BGEM™: Assessing Elevated Blood Glucose Levels Using Machine Learning and Wearable Photoplethysmography Sensors

Bohan Shi [1,2,*], Satvinder Singh Dhaliwal [3,4,5,6,7], Marcus Soo[1,2], Cheri Chan[8], Jocelin Wong[1,2], Natalie W.C. Lam[1,2], Entong Zhou[1,2], Vivien Paitimusa[1,2], Kum Yin Loke[1,2], Joel Chin[1,2], Mei Tuan Chua[8], Kathy Liaw Chiew Suan[8], Amos WH Lim [1,2], Fadil Fatin Insyirah[8], Shih-Cheng Yen[9], Arthur Tay [10], and Seng Bin Ang[11,12]

[1]Actxa Pte. Ltd., Singapore
[2]Activate Interactive Pte. Ltd., Singapore
[3]Curtin Health Innovation Research Institute, Curtin University, Australia
[4]Faculty of Health Sciences, Curtin University, Australia
[5]Duke NUS Graduate Medical School, National University of Singapore, Singapore
[6]Singapore University of Social Sciences, Singapore
[7]Institute for Research in Molecular Medicine, Universiti Sains Malaysia, Malaysia
[8]KK Women's and Children's Hospital, Singapore
[9]Innovation and Design Programme, Faculty of Engineering, National University of Singapore
[10]Department of Electrical and Computer Engineering, National University of Singapore
[11]Family Medicine Academic Clinical Program, Duke-NUS Medical School, Singapore
[12]Menopause Unit, KK Women's and Children's Hospital, Singapore

Corresponding Author:
Bohan Shi, PhD
[*]bohan.shi@actxa.com

Abstract

Background: Diabetes mellitus (DM) is the most challenging and fastest-growing global public health challenge. An estimated 10.5% of the global adult population suffers from diabetes, and almost half of them are undiagnosed. The growing at-risk population exacerbated the shortage of health resources, with an estimated 10.6% and 6.2% of adults worldwide having impaired glucose tolerance (IGT) and impaired fasting glycemia (IFG) respectively. All the current diabetes screening methods are invasive and opportunistic and must be conducted in a hospital or a laboratory by trained professionals. At-risk subjects might remain undetected for years and miss the precious time window for early intervention to prevent or delay the onset of diabetes and its complications.

Objective: We aimed to develop an AI solution to recognise elevated blood glucose levels ($\geqslant 7.8 mmol/L$) non-invasively and evaluate diabetic risk based on repeated measurements.

Methods: This study was conducted at KK Women's and Children's Hospital of Singapore, and five hundred (n=500) participants were recruited (mean age 38.73 ± 10.61 years; mean BMI 24.4 ± 5.1 kg/m2). The blood glucose levels for most participants were measured before and after 75g of sugary drink using both the conventional glucometer (Accu-Chek Performa) and the wrist-worn wearable. The results obtained from the glucometer were used as the ground truth measurements. We performed extensive feature engineering on the photoplethysmography (PPG) sensor data and identified features sensitive to glucose changes. These selected features were further analysed using an explainable AI approach to understand their contribution to our predictions.

Results: Multiple machine learning models were trained and assessed with 10-fold cross-validation using subject demographic data and critical features extracted from the PPG measurements as predictors. Support vector machine (SVM) with a radial basis function (RBF) kernel had the best detection performance with an average accuracy of 84.7%, a sensitivity of 81.05%, a specificity of 88.3%, a precision of 87.51%, a geometric mean of 84.54% and F-score of 84.03%.
Conclusions: Our findings suggested PPG measurements can be utilized to identify subjects with elevated blood glucose measurements and assist in the screening of subjects for diabetes risk.
Trial Registration: ClinicalTrials.gov NCT05504096;
https://clinicaltrials.gov/ct2/show/NCT05504096

Keywords: Diabetes mellitus; explainable AI; feature engineering; machine learning; photoplethysmography; wearable sensor.

Introduction

Diabetes mellitus (DM) is a chronic and heterogeneous metabolic disorder characterised by the presence of hyperglycemia due to deterioration of insulin secretion, defective insulin action, or both [1,2]. There are three main types of DM: type-1 DM (T1DM), type-2 DM (T2DM) and gestational diabetes. T2DM is the most prevalent type of diabetes, affecting over 95% of people with diabetes worldwide [3,4].
The prevalence of DM has been proliferating in recent decades, and it is now the most prominent and fastest-growing global public health challenge [5,6]. Uncontrolled diabetes is associated with an increased risk of complications such as cardiovascular disease, kidney failure, vision loss, nerve damage, and overall mortality [7–9]. Based on the latest diabetes prevalence estimate, 10.5% of the global adult population has been suffering from diabetes, and almost half of them are undiagnosed [10]. The growing at-risk population has further strained scarce health resources. Globally, around 10.6% of adults have impaired glucose tolerance (IGT), and 6.2% have impaired fasting glycaemia (IFG) [4]. IGT and IFG are reversible transitional conditions between normality and diabetes. These conditions, also known as prediabetes, are those with elevated blood glucose levels that are not high enough to be classified as diabetic. However, individuals with IGT or IFG are at increased risk of developing cardiovascular disease, coronary heart disease, stroke, and even mortality [11]. One of the challenges with IGT and IFG is that they often do not have any obvious symptoms, which means they can go undetected and undiagnosed for years. Moreover, a follow-up study conducted in Singapore reported that a third of these prediabetic individuals would likely develop T2DM within eight years without lifestyle changes [12]. A similar study with data from the United Kingdom has also reported that a substantial proportion of prediabetes could progress to T2DM within five years [13]. Therefore, predicting the risk of diabetes in the asymptomatic population is a significant health challenge that must be addressed. Early recognition of prediabetes and undiagnosed T2DM will result in a better health outcome or a more favourable long-term prognosis [14].
Currently, the diagnosis of diabetes and prediabetes is well-established. The T2DM and prediabetes can be detected through one of the four methods: 1) the fasting plasma glucose (FPG) value, 2) the 2-hour plasma glucose value during a 75-g oral glucose tolerance test (OGTT), 3) Hemoglobin A1c (HbA1c) and 4) a random plasma glucose test [3]. All these diagnostic screening methods are invasive and opportunistic in nature and must be conducted in a hospital or a laboratory by trained professionals. A confirmed diagnosis usually requires repeated testing. As all the tests are single-time-point screenings, adults above 35 are recommended to have regular

screening every three years. Nevertheless, at-risk individuals hardly comply with this recommendation, especially in developing countries, due to the cost of diagnostic tests and the scarce medical resources [15,16].

Unlike T1DM and gestational diabetes, the development of T2DM and its complications are preventable or at least controllable. A considerable number of studies have shown that lifestyle and behavioural interventions help diabetes patients achieve adequate glycaemic control [17,18]. Recent evidence also suggested that early lifestyle adjustment will help prediabetes subjects get back to normoglycaemia and reduce the risk of developing T2DM [19–21]. Frequent diabetes screening identifies individuals with a high risk of T2DM 2.2 years earlier [22], creating a precious time frame and opportunity for taking early intervention to prevent or delay the onset of diabetes and its complications and improve the overall clinical outcomes.

For established diabetics, constant monitoring of their blood glucose concentration is crucial so that appropriate insulin dosage can be administered timely to avoid acute and chronic complications and delay the disease progression. The conventional blood glucose measurement requires patients to prick their fingers several times a day, which causes the development of massive scarring and loss of sensation at the fingertips over the year [23]. This measurement method is invasive, inconvenient and expensive, which are the main barriers of effective self-management of diabetes in the elderly group [24,25]. To improve the diabetes outcome and assist patients in self-managing the disease, continuous glucose monitoring (CGM) devices have entered the market and are made available for some diabetic patients. However, most CGM sensors currently available are still invasive, which measures glucose concentration in the subcutis by an electrochemical needle sensor [26]. Users need to replace the sensor frequently and purchase different components of the system regularly, which will cost from $2,500 to $6,000 per year [27,28].

In recent years, the advancement and utilisation of wearable technologies and artificial intelligence (AI) have gradually changed our daily lives as many people use wrist-worn wearables daily for fitness and health monitoring [29]. Most consumer wearables have incorporated green light reflection photoplethysmography (PPG) sensors into their products. Wearable technology has the potential to greatly expand the impact of public health initiatives by employing a proactive approach to identify abnormal physiological signals, assessing disease risk factors, and helping patients manage chronic conditions and recovery [30–33].

