



**INFORMATICS  
INSTITUTE OF  
TECHNOLOGY**

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

**DiabetCare - An Artificial Intelligence based Mobile Application  
for Predicting the Risk of Prediabetes Mellitus, Gestational  
Diabetes Mellitus and Type 2 Diabetes Mellitus**

A Dissertation by

Ms. Nirmani Pathirana

Supervised by

Mr. Asanka Suraweera

Submitted in partial fulfillment of the requirements for the MSc in Advanced Software  
Engineering degree at the University of Westminster

**May 2023**

## **ABSTRACT**

Diabetes mellitus, simply known as diabetes, is a crucial public health issue worldwide. Ongoing global estimates of International Diabetes Federation states that, while 415 million of population are affected by this condition, it is set to rocket up to 642 million by 2040. It is a long-term illness which hinders the body's capability to absorb sugar/glucose. There are 3 commonly known diabetes mellitus types: prediabetes, type 2 diabetes, and gestational diabetes. Treatments are specific and varied for each diabetes type. The existence of glucose in the blood flow can rise and progress to severe stages if the specific diabetes type is not predicted and treated earlier. For diagnosis and to prevent diabetes mellitus, early risk prediction is critical.

The purpose of this study is to introduce an innovative way to early predict the risk of 3 different types of diabetes categorically with an artificial intelligence approach, so they can be confirmed through blood tests and can be treated as early as possible to avoid future complications. Medical datasets from University of Kelaniya have been used to apply machine learning in a more vigorous and acknowledgeable approach. This inspection will be based upon implementing 2 separate machine learning-based ensemble learning approaches and will be presented to the users in the face of an automated mobile application named "DiabetCare", where users may predict the risk of having a specific diabetes type earlier without age limit.

Experiments were carried out by performing model training with a set of 12 machine learning base classifiers comparatively, along with 3 ensemble techniques for each component. Best parameters were chosen for each algorithm through random search cv hyper-parameter tuning technique. Finally, a base model and an ensemble model with the best evaluation metrics were chosen to be deployed in the final product. For type 2/prediabetes risk prediction component, Random Forest classifier with a training accuracy of 1.0000, testing accuracy of 0.8715, AUC-ROC score of 0.9811 and Cohen's Kappa of 0.8037 was chosen, while for gestational diabetes risk prediction component, Stacking classifier with a training accuracy of 1.0000, testing accuracy of 0.9650, AUC-ROC score of 0.9963 and Cohen's Kappa of 0.9288 was chosen. Implementing such a system will honestly contribute to the diabetes medical domain and also will be capable of being used as a verification mechanism for medical professionals.

**Keywords:** *Type 2 Diabetes Mellitus, Prediabetes Mellitus, Gestational Diabetes Mellitus, Risk Prediction, Machine Learning, Ensemble Stacks*