NON-INVASIVE PPG-BASED ESTIMATION OF BLOOD GLUCOSE LEVEL

Enikö Vargová[1], Andrea Němcová[1], Zuzana Nováková[2]

[1]Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czech Republic

[2]Department of Physiology, Faculty of Medicine, Masaryk University, Brno, Czech Republic

Abstract

This paper focuses on non-invasive blood glucose determination using photoplethysmographic (PPG) signals, which is crucial for managing diabetes. Diabetes stands as one of the world's major chronic diseases. Untreated diabetes frequently leads to fatalities. Current self-monitoring techniques for measuring diabetes require invasive procedures such as blood or bodily fluid sampling, which may be very uncomfortable. Hence, there is an opportunity for non-invasive blood glucose monitoring through smart devices capable of measuring PPG signals. The primary goal of this research was to propose methods for glycemic classification into two groups (low and high glycemia) and to predict specific glycemia values using machine learning techniques. Two datasets were created by measuring PPG signals from 16 individuals using two different smart devices – a smart wristband and a smartphone. Simultaneously, the reference blood glucose levels were invasively measured using a glucometer. The PPG signals were preprocessed, and 27 different features were extracted. With the use of feature selection, only 10 relevant features were chosen. Numerous machine learning models were developed. Random Forest (RF) and Support Vector Machine (SVM) with the radial basis function (RBF) kernel performed best in classifying PPG signals into two groups. These models achieved an accuracy of 76% (SVM) and 75% (RF) on the smart wristband test dataset. The functionality of the proposed models was then verified on the smartphone test dataset, where both models achieved similar accuracy: 74% (SVM) and 75% (RF). For predicting specific glycemia values, RF performed best. Mean Absolute Error (MAE) was 1.25 mmol/l on the smart wristband test dataset and 1.37 mmol/l on the smartphone test dataset.

Keywords

PPG, diabetes, glycemia, smart devices, smartphone, classification, prediction

Introduction

Diabetes is a major worldwide health issue caused by a relative or absolute lack of insulin. The International Diabetes Federation (IDF) reports that around 537 million people aged 20–79 have diabetes. Untreated diabetes can lead to death [1].

It is essential for individuals with diabetes to monitor their blood glucose level (BGL) and keep it within the desired range. Currently, clinical practice relies only on invasive methods for blood glucose estimation, which has several drawbacks. Patients experience both physical and mental discomfort, and there is a notable risk of infection associated with this invasive approach [2].

The primary aim of this paper is to explore the potential of non-invasive measuring BGL using photoplethysmographic (PPG) signals measured by smart devices such as smart wristband (Empatica) and a smartphone. The paper aims to classify glycemia into two groups (low and high glycemia) and predict specific glycemia values using machine learning techniques.

In recent years, PPG signals have become increasingly popular for estimating physiological parameters such as heart rate (HR), blood pressure (BP), or oxygen saturation (SpO_2), perfusion index (PI) or vessel compliance. These signals are one of the most used signals in the field of health assessment through wearables. Considering the rapid expansion of smart devices, there exists a significant opportunity to extend health monitoring to a broader population [3].

Utilizing smart devices in healthcare comes with numerous benefits, such as cost-effectiveness, ease of use, the ability to monitor health in real-time, and suitability for extended self-monitoring at home. Moreover, smart devices make it convenient to

Lekar a technika – Clinician and Technology 2023, vol. 53(1), pp. 19–24, DOI: 10.14311/CTJ.2023.1.04
ISSN 0301-5491 (Print), ISSN 2336-5552 (Online)

remotely share measured data and results with healthcare providers [4].

Photoplethysmography

Photoplethysmography (PPG) is a non-invasive technique that measures changes in blood volume within the microvascular tissue of the skin by utilizing optical properties. While this method has been in use for several years, it has gained notable popularity in the past two decades due to advancements in wearable technology. PPG signals can be obtained from well-vascularized areas of the body, typically from the finger (using a smartphone, smart ring), earlobe, wrist (smartwatches), or foot. The advantages of this method include simplicity, accessibility, and the low cost of wearable electronics for PPG sensing [5–7].

The PPG signal is acquired using a light source, such as Light Emitting Diode (LED), which illuminates the tissue in the desired area. In most cases, infrared LEDs are used because there is a small difference in light absorption between oxygenated and deoxygenated blood. The second fundamental component in this measurement is a photodetector, which can capture light that has either passed through the tissue (transmission mode) or light that has been reflected (reflection mode). In the transmission mode, the monitored area, such as a finger, is positioned between the light source and the photodetector. The reflection mode, on the other hand, involves placing the light source and photodetector side by side [5, 7, 8].