In 2011, Monte-Moreno demonstrated the use of PPG collected with a pulse oximeter to estimate blood glucose levels [34]. From analysing the PPG waveform, features such as respiration frequency, heart rate variability (HRV), and other physiological parameters can be extracted. They are then fed into a random forest model, giving a prediction accuracy of 87.7% based on the Clark Error Grid. Rodin et al. validated the wearable biosensor developed by Zilberstein et al. as an indirect glucometry [35,36]. The biosensor comprises a PPG sensor and an optically sensitive backglass panel that changes its optochemical characteristics according to the concentrations of specific sweat metabolites. Two hundred adult participants were recruited, and each participant wore a smartwatch to extract the PPG data while the blood samples were collected from the antecubital vein concurrently. The estimation of the blood glucose level was derived using Spectrophon's proprietary algorithm and compared against a glucose lactate analyser (YSI 2300). The proposed biosensor was able to detect anteprandial glucose with a mean absolute percentage error (MAPE) of 7.40% and a normalised root mean squared error (NRMSE) of 11.56%, while postprandial glucose measurements gave 7.54% MAPE and 9.79% NRMSE. Zhang et al. used a smartphone, taking a video of the index finger covering the flash, to capture the fluctuation in the light absorption associated with the change in blood volume. The resulting RGB image was then

transformed into PPG data [37]. The Gaussian fitting method was applied to model the components of the PPG waveform, from which twenty-eight time-domain and frequency-domain features were extracted. A support vector machine (SVM) with a Gaussian kernel was trained with data from eighty subjects to classify the user's glucose level as normal, borderline, or warning, giving an accuracy of 81.49%, along with 79.85% sensitivity, 83.19% specificity and 80.2% F-score. The study was conducted in a highly controlled environment with limited subjects, so the generalisability of these results is subject to certain limitations.

Conventional blood glucose monitoring technologies often require invasive measures such as finger pricking or the use of skin sensors and patches. These methods can be uncomfortable and inconvenient for users and can also be financially burdensome. To address these issues, we propose a novel solution called BGEM™ (Blood Glucose Evaluation and Monitoring) that leverages the latest advancements in signal processing, wearable technology, and AI to detect elevated blood glucose levels and evaluate the risk of developing diabetes. With BGEM™, users only need to measure their PPG data using a consumer-grade wrist-worn wearable. The AI model will then compute relevant digital biomarkers and evaluate the risk of prediabetes or T2DM by recognising elevated blood glucose levels ($\geqslant 7.8 mmol/L$). This solution allows for frequent blood glucose testing without the discomfort and inconvenience of current technologies.

Methods

PPG Sensor

PPG is a low-cost, non-invasive technique that measures the volumetric fluctuation in arterial blood flow [38]. The human wrist is one of the sites for measuring the PPG signal since it has a rich arterial source and an excellent sensor placement with minimal interference to one's daily activities. The PPG signal comprises superimposed pulsatile alternating current (AC) components and direct current (DC) voltage components. A PPG signal is obtained by illuminating LED light on the skin surface and measuring the variations in light absorption or reflection that reflects the pulsatile flow patterns as seen in Figure 1.

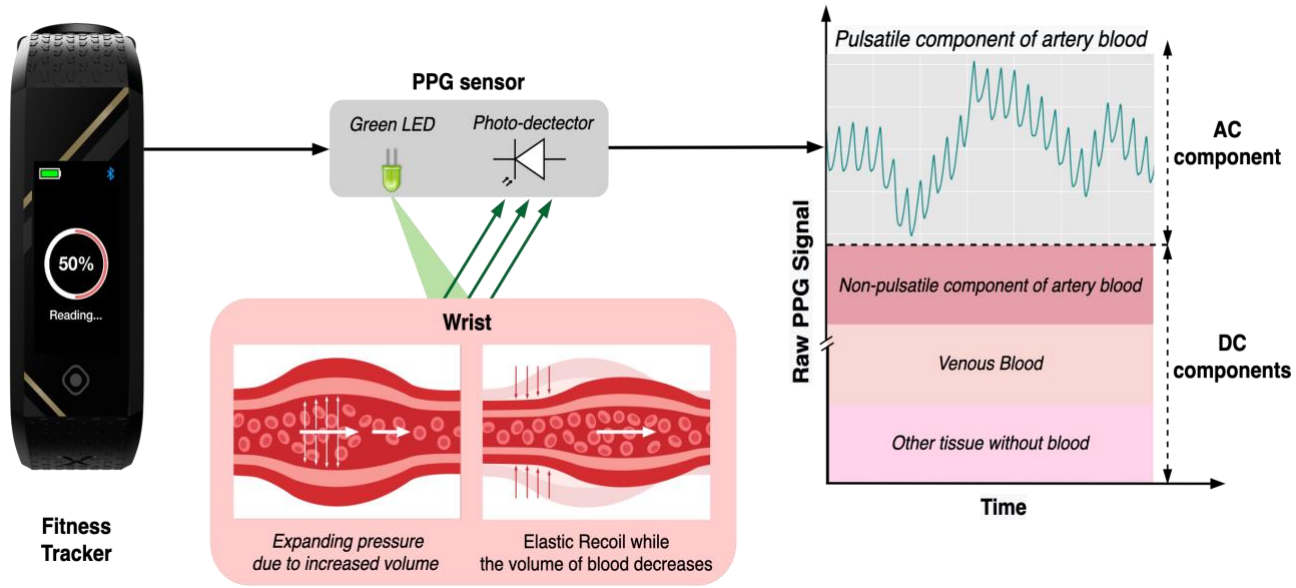


Figure 1. Illustration of the working principle of a PPG sensor. Changes in blood flow represent different phases within the cardiac cycle. During the diastolic phase, blood volume, arterial diameter, and hemoglobin concentration in the measurement site are minimised, leading to

minimum absorption of light by blood and, consequently, an increase in light intensity detected by the sensor system. The reverse is valid for the systolic phase, where a decrease in light intensity is detected instead.

The pulsatile AC component corresponds to the cardiac cycle, characterising that the wrist's blood vessels expand and contract with each heartbeat, while the DC component reflects constant light absorption by venous and arterial blood, as well as other tissues [39]. The PPG signal can detect vascular changes associated with diabetes and contains substantial valuable information from HRV, which is significantly associated with diabetes [40]. Hence, it will be used in this study to extract valuable and meaningful features to identify an individual's glucose status (elevated or normal).

Study Protocol

Before commencing the study, ethical clearance was obtained from the SingHealth Centralised Institutional Review Board of Singapore (CIRB Ref: 2020/2968) on 21/03/2021. All methods were carried out following Singapore's clinical guidelines and regulations. Informed consent was obtained from all trial participants or their legal guardians. The clinical trial was registered on ClinicalTrials.gov (Identifier: NCT05504096) on 17/08/2022.

Five hundred participants were recruited from Singapore KK Women's and Children's Hospital (KKH). The participants' demographic is summarised in Table 1. The blood glucose levels, for most participants, were measured before and after 75g of sugary drink using both the conventional glucometer (Accu-Chek Performa) and the wrist-worn wearable. Subjects who were excluded for the second measurement had high blood glucose measurements $\geqslant 11.1 mmol/L$ on their first measurement and hence were not administered the 75g sugary drink.

Table 1. Description of participants.

Demographic Data			
Age (years)	Mean = 38.73	SD = 10.61	Range = [21, 81]
BMI (kg/m^2)	Mean = 24.4	SD = 5.1	Range = [16.3, 71.1]
Gender	Male	10.2%	
	Female	89.8%	
Diabetes Profile			
Family History of Diabetes	Yes	31.4%	
	No	68.6%	
Pre-diabetes	Yes	3.4%	
	No	96.6%	
Diabetes	Yes	1.6%	
	No	98.4%	
Gestational Diabetes	Yes	4.2%	
	No	85.6%	
	N/A	10.2%	

After consuming the sugary drink, 55.1% of the participants had high blood glucose (≥ 7.8*mmol/L*). The distribution of blood glucose levels before and after drinking the sugary drink is shown in the Figure 2. A statistically significant difference was observed between the two distributions ($p < 0.001$).

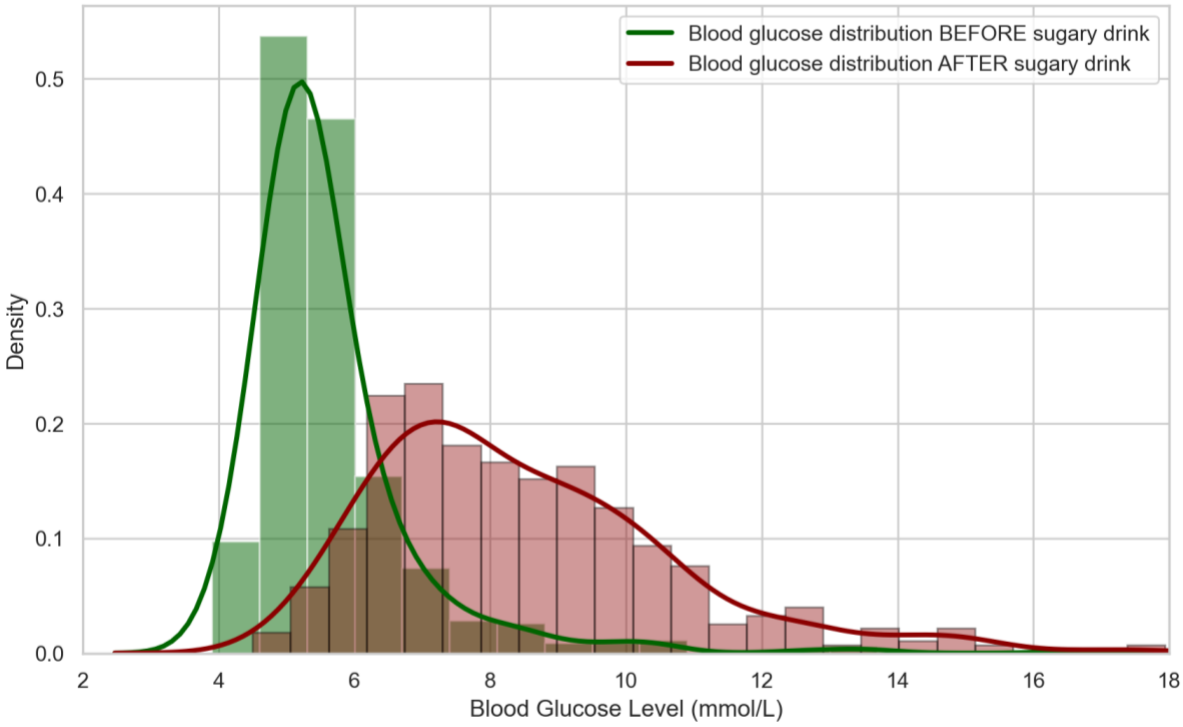


Figure 2. The distribution of ground-truth blood glucose levels before and after sugary drinks ($p < 0.001$).

