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], 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
[*]bohan.shi@actxa.com

ABSTRACT

Diabetes mellitus (DB) 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 in preventing or delaying the onset of diabetes and its complications. This study was conducted at KK Women's and Children's Hospital of Singapore, and five hundred participants were recruited (mean age 38.73 $\pm$ 10.61 years; mean BMI 24.4 $\pm$ 5.1 kg/m^2). 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 propose leveraging photoplethysmography (PPG) sensors and machine learning techniques to incorporate this into an affordable wrist-worn wearable device to detect elevated blood glucose levels ($\geqslant 7.8 mmol/L$) non-invasively. 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 kernel has 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%. Hence, PPG measurements can be utilized to identify subjects with elevated blood glucose measurements and assist in the screening of subjects for diabetes risk.

Introduction

Diabetes mellitus (DB) 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 DB: type-1 DB (T1DB), type-2 DB (T2DB) and gestational diabetes. T2DB is the most prevalent type of diabetes, affecting over 95% of people with diabetes worldwide[3,4].

The prevalence of DB 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 characterised by 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 T2DB 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[3]. 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[6,15,16].

Unlike T1DB and gestational diabetes, the development of T2DB and its complications are preventable or at least controllable. A considerable amount 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 T2DB 2.2 years earlier[22], creating a precious time frame and opportunity for taking early intervention in preventing or delaying the onset of diabetes and its complications and improving 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 it is one of the barriers to effectively self-manage 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 on the market are still invasive and measure 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 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 and other physiological parameters can be extracted, which are then fed into a random forest model, giving a prediction accuracy of 87.7% based on the Clark Error Grid. Rodin et al. proposed using a smartwatch with an integrated biosensor as an indirect glucometry[35]. 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 is 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% MPAE 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[36]. 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 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 artificial intelligence 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 is a low-cost, non-invasive technique that measures the volumetric fluctuation in arterial blood flow[37]. 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 voltage (DC) 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. 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[38]. The PPG signal can detect vascular changes associated with diabetes and contains substantial valuable information from heart rate variability, which is significantly associated with diabetes[39]. Hence, it will be used in this study to extract valuable and meaningful features to identify an individual's glucose status (elevated or normal).

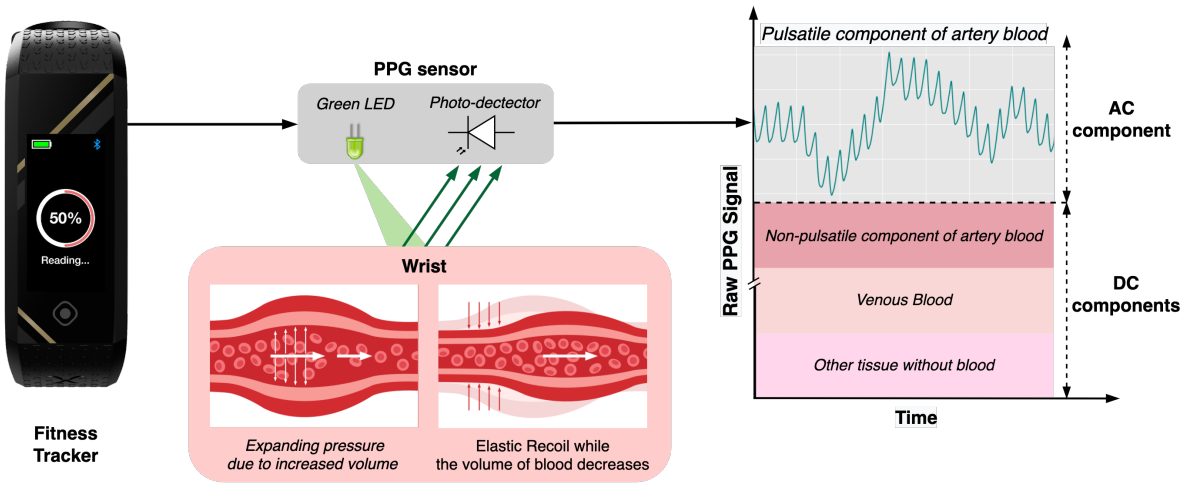


Figure 1. Illustration of the working principle of a PPG sensor. Changes in blood flow represent different phases within the cardiac cycle. During the diastolic phase, blood volume, arterial diameter, and hemoglobin concentration in the measurement site are minimised, leading to minimum absorption of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

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 March 2021. The detailed study protocol was registered on ClinicalTrials.gov (Identifier: NCT05504096).

