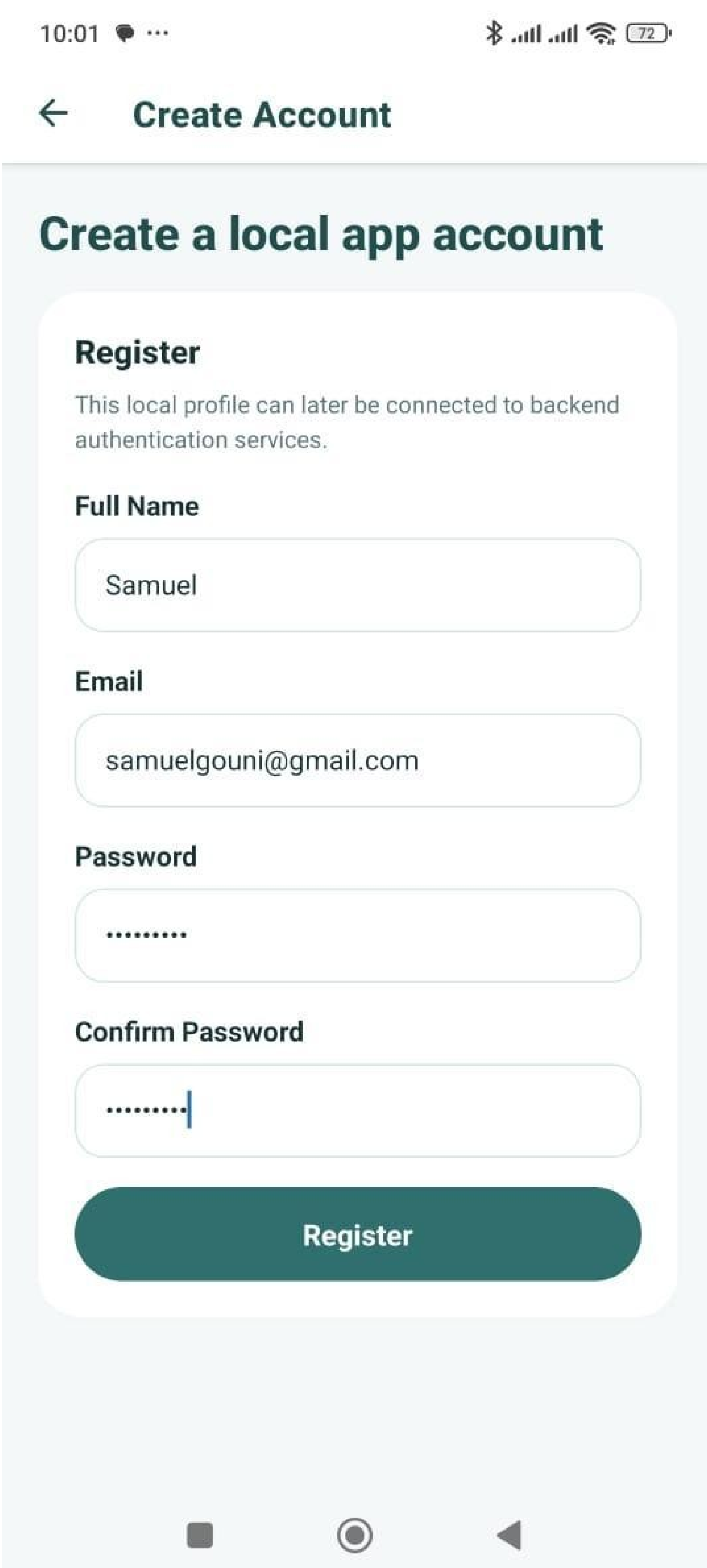
Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77–84. https://doi.org/10.1111/jbl.12010

Winters, B. H., & Hernandez, L. (1972). Automated drug inventory control system. *American Journal of Health-System Pharmacy*, 29(9), 780–785.


10:01
New Risk Assessment
Assessment Instructions
Complete all 21 encoded features in the exact order expected by the backend model.
This form sends a real request to the FastAPI `/predict` endpoint using a numeric `features` array. Authentication, history, reports, and interpretive notes remain local frontend modules for now.
Personal and General Health
Core health indicators used by the prediction model.
High Blood Pressure
Select an option
Has a healthcare provider ever said you have high blood pressure?
High Cholesterol
Select an option
Indicate whether high cholesterol has been diagnosed.
Cholesterol Check
Select an option
Has your cholesterol been checked within the expected period?


10:01
Welcome, Samuel
Explainable Diabetes Risk Assessment
Follow the guided assessment, submit the 21 encoded health features, review the model response, and keep presentation-ready local reports.
System Workflow
Frontend aligned with your thesis system design.
Mobile Interface to Authentication Layer to Prediction API to preprocessing and CNN-Transformer inference to interpretive result review, recommendations, and local history/ report storage.
New Risk Assessment
Collect the 21 clinical and lifestyle indicators.
Prediction History
Review locally stored previous assessments.
Dashboard
History
Reports
Profile


10:01
Mobile Interface to Authentication Layer to Prediction API to preprocessing and CNN-Transformer inference to interpretive result review, recommendations, and local history/ report storage.
New Risk Assessment
Collect the 21 clinical and lifestyle indicators.
Prediction History
Review locally stored previous assessments.
Health Reports
Read or share the latest local report.
Profile
Manage local user information and session.
Latest Prediction Snapshot
This summary updates after each successful backend prediction.
No prediction has been saved yet. Start a new assessment to generate a real backend result.
Dashboard
History
Reports
Profile


10:01
Create Account
Create a local app account
Register
This local profile can later be connected to backend authentication services.
Full Name
Samuel
Email
samuelgouni@gmail.com
Password
Confirm Password
Register