Study Device

The Actxa Spark+ Series 2, a low-cost and commercially available wrist-worn wearable, was used in the project. This multi-functional device, built for everyday activities, fitness, and preventive health monitoring, provided adequate PPG signal quality at 50 Hz. The wearable is equipped with advanced PPG technology that enables accurate and reliable measurement of heart rate (HR) and other physiological parameters. This is similar to the devices used in Singapore's nation-wide healthcare campaigns, such as the National Steps Challenge™. It is also worth noting that our proposed solution is device-agnostic and can be easily integrated into other wearables with PPG capabilities, allowing for a scalable and cost-effective assessment of risk-based populations, including high-risk subjects, subjects with undiagnosed diabetes, and patients in need of primary prevention interventions.

Pre-processing

The raw PPG signal was collected using both wrist-worn wearables in 16-bit binary format. We first perform a Digital-to-Analog Conversion (DAC) using the formula:

$$V_i = 5 \times \frac{Signal_i}{2^{16}} \quad (1)$$

Liang et al. suggested that a fourth-order Chebyshev II filter provides the optimal processing performance for short PPG signals [41]. Hence, we adopted the recommended filter design to remove the low-frequency drift and high-frequency noise using a band-pass Chebyshev II filter.

The proposed band-pass filter has a lower cut-off frequency of 0.3 Hz and an upper cut-off frequency of 4 Hz.
The filtered PPG signals still contain various forms of outliers, such as peaks with abnormally high amplitudes or distortion in the oscillating waveform, which can be caused by movement from the upper extremity or improper contact between the sensor and skin. Features derived from signals that possess outliers may not be accurate, so a Z-scores outlier detection with a cut-off value of three standard deviations of the mean was applied. The identified outliers or regions of outliers were replaced with a reasonable estimate via a nearest neighbour interpolation for the HRV feature extraction afterwards. As PPG signals do not change drastically in such a short duration, this method is determined to be an appropriate approach to the problem. Further, the number of outliers were minimal in our dataset, and hence should not have affected the features that we generate later. The data pre-processing steps are illustrated in Figure 3.

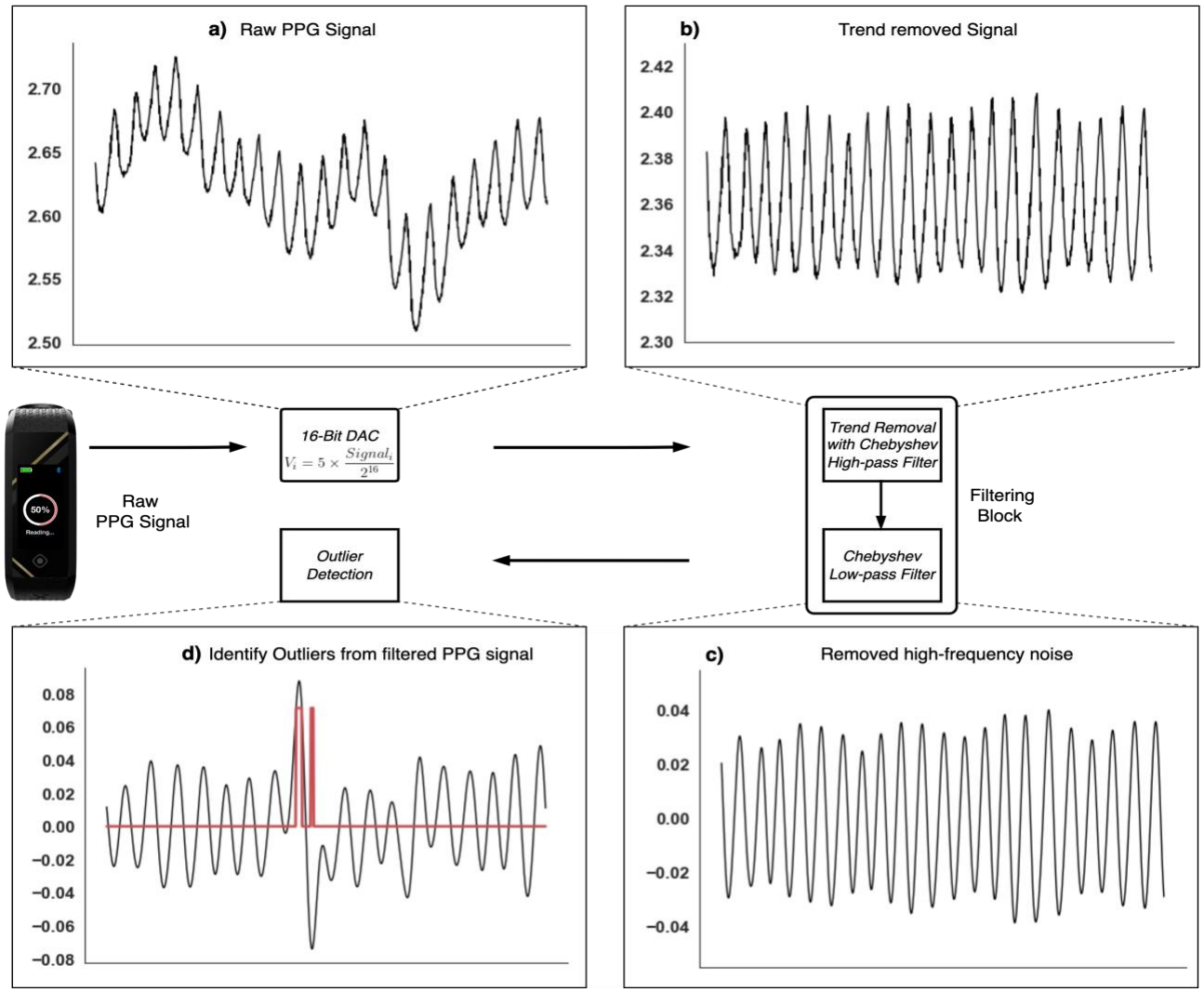


Figure 3. The workflow of data pre-processing.

Feature Extraction

The pre-processed data were suitable for generating reliable features, and a total of 248 features were generated. These features can be classified under seven categories: 1) HRV features, which

encompass time-domain, frequency-domain and non-linear HRV features, 2) waveform features, 3) HR features, 4) energy measures features, 5) complexity measure features, 6) continuous wavelet transform (CWT) features, and 7) patient demographics. The complete set of features analysed in this study is summarised in Table 2. However, these 248 feature candidates are not all relevant to the change in glucose level, and the redundant features might cause prediction performance deterioration. The details of the feature engineering and feature selection process is discussed in the Feature Selection section (under Results).

Table 2. Features Summary

Index	Feature	Description/Equation
HRV time-domain features		
F1	SDNN	The standard deviation of peak-to-peak intervals (PPIs)
F2	SDSD	The standard deviation of the successive difference between adjacent PPIs
F3	RMSSD	Root mean square of successive HRV
F4	pNN20	Percentage of successive PPIs that differ more than 20 ms
F5	pNN50	Percentage of successive PPIs that differ more than 50 ms
F6	BPM	Beats per minute
F7-15	PPI[1]	Time difference between two consecutive systolic peaks
F16-F24	HRV[1]	Time difference between two consecutive PPI
HRV frequency-domain features		
F25-F31	Autoregressive (AR) coefficients	AR coefficients were used to represent the change in the shape of the pulse occurring due to a change in blood flow. To ensure that AR model accurately captures the shape of the pulse, we use AR model of order 7 and compute the coefficients using Yule-Walker equation, which is derived from sample covariance: $\sum_{K=1}^{N} a_k \gamma_{xx}[l-k] = -\gamma_{xx}[l]$
F32-F44	Welch power	Absolute, relative, log, and normalized power of the VLF, LF and HF bands computed using Welch method. The total power across all frequency bands was also computed.
F45-F57	AR power	Absolute, relative, log, and normalized power of the VLF, LF and HF bands computed using AR method. The total power across all frequency bands was also computed.
F58-F60	Welch peak	The peak frequency of the VLF, LF and HF bands were computed using the Welch method
F61-F63	AR peak	The peak frequency of the VLF, LF and HF bands were computed using AR model
F64-F65	LF/HF	The ratio of LF-to-HF power was computed using AR and Welch method
HRV non-linear domain features		
F66	Area	Area of the ellipse which represents total HRV
F67	SD1	Poincaré plot standard deviation perpendicular to the line of identity
F68	SD2	Poincaré plot standard deviation along the line of identity
F69	SD1/SD2	Ratio of SD1-to-SD2