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 $\geqslant 11.8 mmol/L$ on their first measurement and hence were not administered the 75g sugary drink.

Demographic Data		
Age (years)	mean = 38.73	SD = 10.61
BMI (kg/m^2)	mean = 24.4	SD = 5.1
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%

Table 1. Description of participants.

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 and other physiological parameters. This is similar to the devices used in Singapore's national-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:

$$V_i = 5 \times \frac{Signal_i}{2^{16}} \tag{1}$$

Liang et al. suggested that a fourth-order Chebyshev II filter provides the optimal processing performance for short PPG signals[40]. 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 third standard deviations of the mean. The identified outliers or regions of outliers were excluded from the HRV feature extraction afterwards. The data pre-processing steps are illustrated in Figure 2.

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) heart rate variability (HRV) features, which encompass time-domain, frequency-domain and non-linear HRV features, 2) waveform features, 3) heart rate features, 4) energy measures features, 5) complexity measure features, 6) continuous wavelet transform 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 the feature selection process are discussed in the section.

Index	Feature	Description/Equation	Index	Feature	Description/Equation
HRV Time Domain			***Heart Rate Features***		
F1	SDNN	The standard deviation of PPIs	F72-F81	HR statistics[1]	Heart Rate Statistics
F2	SDSD	The standard deviation of the successive difference between adjacent PPIs	***Continuous Wavelet Transform***		
F3	RMSSD	Root mean square of successive HRV	F82-F84	CWT	The continuous wavelet transform (CWT) was performed on the PPG signal using the Mexican Hat Wavelet. The mean, standard deviation and maximum value of the resulting CWT values.
F4	pNN20	Percentage of successive PPIs that differ more than 20 ms	***Waveform Features***		
F5	pNN50	Percentage of successive PPIs that differ more than 50 ms	F85-94	RP[1]	Magnitude of rising edge peak for PPG signal
F6	BPM	Beats per minute	F95-104	FN[1]	Magnitude of falling edge notch for PPG signal
F7-15	PPI[1]	Time difference between two consecutive systolic peaks	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.)
F16-F24	HRV[1]	Time difference between two consecutive PPI	F115-F124	FT[1]	Falling time $FT = t_e - t_p$ (t_e : end time of the current waveform.)
HRV Frequency Domain			F125-F134	AUR[1]	Area under rising edge
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 : $\sum_{K=1}^{N} a_k \gamma_{xx}[l-k] = -\gamma_{xx}[l]$	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]	$Timediff = RT - FT$
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.	F195-F196	Eigenvalue	The first and second Eigenvalue of the first derivative of PPG signal.
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.	***Energy Features***		
F58-F60	Welch peak	The peak frequency of the VLF, LF and HF bands were computed using the Welch method	F197-F206	KTE[1]	Kaiser-Teager Energy
F61-F63	AR peak	The peak frequency of the VLF, LF and HF bands were computed using AR model	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
F64-F65	LF/HF	The ratio of LF-to-HF power was computed using AR and Welch method	***Complexity Measures***		
HRV Non-linear Domain			F224	SampEn	Sample Entropy
F66	Area	Area of the ellipse which represents total HRV	F225-F244	MSE	Multiscale Entropy reveals the confidence of entropy measures on the scale by quantifying the time series' complexity
F67	SD1	Poincaré plot standard deviation perpendicular to the line of identity	***Patient Demographics***		
F68	SD2	Poincaré plot standard deviation along the line of identity	F245	Age	Age of the subject
F69	SD1/SD2	Ratio of SD1-to-SD2	F246	BMI	Body Mass Index
F70	DFA $\alpha 1$	Detrended fluctuation analysis (DFA), which describes short-term fluctuations	F247	Family History	If an ancestor had diabetes
F71	DFA $\alpha 2$	DFA, which describes long-term fluctuations	F248	Gender	Gender of the subject

[1] Statistical parameters 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) was computed.