In general, the PPG signal consists of a pulsatile component (AC) and a non-pulsatile component (DC). The AC component originates from the heart's rhythmic changes in blood volume associated with each heartbeat. On the other hand, the non-pulsatile component, often referred to as the direct current (DC) component, changes very slowly, contains low-frequency components, and is affected by factors like respiration, thermoregulation, and possible vasomotor activity [3, 6, 8].

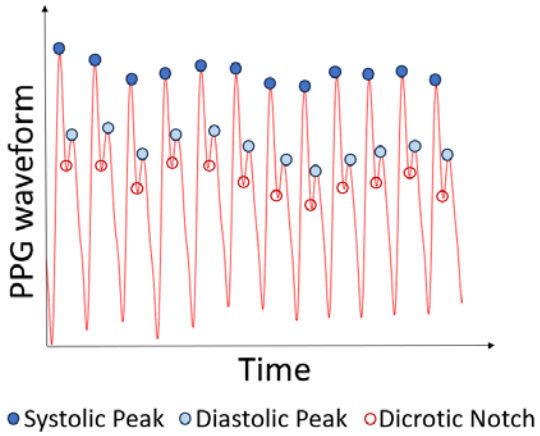

Fig. 1: PPG waveform.

In a typical PPG waveform of healthy individuals, a systolic peak, a diastolic peak, and a dicrotic notch are present (Fig. 1). The systolic peak is the highest point on the PPG waveform, corresponding to the maximum blood volume in the microvascular bed during systole. Diastolic peak reflects the lowest blood volume in the microvascular bed during diastole. In individuals with healthy arteries, the PPG waveform can exhibit a dicrotic notch, which varies with vascular compliance and vascular tone [6].

Materials and methods

Data recording

For this experiment, two datasets of PPG signals were created. The first dataset came from the Empatica E4 wristband (Empatica Inc., United States). The PPG signal derived from Empatica is generated using a patented algorithm that combines both green and red light. The signals are sampled at a frequency of 64 Hz. The second one was measured using a Samsung Galaxy Note 20 smartphone (Samsung Electronics, Vietnam). The video resolution was set to 1080×1920 pixels, and the frame rate was 30 frames per second.

Sixteen people (11 females and 5 males) with ages ranging from 22 to 79, volunteered for the study. Out of these participants, 5 had diabetes or prediabetes. All participants provided written informed consent, and the research was approved by the Ethical Committee Faculty of Electrical Engineering and Communication for Biomedical Research, Brno University of Technology.

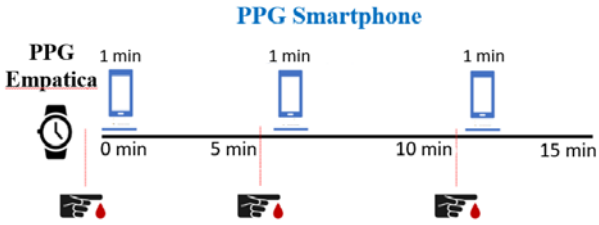

Fig. 2: Measurement scheme: Acquisition of a 15-min PPG signal using the Empatica wristband, and three one-min PPG recordings using a smartphone. Reference values were collected three times, each time after a 5-min interval.

The measurement process is depicted in Fig. 2. Each measurement lasted for 15 minutes. Throughout the measurement, the PPG signal was continuously recorded using the Empatica smart wristband. Additionally, one-minute PPG signals were measured three times using a smartphone during the whole measurement. The participants placed their finger on the smartphone's camera, covering both the lens and the LED light source. From the video of the finger, the

Lekar a technika – Clinician and Technology 2023, vol. 53(1), pp. 19–24, DOI: 10.14311/CTJ.2023.1.04
ISSN 0301-5491 (Print), ISSN 2336-5552 (Online)

average red component corresponding to the raw PPG signal was extracted. Simultaneously, the reference BGL was measured three times using a certified Fora Diamond Mini glucometer. To increase the amount of data obtained from individuals with higher BGL, diabetic participants underwent the entire measurement process twice. The measured BGL ranged from 4.3 to 13.2 mmol/l.

Data preprocessing

To increase the amount of data for glycemic classification and prediction, a 15-min measurement from a smart wristband was divided into three separate 5-min segments, each associated with three reference glycemia values. These 5-min segments were shortened to 4.5 min for individuals with diabetes or prediabetes and to 2.5 min for healthy individuals to ensure an approximately even distribution of data in both classification groups. All shortened signals were divided into 10-s segments (Table 1). The classification threshold for dividing the data into two groups was set at 7.2 mmol/l, according to the glucometer manual.