F70	DFA $\alpha 1$	Detrended fluctuation analysis (DFA), which describes short-term fluctuations
F71	DFA $\alpha 2$	DFA, which describes long-term fluctuations
Heart rate features		
F72-F81	HR statistics[1]	Heart Rate Statistics
Continuous wavelet transform features		
F82-F84	CWT	The CWT was performed on the PPG signal using the Mexican Hat Wavelet. The mean, standard deviation and maximum value of the resulting CWT values were used.
Waveform features		
F85-94	RP[1]	Magnitude of rising edge peak for PPG signal
F95-104	FN[1]	Magnitude of falling edge notch for PPG signal
F105-F114	RT[1]	Rising time $RT = t_p - t_s$ (t_s: start time of the current waveform; t_p: the peak time of the current waveform.)
F115-F124	FT[1]	Falling time $FT = t_e - t_p$ (t_e: end time of the current waveform.)
F125-F134	AUR[1]	Area under rising edge
F135-F144	AUF[1]	Area under falling edge
F145-F154	Apulse[1]	Area under one pulse/waveform $Apulse = AUR+AUF$
F155-F164	Aratio[1]	$Aratio = \frac{AUR}{AUF}$
F165-F174	Rslope[1]	Slope of rising edge $Rslope = \frac{f(t_p)-f(t_s)}{t_p-t_s}$
F175-F184	Fslope[1]	Slope of falling edge $Fslope = \frac{f(t_e)-f(t_p)}{t_e-t_p}$
F185-F194	Timediff[1]	$Timediff = RT - FT$
F195-F196	Eigenvalue	The first and second Eigenvalue of the first derivative of PPG signal.
Energy features		
F197-F206	KTE[1]	Kaiser-Teager Energy
F207-F223	LogE[1]	On top of the statistical parameters of log energy (LogEn), $LogE_n$, was used to compute AR coefficients of order 7
Complexity measures		
F224	SampEn	Sample Entropy
F225-F244	MSE	Multiscale Entropy reveals the confidence of entropy measures on the scale by quantifying the time series' complexity
Patient demographics		
F245	Age	Age of the subject
F246	BMI	Body Mass Index
F247	Family History	If an ancestor had diabetes
F248	Gender	Gender of the subject
[1] Statistical parameter such as mean, median, standard deviation, skewness, kurtosis, minimum, maximum, interquartile range, mean absolute difference (MAD) and the difference between the mean and median (MNMD) were computed.		

HRV Features

HRV is the variation in time intervals between consecutive heartbeats and is widely used as the non-invasive physiological biomarker of autonomic nervous system (ANS) response [42–44]. HRV provides a proxy to measure sympathetic (SNS) and parasympathetic (PNS) activity, which reflect the ability to respond to and recover from abrupt physical, psychological, and environmental changes [44–46]. As HR estimated at any given time represents the net effect of the neural output of the PNS, which slows HR, and SNS, which accelerates HR, HRV also detects imbalance in the ANS resulting from over or under stimulation of SNS and PNS. Therefore, the fluctuation in HRV values contains useful insights into many clinical applications, such as mental stress, exercise and rehabilitation, cardiovascular fitness, pathological state, the progression of chronic disease, and even predicting the onset of diseases [47–51]. Depending on the application, HRV features are usually extracted from an ultra-short-term (< 5 minutes), short-term (around 5 minutes), or whole day 24-hour time frame [52]. Most HRV features can be grouped under time-domain, frequency-domain, or non-linear categories. In this project, most of the widely used HRV features were included in our analysis and were extracted using a 5-minute time frame. These HRV features are briefly explained in Table 2 with feature indices (F1-F71).

Heart Rate Features

Prior studies have noted the influence of impaired blood glucose on HR, especially the resting HR [53,54]. Hence, HR was extracted by finding the number of peaks for every 10 seconds of filtered PPG signal. The statistical features of HR were then calculated and used as part of the feature inputs (F72-F81).

Wavelet Analysis

A considerable amount of literature has applied wavelet transformation to analyse the HRV data associated with a wide variety of healthcare applications. Earlier research has utilised features derived from CWT to predict blood glucose level [55]. In this project, we applied CWT to the PPG signal using the Mexican Hat mother wavelet. The mean, standard deviation, and maximum value of the resulting CWT matrix were included in the feature vector (F82-F84).

Waveform Features

Previous studies have reported that the characteristics of the PPG waveform extracted from healthy and diabetic subjects exhibited statistical differences [37,56]. Nirala et al. also suggested that the first and second eigenvalues derived from the first derivative of the PPG signal are the top features to identify T2DM. In addition, several studies thus far have revealed a functional relationship between the PPG signal and blood glucose levels [34,57]. Likewise, respiratory information can also be extracted from the PPG waveform [33,58]. However, PPG waveforms derived from signals using a wrist-worn PPG sensor often have a non-detectable diastolic peak and dicrotic notch, unlike signals collected using fingertip PPG.

Waveform features (F85-F196) derived from the PPG waveform were included in the feature set, and the definition of the waveform features are illustrated in Figure 4.

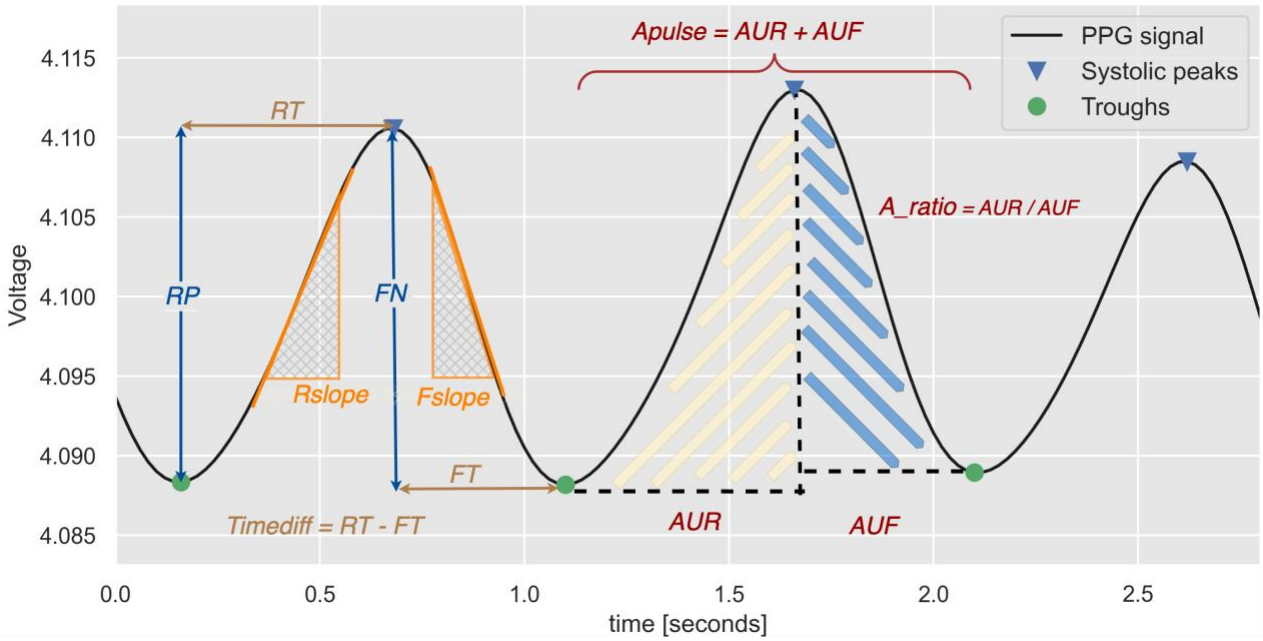


Figure 4. Definition of the PPG waveform features. (RP: Magnitude of rising edge, RT: Rise time, Rslope: Slope of rising edge, FN: Magnitude of falling edge, Fslope: Slope of falling edge, FT: Fall time, Apulse: Area under a PPG wave, AUR: Area under the rising edge, AUF: Area under the falling edge.)

Energy Measures

Several studies have utilised the energy features extracted from PPG signals to estimate blood glucose [34,59,60]. The Kaiser-Teager Energy (KTE) operator and Logarithmic Energy (LogE) are two commonly used methods to analyse the energy profile. These features were computed from a 5-second sliding window as it ensures that the PPG signals within each window would be long enough to contain several heart beats but short enough such that the wave amplitude changes are negligible.

KTE operator is a well-known method to provide time-frequency analysis on the instantaneous energy of the PPG signal from the amplitude and frequency. Using the implementation strategy explained by Monte-Moreno [34], we computed the energy profile of the PPG signal at each sliding window frame, and the KTE operator for the n-th frame is computed using the following equation:

$$KTE_n(i) = x_{frame}(i)^2 - x_{frame}(i+1) * x_{frame}(i-1), \quad \textit{which holds for } i = 2,3,\ldots,(L_{frame}-1) \qquad (2)$$

where x_{frame} is the filtered PPG signal within each sliding window frame.

The statistical metrics were computed for each frame, and the average of the metric for the n-th frame was then calculated and represented as F197 to F206.

To estimate the respiration rate from the PPG signal, we used the LogE value calculated at the frame level with the equation below:

$$LogE_n = log_2(\sum_{\tau=1}^{L_{frame}} x_{frame}(\tau, n)^2) \qquad (3)$$

where x_{frame} is the filtered PPG signal within each sliding window frame.