Table 2. Features Summary

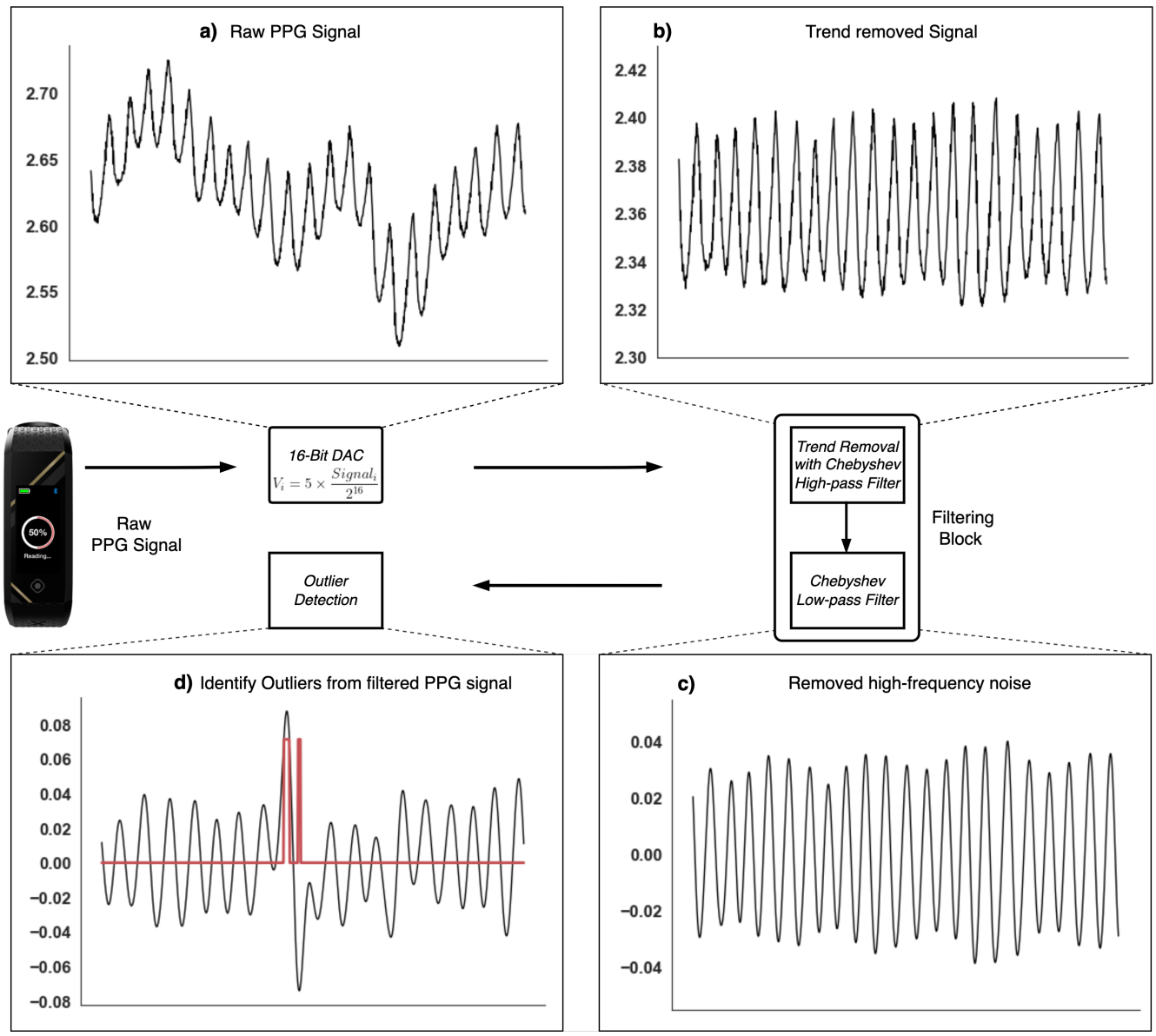


Figure 2. The workflow of data pre-processing.

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 response[41–43]. 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[43–45]. 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 autonomous nervous system 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, an indication of the pathological state, the progression of chronic disease, and even predicting the onset of diseases[46–50]. 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 hours time frame[51]. 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 heart rate, especially the resting heart rate[52,53]. Hence, HR was extracted by finding the number of peaks for every 30 seconds of filtered PPG signal. The statistical features of HR were then calculated and used as part of the feature inputs (F72 to 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 continuous wavelet transform (CWT) to predict blood glucose level[54]. 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- 84).

Waveform Features

Previous studies have reported that the characteristics of the PPG waveform extracted from healthy and diabetic subjects exhibited statistical differences[36,55]. 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[55]. In addition, several studies thus far have revealed a functional relationship between the PPG signal and blood glucose levels[34,56]. Likewise, respiratory information can also be extracted from the PPG waveform[33,57]. 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 3.

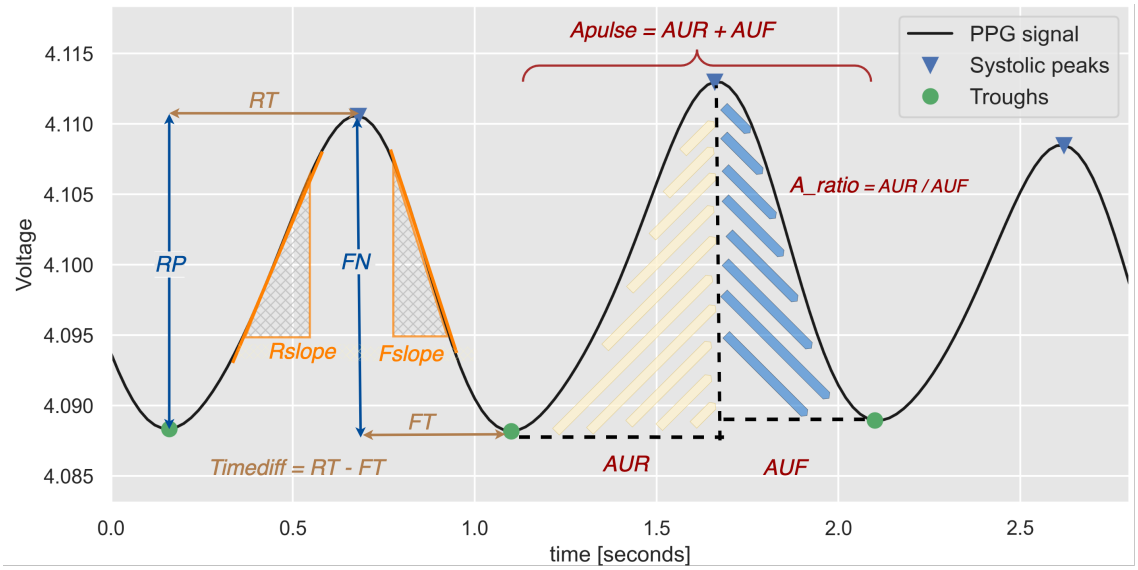


Figure 3. Definition of the PPG waveform features. An additional explanation of each feature can be found in Table 2.

Energy Measure

Several studies, for example,[34,58,59], have utilised the energy features extracted from PPG signals to estimate blood glucose. The Kaiser-Teager Energy (KTE) operator and Logarithmic Energy (LogE) are two commonly used methods to analyse the energy profile.

Kaiser-Teager Energy Operator 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 the frame level, and the KTE operator for the n-th frame is computed using the following equation :

$$KTE_n(i) = x^2_{frame}(i) - x_{frame}(i+1) * x_{frame}(i-1), \;\; which\ holds\ for\ i = 2,3,...,L_{frame}-1 \tag{2}$$

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

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

$$LogE_n = \left(\sum_{\tau=1}^{L_{frame}} S^2_{frame}(\tau, n) \right) \tag{3}$$

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 Sample entropy (SampEn, F224) measures the unpredictability of physiological signals[60] and is commonly used in heart rate variability analysis. 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[61]:

$$SampEn(m, r, N) = ln[\Phi^m(r) - \Phi^{m+1}(r)], \ \ where \ \ \Phi^m(r) = \frac{1}{N-m-1} \sum_{i=1}^{N-m+1} {C_r}^m(i) \tag{4}$$

m denotes the embedding dimension, tolerance r equals to $0.2 * standard\ deviation$, and the number of data points is represented as N.

Multiscale Entropy 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. 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.