The autoregressive model coefficients of order seven were estimated using the Yule-Walker method, and the python function *aryule* was used for this purpose. In addition, other statistical parameters were also computed (F207-F223).

Complexity Measures

Sample entropy (SampEn, F224) measures the unpredictability of physiological signals and is commonly used in heart rate variability analysis [61]. The lower the SampEn, the more regular the signal is.

SampEn can be defined after first calculating the template vector ϕ^m which is the probability that two sequences will match for m points without allowing self-counting [62] :

$$SampEn(m, r, N) = -ln\,[\phi^m(r) - \phi^{m+1}(r)], \quad where \quad \phi^m(r) = \frac{1}{N-m-1}\sum_{i=1}^{N-m+1} C_i^m(r) \qquad (4)$$

Where m denotes the embedding dimension, tolerance r equals to $0.1 * standard\ deviation$, N denotes the number of data points, and C_i^m counts, within the tolerance resolution r, the number of matching blocks across different embedding dimensions.

SampEn is a tool to analyse physiological time-series data, but it does not evaluate the data's complexity in different time scales. Hence, we applied the multiscale entropy (MSE) analysis on raw PPG signals to evaluate the hypothetical difference in signal complexity across various time scales for normoglycemia and elevated glucose levels. However, the scale factor is inversely proportionate to the number of data points. From our empirical results, we found that a minimum of 240 pulse waves were required in order to correctly compute MSE values over all the time scale factors (τ =20). We found that the sample entropy calculated from PPG signals during periods of elevated blood glucose was significantly higher than that of blood glucose in the normal range at time scale factors between 8 and 14 (τ). This information was then used to create features for detecting elevated blood glucose. Each time scale factor between 8 and 14 was used as a separate feature. Additionally, the mean of adjacent time scale factors was derived to create additional features. These MSE features were represented in the feature vector with feature indices F225 to F244.

Results

Software

All experiments and analyses were performed using Python 3.9 and relevant libraries (Table 3). The final model was deployed on Amazon Web Services (AWS).

Table 3. A list of the software, and relevant libraries, along with the versions used.

Library	Version	Library	Version	Library	Version
Python	3.9.10	Neurokit2	0.2.2	Scikit-learn	1.1.3
Imbalanced-learn	0.10.1	Nolds	0.5.2	Scipy	1.8.0
Joblib	1.2.0	Numba	0.53.1	Seaborn	0.12.1
Jupyter	1.0.0	Numpy	1.23.5	Spectrum	0.8.1
Lightgbm	3.3.4	Pandas	1.5.2	Statsmodels	0.13.5
Matplotlib	3.6.2	Pillow	9.3.0	Xgboost	1.7.2

Feature Selection

Considering the AI ethics and the practicality of implementing the algorithm, some demographic data, such as skin colour, race, and personal lifestyle habits, were not used as input to the models. However, other general personal particulars associated with the risk of developing T2DM, such as age, gender, BMI, and family health history of diabetes, were added to the feature vector before the feature selection process.

The redundant or irrelevant features might hinder the prediction model's performance. To reduce the dimensionality of the input features, we applied an ensemble strategy which utilises multiple feature selection algorithms. This creates an optimal feature subset that minimises the prediction error rate and is most relevant in predicting the target variable. The ensemble feature selection steps are summarised below:

- Six feature selection methods, including ANOVA correlation coefficient, mutual information (MI), dispersion ratio, recursive feature elimination (RFE), lasso regression, and eXtreme Gradient Boosting (XGBoost), were used to choose the 30 best features independently.
- We combined the features obtained from each feature selection method and ranked the features using a majority vote approach to find the common features selected by more than one model.
- The highly correlated features were dropped from the selected feature subset.

Twelve features were selected from the entire feature set and ranked based on the feature selection strategy's result (shown in Table 4). In our study, these selected features are the most sensitive predictors to capture the characteristics of a subject's elevated blood glucose.

Table 4. The selected top features after the ensemble feature selection method. [63]

Rank	Feature	Rank	Feature	Rank	Feature	Rank	Feature
1	Welch_hf_rel	5	age	8	LOG_std	11	familyHistory
2	AR_hf_rel	6	A_Pulse_iqr	9	BMI	12	A_ratio_max
3	A_FE_mean	7	KTE_skew	10	MSE_sum_13_14	13	Gender[a]
4	A_ratio_mean						

[a]Note that gender was not selected as a top feature in our feature selection algorithm. However, it was previously identified as a sensitive predictor for T2DM, where the prevalence of T2DM in men is higher than in women [63]. This discrepancy could be attributed to gender imbalance in the dataset (male 10.2% and female 89.8%). Therefore, we included gender as one of the top features to provide a complete user profile for future investigation and development.

The selected features could be further divided into 4 main categories. Under the time-domain features, the features selected were the area under the PPG curves. A_FE_mean refers to the average area under the falling edge of each pulse (Figure 4, AUF). A_ratio refers to the ratio of the area under the rising edge to the area under the falling edge of each pulse (Figure 4, A_ratio), and both the average and the maximum values were deemed relevant to the model's predictions. A_pulse_iqr refers to the interquartile range of the total area under each pulse (Figure 4, Apulse). In the frequency-domain, the selected features were the relative powers of the high frequency bands in both the welch PSD (Table 2, F32-F44) and the autoregressive PSD (Table 2, F45-F57). In the non-linear domain, the features selected were either related to the energy or the complexity of the signal. LOG_std refers to the standard deviation of the log energy entropy (Eq. 3), while the KTE_skew refers to the skewness of the KTE energy measure of each sliding window

(Eq. 2). Further, the complexity feature that was selected was the sum of the MSE over 2 scales, 13 and 14.
Lastly, the remaining selected features were demographic features that describes the age and BMI of the participants, as well as if they had any family history of diabetes.

Machine Learning Model Performance

Seven widely used machine learning (ML) algorithms, including the Naïve Bayes (NB) classifier, K-nearest neighbours (KNN) algorithm, logistic regression (LR), random forest (RF), support vector machine (SVM), eXtreme Gradient Boosting (XGB), and light gradient boosting machine (LGBM), were trained with the selected features as input. We fine-tuned the hyperparameters of each model and then validated their performance using the stratified 10-fold cross-validation method. To prevent overfitting during the model training, we adopted multiple regularization techniques across various models. Six evaluation metrics, accuracy, sensitivity, specificity, precision, geometric mean (G-mean) and F-score, were used to evaluate the model's performance, as accuracy alone cannot provide a comprehensive examination of the model performance due to data imbalance. G-mean and F-score are the critical evaluation criteria to assess the models' performance as they are robust to significant label imbalance.
The prediction result from each model is reported with the mean and standard deviation of the evaluation metrics, and Table 5 shows the summary of the results. SVM with the Radial Basis Function (RBF) kernel showed the best prediction performance with an average accuracy of 84.7%, a sensitivity of 81.05%, a specificity of 88.35%, and a precision of 87.51%. Most especially, an average G-mean of 84.54% and F-score of 84.03%.

Table 5. The prediction results obtained from 10-fold cross-validation using various ML models.

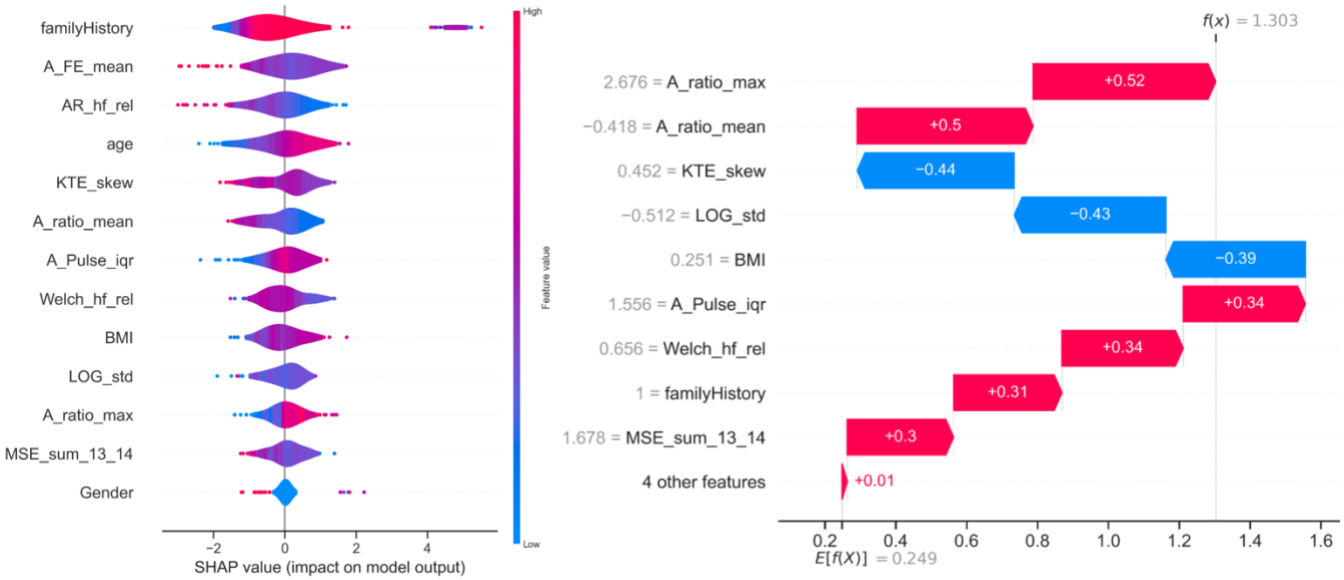
	Accuracy		Sensitivity		Specificity		Precision		G-mean		F-score	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
NB	60.51	4.63	66.17	7.44	54.87	5.78	59.43	4.12	60.08	4.6	62.51	5.19
KNN	76.7	3	90.45	4.30	62.94	4.15	70.97	2.47	75.4	3.09	79.5	2.68
LR	63.1	4.65	64.56	7.07	61.66	4.30	62.65	4.16	63	4.67	63.52	5.37
RF	76.76	5.73	76.84	8.18	76.69	6.42	76.81	6.08	76.64	5.72	76.68	6.23
SVM	**84.7**	4.14	81.05	6.77	**88.34**	4.19	**87.51**	4.26	**84.54**	4.18	**84.03**	4.58
XGB	78.06	4.91	77	6.58	79.12	4.98	78.7	4.88	78	4.89	77.77	5.15
LGBM	77.9	3.98	75.54	7.36	80.27	4.45	79.35	4.1	77.74	4.07	77.24	4.81