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 the 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, are 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 3). In our study, these selected features are the most sensitive predictors to capture the characteristics of a subject's elevated blood glucose.

Results

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

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*
4	A_ratio_mean						

Table 3. The selected top features after the ensemble feature selection method. *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[62]. 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.

	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 (RBF)	**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

Table 4. The prediction results obtained from 10-fold cross-validation using various machine learning models.

ML Model Performance

Seven widely used machine learning algorithms, including the naive 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. Each model was fine-tuned accordingly and validated under the 10-fold cross-validation scheme. 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.

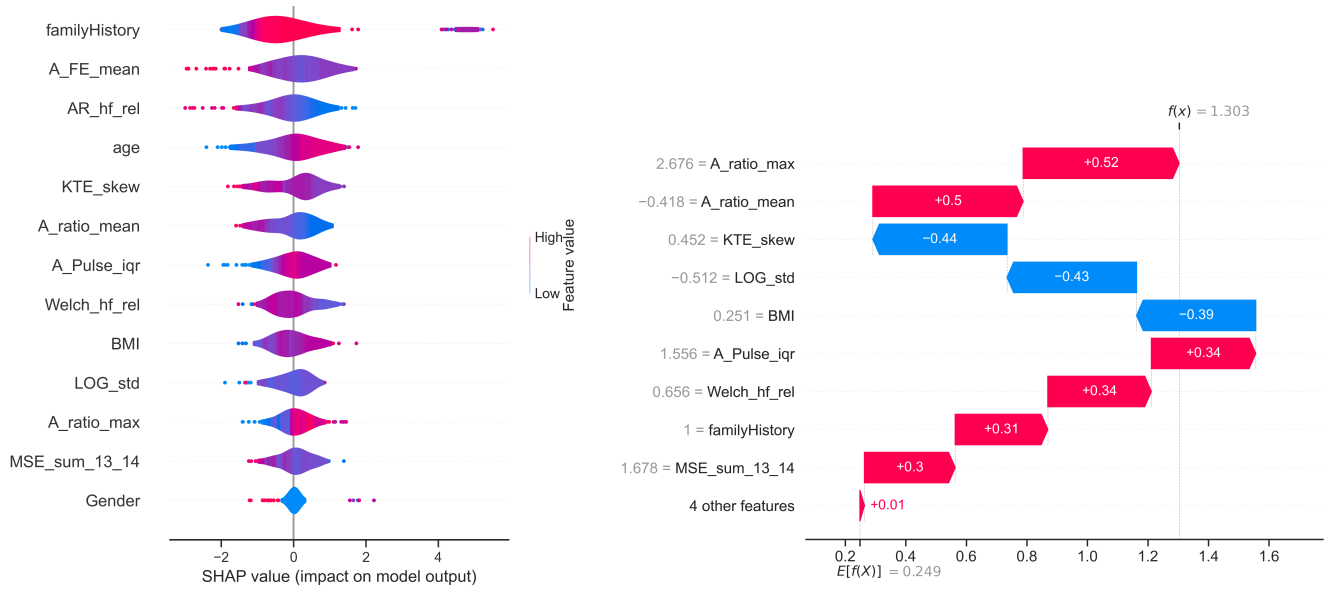
The prediction result from each model is reported with the mean and standard deviation of the evaluation metrics, and Table 4 shows the summary of the results. As shown in Table 4, SVM with the Radial Basis Function (RBF) kernel has shown the best prediction performance with an average accuracy of 84.7%, a sensitivity of 81.05%, a specificity of 88.35% and a precision of 87.51%. Most especially, the average G-mean of 84.54% and F-score of 84.03%.

Model interpretation using explainable AI approach

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[63,64]. 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. 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 chapter 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[62,65,66]. 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[67].

Figure 4 (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



(a) SHAP beeswarm plot.

(b) SHAP waterfall plot

Figure 4. The SHAP explanation plots indicate the association between the selected features and their impact on the predicted outcome.

predictions.

In Figure 4 (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[68].

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.

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.

Our study has limitations. Non-fasting blood glucose measurements were collected from subjects, and most of the participants were female. 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.

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[36]. 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.

Data availability

The data are not publicly available because they contain information that could compromise the participant's privacy and the authors' background IP. Sample data are available from the corresponding author upon reasonable request.

References

1. National Diabetes Data Group. Classification and diagnosis of diabetes mellitus and other categories of glucose intolerance. *Diabetes* **28**, 1039–1057 (1979).
2. Kerner, W. & Brückel, J. Definition, classification and diagnosis of diabetes mellitus. *Exp. Clin. Endocrinol. & Diabetes* **122**, 384–386 (2014).
3. American Diabetes Association. 2. classification and diagnosis of diabetes: Standards of medical care in diabetes-2020. *Diabetes care* **43**, S14–S31 (2020).
4. International Diabetes Federation. Idf diabetes atlas, 10th edn (2021).
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* **375**, 2215–2222 (2010).
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. *Sci. Reports 2020 10:1* **10**, 1–11 (2020).
7. Li, S., Wang, J., Zhang, B., Li, X. & Liu, Y. Diabetes mellitus and cause-specific mortality: A population-based study. *Diabetes & Metab. J.* **43**, 319 (2019).
8. Saran, R. *et al.* Us renal data system 2014 annual data report: Epidemiology of kidney disease in the united states. *Am. J. Kidney Dis.* **66**, A7 (2015).
9. Lau, L. H., Lew, J., Borschmann, K., Thijs, V. & Ekinci, E. I. Prevalence of diabetes and its effects on stroke outcomes: A meta-analysis and literature review. *J. Diabetes Investig.* **10**, 780–792 (2019).
10. Sun, H. *et al.* Idf diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes research clinical practice* **183**, 109119 (2022).
11. Huang, Y., Cai, X., Mai, W., Li, M. & Hu, Y. Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis. *BMJ (Clinical research ed.)* **355**, i5953 (2016).
12. Wong, M.-S. *et al.* The singapore impaired glucose tolerance follow-up study: does the ticking clock go backward as well as forward? *Diabetes care* **26**, 3024–30 (2003).
13. Tabák, A. G., Herder, C., Rathmann, W., Brunner, E. J. & Kivimäki, M. Prediabetes: a high-risk state for diabetes development. *Lancet (London, England)* **379**, 2279–90 (2012).
14. Force, U. P. S. T. *et al.* Screening for prediabetes and type 2 diabetes: Us preventive services task force recommendation statement. *JAMA* **326**, 736–743 (2021).

15. Misra, A. *et al.* Diabetes in developing countries. *J. diabetes* **11**, 522–539 (2019).

16. 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 medicine* **16**, e1002751 (2019).

17. García-Molina, L. *et al.* Improving type 2 diabetes mellitus glycaemic control through lifestyle modification implementing diet intervention: a systematic review and meta-analysis. *Eur. journal nutrition* **59**, 1313–1328 (2020).

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 Heal.* **18** (2021).

19. Tuso, P. Prediabetes and lifestyle modification: time to prevent a preventable disease. *The Perm. journal* **18**, 88–93 (2014).

20. Bansal, N. Prediabetes diagnosis and treatment: A review. *World journal diabetes* **6**, 296–303 (2015).

21. Magkos, F., Hjorth, M. F. & Astrup, A. Diet and exercise in the prevention and treatment of type 2 diabetes mellitus. *Nat. Rev. Endocrinol. 2020 16:10* **16**, 545–555 (2020).

22. Simmons, R. K., Griffin, S. J., Lauritzen, T. & 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* **60**, 2192–2199 (2017).

23. Heinemann, L. Finger pricking and pain: A never ending story. *J. Diabetes Sci. Technol.* **2**, 919–921 (2008).

24. Hambling, C. E., Seidu, S. I., Davies, M. J. & Khunti, K. Older people with type 2 diabetes, including those with chronic kidney disease or dementia, are commonly overtreated with sulfonylurea or insulin therapies. *Diabet. Medicine* **34**, 1219–1227 (2017).

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* **9**, e032037 (2019).

26. Vettoretti, M., Cappon, G., Acciaroli, G., Facchinetti, A. & Sparacino, G. Continuous glucose monitoring: Current use in diabetes management and possible future applications. *J. Diabetes Sci. Technol.* **12**, 1064–1071 (2018).

27. Funtanilla, V. D., Caliendo, T. & Hilas, O. Continuous glucose monitoring: A review of available systems. *P T* **44**, 550–553 (2019).