Model Interpretation

The use of deep learning in the medical and healthcare domain has shown great potential for solving a range of problems, such as detecting specific symptoms or abnormalities [64,65]. However, the interpretability of deep learning models remains a significant challenge, and it is often difficult for clinicians to trust the decisions made by a black-box system. The lack of model interpretability also raises ethical concerns, especially when the decision fails. Furthermore, our current dataset is considerably small (500 subjects) as compared to typical deep learning models in other domains, which are trained with thousands of data points. Deep learning models are known to perform well with a larger dataset and fail to learn meaningful representations when there is a lack of data[66]. For these reasons, we did not investigate the use of deep learning in this study.

As the proposed ML model is designed to complement the existing diabetes detection solution and is relatively new to the clinical community, the features selected in the previous section must be interpretable and exhibit a certain level of agreement with existing findings. A family history of diabetes, being a male, being over 45 years old, and having an increased BMI have been identified as major risk factors in the literature for developing prediabetes, or T2DM [63,66,67]. These four risk factors were part of the selected predictors, and this paper provides a preliminary attempt to explain how the selected predictors contribute to detecting elevated blood glucose using the SHapley Additive exPlanations (SHAP) framework. SHAP is a game theoretic approach that provides global and local explanations of the association between ML output and input features [68].

Figure 5 (a) illustrates the SHAP values of each feature across all predictions from the training set. The features are ranked by their mean SHAP values, with larger values shown in red and smaller values shown in blue. The beeswarm plot reveals that a family history of diabetes, increasing age, and higher BMI are associated with a higher probability of elevated blood glucose levels. These observations are consistent with previous research and demonstrate that the ML algorithm has successfully captured the relationship between these features and elevated blood glucose. In addition, other proposed features also showed varying levels of impact on the model's output. However, the gender feature did not have any apparent effect on the model's predictions.



(a) SHAP beeswarm plot.

(b) SHAP waterfall plot.

Figure 5. The SHAP explanation plots indicate the association between the selected features and their impact on the predicted outcome.

In Figure 5 (b), each row in the plot shows how the contributions of different features move the model's output from the expected value ($E[f(x)]$) to the actual prediction output $f(x)$ for a single sample with a positive class prediction (blood glucose level $\geqslant 7.8 mmol/L$) in the test set. The expected value, $E[f(x)]$, is determined by the entire training dataset. As expected, most features give positive SHAP values in this sample, which collectively push the model's output towards the correct prediction. However, this specific test subject's BMI was in the healthy range, which pushed the model's output towards the normal class and might result in a false negative

prediction. This shows that relying on a single feature or demographic data alone may not give an accurate prediction of blood glucose levels.
Using the SHAP values, we can understand the model's overall behaviours and how features affect the output positively or negatively, which can help improve the prediction model in the future.

Assessment of the elevated blood glucose levels from multiple measurements

Diagnostic tests are generally not both highly sensitive and highly specific. For this reason, repeated measurements of the wrist-worn wearable were combined and assessed in an optimum fashion to maximise sensitivity, specificity, and precision.
Consecutive measures of blood glucose were combined in parallel using the "AND" and "OR" rules to assist in the detection of elevated blood glucose measurement levels. The "OR" rule increases the overall sensitivity, and the "AND" rule increases the overall specificity, greater than that of either test alone [69].

Discussion

While the healthcare landscape is changing, the rapidly ageing society and the need for improved population health outcomes call for new models of care to effectively prevent the onset and delay the progression of chronic diseases. Furthermore, short-term health behaviours contribute significantly towards longer-term health outcomes, while unattended and frequent glucose spikes might result in prediabetes and, eventually, diabetes. The availability of non-invasive and device-agnostic blood glucose detection solution will allow for more frequent and better monitoring of blood glucose levels, hence reducing the risk of developing T2DM. This study demonstrates that a non-invasive method of assessing diabetic risk using PPG is a viable option to provide a cheaper and accessible modality for population-wide screening of blood glucose levels. This population wide screening would allow the earlier detection of diabetes mellitus in the population, especially in those subjects who are unaware of their elevated blood glucose levels. Hence, timely and appropriate lifestyle advice and medical intervention can be provided to prevent the complications of diabetes. This will subsequently reduce the healthcare burden to both the individual and the society.
BGEM™ is a cloud-based solution that can frequently monitor multiple digital biomarkers with minimal disruption to daily life. Developed using the advanced Machine Learning Operations (MLOps) practice, BGEM™ can easily scale to meet the increasing demand for healthcare services. The solution includes a user-friendly mobile app that can screen a large population to identify high-risk individuals, people with undiagnosed diabetes, and those who need primary prevention intervention. It also provides timely feedback to users through the app, informing them of their diabetes risk and providing targeted, actionable insights to empower them to take a proactive approach to monitor their glucose levels.
As a pilot study, our study has certain limitations. Since fasting blood glucose measurements were excluded, and the criteria to define normal and abnormal levels under fasting conditions differed from our current cut-off, we must refrain from definitively concluding that our model is applicable to fasting conditions. Regarding gender, our feature selection model did not specifically incorporate it, and our analysis utilizing SHAP demonstrated that gender exerted minimal influence on model predictions. Moreover, all analyses were adjusted for the covariate gender, as required. As such, we consider gender to have a limited impact and is not a primary limitation of our findings. To address these limitations, we are actively planning the subsequent phase of data collection. This phase will involve collecting fasting blood glucose measurements in a primary care setting, also allowing for a more balanced gender distribution. More importantly, we can

expand our participant pool to encompass pre-diabetes and diabetes subjects. By addressing these gaps, we aim to offer a more comprehensive and robust assessment of our model's applicability and effectiveness.

There was also no longitudinal follow-up of participants. External validation of our model on an independent sample needs to be undertaken to further assess the detection accuracy and the generalizability of the results. Nevertheless, as a preliminary investigation, the potential implications of our findings are significant, as they might offer a means to identify previously undiagnosed pre-diabetes or diabetes cases at the population level. We anticipate that our study will serve as a foundational stepping stone, paving the way for more comprehensive diabetes research utilizing AI and wearables. To the best of our knowledge, no publicly available dataset exists that systematically examines the relationship between PPG data and blood glucose levels. Acquiring a substantial volume of data is imperative, encompassing a diverse and representative sample that spans the entire spectrum of glucose values and incorporates relevant sociodemographic factors. Such comprehensive data can be attained through a collaborative effort involving research institutions and industry partners while ensuring strict adherence to local ethical considerations and data privacy regulations.

We demonstrated that the deployed cloud-based ML model was able to detect elevated blood glucose levels, where consecutive measurements could be combined in an optimal manner to provide high sensitivity, specificity, and precision. Further research is required to address the limitations discussed.

Conclusion

In this study, we performed sophisticated feature engineering, and we found that the features derived from multiscale entropy analysis of PPG signals effectively detect blood glucose changes. We will discuss this set of novel features in more detail in a separate paper. To reduce bias and evaluate the model's generalizability, we used 10-fold cross-validation to assess its performance. The SVM with RBF model performed the best, with an average accuracy of 84.7%, a G-mean of 84.54%, and an F-score of 84.03%. Previous models were developed using smaller samples and had lower model performance measures. Our model was developed with a larger sample of 500 subjects, and most subjects were assessed before and after consumption of a sugary drink. It also achieved better detection accuracy.