28. Robertson, S. L., Shaughnessy, A. F. & Slawson, D. C. Continuous glucose monitoring in type 2 diabetes is not ready for widespread adoption. *Am. Fam. Physician* **101**, 646–646 (2020).

29. Sabry, F., Eltaras, T., Labda, W., Alzoubi, K. & Malluhi, Q. Machine learning for healthcare wearable devices: The big picture. *J. Healthc. Eng.* **2022** (2022).

30. Patel, S., Park, H., Bonato, P., Chan, L. & Rodgers, M. A review of wearable sensors and systems with application in rehabilitation. *J. NeuroEngineering Rehabil.* **9**, 1–17 (2012).

31. Rodgers, M. M., Alon, G., Pai, V. M. & Conroy, R. S. Wearable technologies for active living and rehabilitation: Current research challenges and future opportunities. *J. rehabilitation assistive technologies engineering* **6**, 2055668319839607 (2019).

32. Xie, Y. *et al.* Integration of artificial intelligence, blockchain, and wearable technology for chronic disease management: A new paradigm in smart healthcare. *Curr. Med. Sci. 2021 41:6* **41**, 1123–1133 (2021).

33. Iqbal, S. M., Mahgoub, I., Du, E., Leavitt, M. A. & Asghar, W. Advances in healthcare wearable devices. *npj Flex. Electron. 2021 5:1* **5**, 1–14 (2021).

34. Monte-Moreno, E. Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques. *Artif. intelligence medicine* **53**, 127–38 (2011).

35. Rodin, D. *et al.* Comparative accuracy of optical sensor-based wearable system for non-invasive measurement of blood glucose concentration. *Clin. Biochem.* **65**, 15–20 (2019).

36. Zhang, G. *et al.* A noninvasive blood glucose monitoring system based on smartphone ppg signal processing and machine learning. *IEEE Transactions on Ind. Informatics* **16**, 7209–7218 (2020).

37. Challoner, A. V. & Ramsay, C. A. A photoelectric plethysmograph for the measurement of cutaneous blood flow. *Phys. Medicine Biol.* **19**, 317–328 (1974).

38. Zhao, D., Sun, Y., Wan, S. & Wang, F. Sfst: A robust framework for heart rate monitoring from photoplethysmography signals during physical activities. *Biomed. Signal Process. Control.* **33**, 316–324 (2017).

39. Schroeder, E. B. *et al.* Diabetes, glucose, insulin, and heart rate variability: The atherosclerosis risk in communities (aric) study. *Diabetes Care* **28**, 668–674 (2005).

40. Liang, Y., Elgendi, M., Chen, Z. & Ward, R. An optimal filter for short photoplethysmogram signals. *Sci. Data* **5**, 180076 (2018).

41. van Ravenswaaij-Arts, C. M., Kollée, L. A., Hopman, J. C., Stoelinga, G. B. & van Geijn, H. P. Heart rate variability. *Annals internal medicine* **118**, 436–47 (1993).

42. Xhyheri, B., Manfrini, O., Mazzolini, M., Pizzi, C. & Bugiardini, R. Heart rate variability today. *Prog. cardiovascular diseases* **55**, 321–31 (2012).

43. Thomas, B. L., Claassen, N., Becker, P. & Viljoen, M. Validity of commonly used heart rate variability markers of autonomic nervous system function. *Neuropsychobiology* **78**, 14–26 (2019).

44. Obrist, P. A. *Cardiovascular Psychophysiology* (Springer US, 1981).

45. Singh, N. *et al.* Heart rate variability: An old metric with new meaning in the era of using mhealth technologies for health and exercise training guidance. part one: Physiology and methods. *Arrhythmia & electrophysiology review* **7**, 193–198 (2018).

46. Prinsloo, G. E., Rauch, H. G. L. & Derman, W. E. A brief review and clinical application of heart rate variability biofeedback in sports, exercise, and rehabilitation medicine. *The Physician sportsmedicine* **42**, 88–99 (2014).

47. Billman, G. E., Huikuri, H. V., Sacha, J. & Trimmel, K. An introduction to heart rate variability: methodological considerations and clinical applications. *Front. physiology* **6**, 55 (2015).

48. Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H. & Koo, B.-H. Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry investigation* **15**, 235–245 (2018).

49. Taye, G. T., Hwang, H.-J. & Lim, K. M. Application of a convolutional neural network for predicting the occurrence of ventricular tachyarrhythmia using heart rate variability features. *Sci. reports* **10**, 6769 (2020).

50. Mosley, E. & Laborde, S. A scoping review of heart rate variability in sport and exercise psychology. *Int. Rev. Sport Exerc. Psychol.* 1–75 (2022).

51. Shaffer, F. & Ginsberg, J. P. An overview of heart rate variability metrics and norms. *Front. Public Heal.* **5** (2017).

52. Valensi, P. *et al.* Influence of blood glucose on heart rate and cardiac autonomic function. the desir study. *Diabet. medicine: a journal Br. Diabet. Assoc.* **28**, 440–9 (2011).

53. Inamdar, A. Correlation between fasting heart rate and fasting plasma glucose level in rural indians. *Eur. Hear. J.* **43** (2022).

54. Gupta, S. S., Kwon, T.-H., Hossain, S. & Kim, K.-D. Towards non-invasive blood glucose measurement using machine learning: An all-purpose ppg system design. *Biomed. Signal Process. Control.* **68**, 102706 (2021).

55. Nirala, N., Periyasamy, R., Singh, B. K. & Kumar, A. Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine. *Biocybern. Biomed. Eng.* **39**, 38–51 (2019).

56. Philip, L. A., Rajasekaran, K. & Jothi, E. S. J. Continous monitoring of blood glucose using photophlythesmograph signal. vol. 2017-Janua, 187–191 (IEEE, 2017).

57. Moraes, J. L. *et al.* Advances in photopletysmography signal analysis for biomedical applications. *Sensors (Basel, Switzerland)* **18** (2018).

58. Habbu, S., Dale, M. & Ghongade, R. Estimation of blood glucose by non-invasive method using photoplethysmography. *Sadhana - Acad. Proc. Eng. Sci.* **44**, 1–14 (2019).

59. Hina, A., Nadeem, H. & Saadeh, W. A single led photoplethysmography-based noninvasive glucose monitoring prototype system. vol. 2019-May, 1–5 (IEEE, 2019).

60. Castiglioni, P. *et al.* Assessing sample entropy of physiological signals by the norm component matrix algorithm: application on muscular signals during isometric contraction. *Annu. Int. Conf. IEEE Eng. Medicine Biol. Soc. IEEE Eng. Medicine Biol. Soc. Annu. Int. Conf.* **2013**, 5053–6 (2013).

61. Delgado-Bonal, A. & Marshak, A. Approximate entropy and sample entropy: A comprehensive tutorial. *Entropy (Basel, Switzerland)* **21** (2019).

62. Nordström, A., Hadrévi, J., Olsson, T., Franks, P. W. & Nordström, P. Higher prevalence of type 2 diabetes in men than in women is associated with differences in visceral fat mass. *The J. clinical endocrinology metabolism* **101**, 3740–3746 (2016).

63. Shi, B. *et al.* Convolutional neural network for freezing of gait detection leveraging the continuous wavelet transform on lower extremities wearable sensors data. vol. 2020-July, 5410–5415 (Institute of Electrical and Electronics Engineers Inc., 2020).

64. Shi, B. *et al.* Detection of freezing of gait using convolutional neural networks and data from lower limb motion sensors. *IEEE Transactions on Biomed. Eng.* **69**, 2256–2267 (2022).

65. Gulcher, J. & Stefansson, K. Clinical risk factors, dna variants, and the development of type 2 diabetes. *The New Engl. journal medicine* **360**, 1360; author reply 1361 (2009).

66. Centers for Disease Control and Prevention. Diabetes risk factors - cdc (2022).

67. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. 4768–4777 (Curran Associates Inc., 2017).

68. Zhou, X.-H., Obuchowski, N. A. & McClish, D. K. *Statistical Methods in Diagnostic Medicine* (John Wiley & Sons, Inc., 2011).

Author contributions statement

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. 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.

Competing interests

We would like to disclose that B.H.S., M.S., J.W. 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.

Funding

This research was sponsored by Actxa Pte. Ltd., but data collection was done independently at KK Women's and Children's Hospital, Singapore.