References

1. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. Diabetes; Dec. 1979;28(12):1039–1057. doi: 10.2337/DIAB.28.12.1039.
2. Kerner W and Brückel J. Definition, classification and diagnosis of diabetes mellitus. Experimental and Clinical Endocrinology & Diabetes; Jul. 2014;122(7):384–386. doi: 10.1055/s-0034-1366278.
3. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2020. Diabetes Care; Jan. 2020;43(Suppl 1):S14–S31. doi: 10.2337/dc20-S002.
4. International Diabetes Federation, "IDF diabetes atlas," Brussels, Belgium, 2021. https://diabetesatlas.org/idfawp/resource-files/2021/07/IDF_Atlas_10th_Edition_2021.pdf (accessed Jun., 2022).
5. The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102

prospective studies. The Lancet; Jun. 2010;375(9733):2215–2222. doi: 10.1016/S0140-6736(10)60484-9.
6. Lin X, et al. Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025. Scientific Reports; Sep. 2020;10(1):1–11. doi: 10.1038/s41598-020-71908-9.
7. Li S, Wang J, Zhang B, Li X, and Liu Y. Diabetes mellitus and cause-specific mortality: a population-based study. Diabetes Metab J; Jun. 2019;43(3):319. doi: 10.4093/dmj.2018.0060.
8. Saran R, et al. US renal data system 2014 annual data report: epidemiology of kidney disease in the United States. American Journal of Kidney Diseases; Jul. 2015;66(1):A7. doi: 10.1053/j.ajkd.2015.05.001.
9. Lau L H, Lew J, Borschmann K, Thijs V, and Ekinci E I. Prevalence of diabetes and its effects on stroke outcomes: A meta-analysis and literature review. J Diabetes Investig; May 2019;10(3): 780–792. doi: 10.1111/JDI.12932.
10. Sun H, et al. IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. Diabetes Res Clin Pract; Jan. 2022;183. doi: 10.1016/j.diabres.2021.109119.
11. Huang Y, Cai X, Mai W, Li M, and Hu Y. Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. BMJ; Nov. 2016;355. doi: 10.1136/bmj.i5953.
12. Wong M-S, Gu K, Heng D, Chew S-K, Chew L-S, and Tai E S. The Singapore impaired glucose tolerance follow-up study: does the ticking clock go backward as well as forward? Diabetes Care; Nov. 2003;26(11):3024–30. doi: 10.2337/diacare.26.11.3024.
13. Tabák A G, Herder C, Rathmann W, Brunner E J, and Kivimäki M. Prediabetes: a high-risk state for diabetes development. Lancet; Jun. 2012;379(9833):2279–90. doi: 10.1016/S0140-6736(12)60283-9.
14. US Preventive Services Task Force, et al. Screening for prediabetes and type 2 diabetes: US Preventive Services Task Force recommendation statement. JAMA; Aug. 2021;326(8):736–743. doi: 10.1001/jama.2021.12531.
15. Manne-Goehler J, et al. Health system performance for people with diabetes in 28 low- and middle-income countries: a cross-sectional study of nationally representative surveys. PLoS Med; Mar. 2019;16(3). doi: 10.1371/journal.pmed.1002751.
16. Misra A, et al. Diabetes in developing countries. J Diabetes; Jul. 2019;11(7):522–539. doi: 10.1111/1753-0407.12913.
17. García-Molina L, Lewis-Mikhael A-M, Riquelme-Gallego B, Cano-Ibáñez N, Oliveras-López M-J, and Bueno-Cavanillas A. Improving type 2 diabetes mellitus glycaemic control through lifestyle modification implementing diet intervention: a systematic review and meta-analysis. Eur J Nutr; Jun. 2020;59(4):1313–1328. doi: 10.1007/s00394-019-02147-6.
18. O'donoghue G, et al. Lifestyle interventions to improve glycemic control in adults with type 2 diabetes living in low-and-middle income countries: a systematic review and meta-analysis of randomized controlled trials (rcts). Int J Environ Res Public Health; Jun. 2021;18(12). doi: 10.3390/IJERPH18126273/S1.
19. Tuso P. Prediabetes and lifestyle modification: time to prevent a preventable disease. Perm J; Jun. 2014;18(3):88–93. doi: 10.7812/TPP/14-002.
20. Bansal N. Prediabetes diagnosis and treatment: a review. World J Diabetes; Mar. 2015;6(2):296–303. doi: 10.4239/wjd.v6.i2.296.

21. Magkos F, Hjorth M F, and Astrup A. Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. Nature Reviews Endocrinology; Jul. 2020;16(10):545–555. doi: 10.1038/s41574-020-0381-5.
22. Simmons R K, Griffin S J, Lauritzen T, and Sandbæk A. Effect of screening for type 2 diabetes on risk of cardiovascular disease and mortality: a controlled trial among 139,075 individuals diagnosed with diabetes in Denmark between 2001 and 2009. Diabetologia; Nov. 2017;60(11):2192–2199. doi: 10.1007/s00125-017-4299-y.
23. Heinemann L. Finger pricking and pain: a never ending story. J Diabetes Sci Technol; Sep. 2008;2(5):919–921. doi: 10.1177/193229680800200526.
24. Hambling C E, Seidu S I, Davies M J, and Khunti K. Older people with type 2 diabetes, including those with chronic kidney disease or dementia, are commonly overtreated with sulfonylurea or insulin therapies. Diabetic Medicine; Sep. 2017;34(9):1219–1227. doi: 10.1111/dme.13380.
25. Mattishent K, et al. Continuous glucose monitoring in older people with diabetes and memory problems: a mixed-methods feasibility study in the UK. BMJ Open; Nov. 2019;9(11). doi: 10.1136/bmjopen-2019-032037.
26. Vettoretti M, Cappon G, Acciaroli G, Facchinetti A, and Sparacino G. Continuous glucose monitoring: current use in diabetes management and possible future applications. J Diabetes Sci Technol; Sep. 2018;12(5):1064–1071. doi: 10.1177/1932296818774078.
27. Funtanilla V D, Caliendo T, and Hilas O. Continuous glucose monitoring: a review of available systems. P and T; Sep. 2019;44(9):550–553.
28. Robertson S L, Shaughnessy A F, and Slawson D C. Continuous glucose monitoring in type 2 diabetes is not ready for widespread adoption. Am Fam Physician; Jun. 2020;101(11):646–646.
29. Sabry F, Eltaras T, Labda W, Alzoubi K, and Malluhi Q. Machine Learning for Healthcare Wearable Devices: The Big Picture. J Healthc Eng; 2022. doi: 10.1155/2022/4653923.
30. Patel S, Park H, Bonato P, Chan L, and Rodgers M. A review of wearable sensors and systems with application in rehabilitation. J Neuroeng Rehabil; Apr. 2012;9(1):1–17. doi: 10.1186/1743-0003-9-21/FIGURES/7.
31. Rodgers M M, Alon G, Pai V M, and Conroy R S. Wearable technologies for active living and rehabilitation: Current research challenges and future opportunities. J Rehabil Assist Technol Eng; Apr. 2019;6. doi: 10.1177/2055668319839607.
32. Xie Y, et al. Integration of artificial intelligence, blockchain, and wearable technology for chronic disease management: a new paradigm in smart healthcare. Current Medical Science; Dec. 2021;41(6):1123–1133. doi: 10.1007/S11596-021-2485-0.
33. Iqbal S M A, Mahgoub I, Du E, Leavitt M A, and Asghar W. Advances in healthcare wearable devices. npj Flexible Electronics; Apr. 2021;5(1):1–14. doi: 10.1038/s41528-021-00107-x.
34. Monte-Moreno E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. Artif Intell Med; Oct. 2011;53(2):127–38. doi: 10.1016/j.artmed.2011.05.001.
35. Zilberstein G, Zilberstein R, Maor U, and Righetti P G. Noninvasive wearable sensor for indirect glucometry. Electrophoresis; Sep. 2018;39(18):2344–2350. doi: 10.1002/elps.201700424.
36. Rodin D, Kirby M, Sedogin N, Shapiro Y, Pinhasov A, Kreinin A. Comparative accuracy of optical sensor-based wearable system for non-invasive measurement of blood glucose concentration. Clinical biochemistry; Mar. 2019;65:15-20. doi: 10.1016/j.clinbiochem.2018.12.014.

37. Zhang G, Mei Z, Zhang Y, Ma X, Lo B, Chen D, Zhang Y. A noninvasive blood glucose monitoring system based on smartphone PPG signal processing and machine learning. IEEE Transactions on Industrial Informatics; Feb. 2020;16(11):7209-18. doi: 10.1109/TII.2020.2975222.
38. Challoner AV, Ramsay CA. A photoelectric plethysmograph for the measurement of cutaneous blood flow. Physics in Medicine & Biology; May. 1974;19(3):317. doi: 10.1088/0031-9155/19/3/003.
39. Zhao D, Sun Y, Wan S, Wang F. SFST: A robust framework for heart rate monitoring from photoplethysmography signals during physical activities. Biomedical Signal Processing and Control; Mar. 2017;33:316-24. doi: 10.1016/j.bspc.2016.12.005.
40. Schroeder EB, Chambless LE, Liao D, Prineas RJ, Evans GW, Rosamond WD, Heiss G. Diabetes, glucose, insulin, and heart rate variability: the Atherosclerosis Risk in Communities (ARIC) study. Diabetes care; Mar. 2005;28(3):668-74. doi: 10.2337/diacare.28.3.668.
41. Liang Y, Elgendi M, Chen Z, Ward R. An optimal filter for short photoplethysmogram signals. Scientific data; May. 2018;5(1):1-2. doi: 10.1038/sdata.2018.76.
42. van Ravenswaaij-Arts CM, Kollee LA, Hopman JC, Stoelinga GB, van Geijn HP. Heart rate variability. Annals of internal medicine; Mar. 1993;118(6):436-47. doi: 10.7326/0003-4819-118-6-199303150-00008.
43. Xhyheri B, Manfrini O, Mazzolini M, Pizzi C, Bugiardini R. Heart rate variability today. Progress in cardiovascular diseases; Nov. 2012;55(3):321-31. doi: 10.1016/j.pcad.2012.09.001.
44. Thomas BL, Claassen N, Becker P, Viljoen M. Validity of commonly used heart rate variability markers of autonomic nervous system function. Neuropsychobiology; May. 2019;78(1):14-26. doi: 10.1159/000495519.
45. Obrist PA. Cardiovascular Psychophysiology. Boston, MA: Springer US, 1981. doi: 10.1007/978-1-4684-8491-5.
46. Singh N, Moneghetti KJ, Christle JW, Hadley D, Froelicher V, Plews D. Heart rate variability: an old metric with new meaning in the era of using mhealth technologies for health and exercise training guidance. part two: prognosis and training. Arrhythmia & electrophysiology review; 2018 Dec;7(4):247. doi: 10.15420/aer.2018.27.2.
47. Prinsloo GE, Rauch HL, Derman WE. A brief review and clinical application of heart rate variability biofeedback in sports, exercise, and rehabilitation medicine. The Physician and sportsmedicine; May. 2014;42(2):88-99. doi: 10.3810/psm.2014.05.2061.
48. Billman GE, Huikuri HV, Sacha J, Trimmel K. An introduction to heart rate variability: methodological considerations and clinical applications. Frontiers in physiology; Feb. 2015;6:55. doi: 10.3389/fphys.2015.00055.
49. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and heart rate variability: a meta-analysis and review of the literature. Psychiatry investigation; Mar. 2018;15(3):235. doi: 10.30773/pi.2017.08.17.
50. Taye GT, Hwang HJ, Lim KM. Application of a convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features. Scientific reports; Apr. 2020;10(1):6769. doi: 10.1038/s41598-020-63566-8.
51. Mosley E, Laborde S. A scoping review of heart rate variability in sport and exercise psychology. International Review of Sport and Exercise Psychology; Jul. 2022:1-75. doi: 10.1080/1750984X.2022.2092884.

52. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Frontiers in public health; 2017:258. doi: 10.3389/fpubh.2017.00258.
53. Valensi P, Extramiana F, Lange C, Cailleau M, Haggui A, Maison Blanche P, Tichet J, Balkau B, DESIR Study Group. Influence of blood glucose on heart rate and cardiac autonomic function. The DESIR study. Diabetic medicine; Apr. 2011;28(4):440-9. doi: 10.1111/j.1464-5491.2010.03222.x.
54. Inamdar A. Correlation between fasting heart rate and fasting plasma glucose level in rural Indians. European Heart Journal; Feb. 2022;43:849-158., doi: 10.1093/eurheartj/ehab849.158.
55. Gupta S, Gupta RK, KulShReStha M, ChaudhaRy RR. Evaluation of ECG abnormalities in patients with asymptomatic type 2 diabetes mellitus. Journal of clinical and diagnostic research: JCDR; Apr. 2017;11(4):OC39. doi: 10.7860/JCDR/2017/24882.9740.
56. Nirala N, Periyasamy R, Singh BK, Kumar A. Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine. Biocybernetics and Biomedical Engineering; Jan. 2019;39(1):38-51. doi: 10.1016/j.bbe.2018.09.007.
57. Philip LA, Rajasekaran K, Jothi ES. Continous monitoring of blood glucose using photophlythesmograph signal. 2017 International Conference on Innovations in Electrical, Electronics, Instrumentation and Media Technology (ICEEIMT); Feb. 2017:187-191. IEEE. doi: 10.1109/ICIEEIMT.2017.8116832.
58. Moraes JL, Rocha MX, Vasconcelos GG, Vasconcelos Filho JE, De Albuquerque VH, Alexandria AR. Advances in photopletysmography signal analysis for biomedical applications. Sensors; Jun. 2018;18(6):1894. doi: 10.3390/s18061894.
59. Habbu S, Dale M, Ghongade R. Estimation of blood glucose by non-invasive method using photoplethysmography. Sādhanā; Jun. 2019;44(6):135. doi: 10.1007/s12046-019-1118-9.
60. Hina A, Nadeem H, Saadeh W. A single led photoplethysmography-based noninvasive glucose monitoring prototype system. 2019 IEEE International Symposium on Circuits and Systems (ISCAS); May. 2019:1-5. IEEE. doi: 10.1109/ISCAS.2019.8702747.
61. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American journal of physiology-heart and circulatory physiology; Jun. 2000;278(6):H2039-49. doi: 10.1152/ajpheart.2000.278.6.h2039.
62. Delgado-Bonal A, Marshak A. Approximate entropy and sample entropy: A comprehensive tutorial. Entropy; May. 2019;21(6):541. doi: 10.3390/e21060541.
63. Nordström A, Hadrévi J, Olsson T, Franks PW, Nordström P. Higher prevalence of type 2 diabetes in men than in women is associated with differences in visceral fat mass. The Journal of Clinical Endocrinology & Metabolism; Oct. 2016;101(10):3740-6. doi: 10.1210/jc.2016-1915.
64. Shi B, Yen SC, Tay A, Tan DM, Chia NS, Au WL. Convolutional neural network for freezing of gait detection leveraging the continuous wavelet transform on lower extremities wearable sensors data. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); Jul. 2020:5410-5415. IEEE. doi: 10.1109/EMBC44109.2020.9175687.
65. Shi B, Tay A, Au WL, Tan DM, Chia NS, Yen SC. Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors. IEEE Transactions on Biomedical Engineering; Jan. 2022;69(7):2256-67. doi: 10.1109/TBME.2022.3140258.

66. Alwosheel A, van Cranenburgh S, Chorus CG. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. Journal of choice modelling; Sep. 2018; (28):167-82. doi: 10.1016/J.JOCM.2018.07.002.
67. Gulcher J, Stefansson K. Clinical risk factors, DNA variants, and the development of type 2 diabetes. The New England journal of medicine; Mar. 2009; 360(13):1360.
68. Centers for Disease Control and Prevention, "Diabetes Risk Factors | CDC," Apr. 2022. https://www.cdc.gov/diabetes/basics/risk-factors.html (accessed Nov., 2022).
69. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Advances in neural information processing systems; 2017;(30):4768–4777.
70. Zhou XH, McClish DK, Obuchowski NA. Statistical methods in diagnostic medicine. John Wiley & Sons; Sep. 2009. doi: 10.1002/9780470906514.

Acknowledgements

B.S. contributed to the study design, conducted data analysis and experiments, developed the algorithms and models, and drafted the manuscript. S.S.D. designed the study, performed statistical data analysis, and drafted the manuscript. J.W. was responsible for the model deployment and developed the data pipeline infrastructure. C.C., C.M.T., K.L.C.S., and F.F.I. contributed to data collection. N.W.C.L., E.Z, K.Y.L., and V.P. assisted with the development of the algorithms and supported the data collection. A.W.H.L performed data analysis and edited the manuscript. M.S. contributed to the study design and supervised the study. J.C., S.C.Y, and A.T. supervised the study. S.B.A. contributed to the study design and supervised the study. All authors reviewed the manuscript.
This research was sponsored by Actxa Pte. Ltd., but data collection was done independently at KK Women's and Children's Hospital, Singapore.

Conflicts of Interest

We would like to disclose that B.H.S., M.S., J.W., A.W.H.L and J.C. are employed by Actxa Pte. Ltd. We have in place an approved plan for managing any potential conflicts arising from the employment. S.B.A. and S.S.D. are on the advisory board of Actxa Pte. Ltd. The remaining authors have no conflicts of interest to declare.

Abbreviations

AC: alternating current
AI: artificial intelligience
ANOVA: analysis of variance
ANS: autonomic nervous system
AR: autoregressive
AWS: Amazon Web Services
BGEM™: Blood Glucose Evaluation and Monitoring
BMI: body mass index
CGM: continuous glucose monitoring
CWT: continuous wavelet transform
DAC: digital-to-analog conversion
DC: direct current
DFA: detrended fluctuation analysis
DM: diabetes mellitus
FPG: fasting plasma glucose
G-mean: geometric mean

HbA1c: Haemoglobin A1c
HF: high frequency
HR: heart rate
HRV: heart rate variability
IGT: impaired glucose tolerance
IFG: impaired fasting glycemia
KKH: Singapore KK Women's and Children's Hospital
KNN: K-nearest neighbours
KTE: Kaiser-Teager energy
LED: light emitting device
LF: low frequency
LGBM: light gradient boosting machine
LogE: logarithmic energy
LR: logistic regression
MAD: mean absolute difference
MAPE: mean absolute percentage error
MI: mutual information
ML: machine learning
MLOps: machine learning operations
MNMD: mean and median difference
MSE: multiscale entropy
NB: Naïve Bayes
NRMSE: normalised root mean squared error
OGTT: oral glucose tolerance test
PNS: parasympathetic nervous system
PPG: photoplethysmography
PPI: peak-to-peak interval
RBF: radial basis function
RF: random forest
RFE: recursive feature elimination
RGB: red, green and blue
SampEn: Sample entropy
SHAP: Shapley additive explanations
SNS: sympathetic nervous system
SVM: support vector machine
T1DM: type-1 diabetes mellitus
T2DM: type-2 diabetes mellitus
VLF: very low frequency
XGBoost: extreme gradient boosting