The video captured using the smartphone camera was converted into a PPG signal. The red video channel was chosen, and an average value was calculated from each video frame. Additionally, the PPG signals derived from the smartphone were divided into 10-s segments.

Table 1: The number of 10-second records and their distribution into two classification groups.

Smart wristband		
number of 10-s records	low glycemia level	high glycemia level
1143	537	606
Smartphone		
number of 10-s records	low glycemia level	high glycemia level
129	60	69

PPG signals measured by the smartphone are often of lower quality compared to wearables due to potential finger movements over the camera lens and pressure variations. For this reason, signal quality was assessed using an algorithm described in [9]. The algorithm is based on the computation of two features: perfusion and Shannon entropy, and the classification of data into high-quality/low-quality groups using a non-hierarchical k-means clustering method. Only high-quality signals were used for further analysis.

All 10-s PPG signals from both devices were filtered and normalized using max-min normalization. Signals measured by smartphone were inverted because they involve reflected light during acquisition. PPG signals obtained from the Empatica wristband are already inverted. Noise was reduced using a second-order Butterworth band-pass filter ranging from 36 beats per minute (bpm) to 216 bpm (Fig. 3).

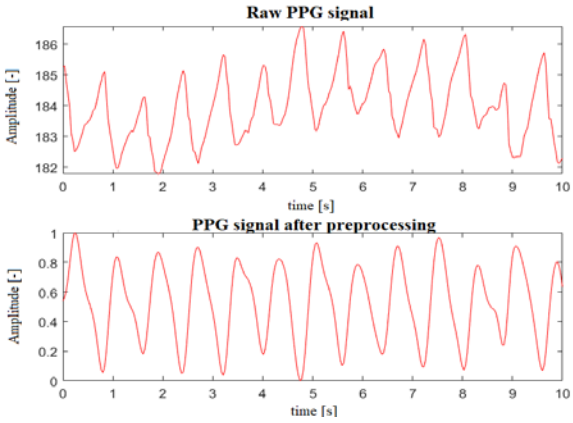

Fig. 3: PPG signal measured by a smartphone before and after preprocessing.

Feature extraction

A total of 27 different features were extracted from the 10-s PPG signals. Some of these features are based on Heart Rate Variability (HRV) analysis, as it has been shown that diabetics exhibit reduced HRV due to decreased parasympathetic activity [10]. Additional features are derived from the PPG waveform morphology, influenced by the varying light absorption of blood with different glucose levels.

Outliers were replaced, all features were normalized using Z-score, and a statistical analysis of the features was conducted. Initially, a normality test was applied to each feature. In cases where the data followed a normal distribution, an unpaired t-test was applied. If the data did not have a normal distribution, the non-parametric Mann-Whitney test was used. The goal of these tests was to identify features that were not statistically significant for classifying the data into the two mentioned groups. In this manner, only 4 features were excluded.

To further reduce the number of features, a selection method called "Maximum Relevance and Minimum Redundancy" (mRMR) was applied to the remaining set of 23 features, reducing the feature space to 10 features (Table 2).

Glycemic classification and prediction

Measured PPG signals are classified into two groups (low/high blood glucose) using a classification threshold of 7.2 mmol/l. Several different machine learning methods were tested for this purpose, including logistic regression (LR), k-nearest neighbors (KNN), support vector machine (SVM), and random forest (RF).

Lekar a technika – Clinician and Technology 2023, vol. 53(1), pp. 19–24, DOI: 10.14311/CTJ.2023.1.04
ISSN 0301-5491 (Print), ISSN 2336-5552 (Online)

Table 2: Description of selected features by mRMR.

Features	Description
kurtosis	describes the distribution of a dataset around the mean and tells us whether the data is distributed symmetrically or not
skewness	is a measurement of the distortion of symmetrical distribution or asymmetry in a data set
entropy	measures the degree to which a probability distribution deviates from a uniform or equal distribution
peak to RMS	returns the ratio of the largest absolute value in signal to the root-mean-square (RMS) value of signal
shortest PP interval	shortest distance of peak-to-peak (PP) intervals
TKEO ratio	the ratio of the minimum and maximum values of the envelope obtained using the Teager-Kaiser Energy Operator (TKEO)
CV	the coefficient of variation (CV) is the standard deviation of a set of RR (PP) intervals divided by the mean of that set of intervals
SDNN	the standard deviation of NN (PP) interval lengths
pNN50	the ratio of NN50 to the total number of NN intervals in the recording
min/max of 1st derivative	the ratio of the minimum to the maximum of the 1st derivative of the PPG

Training and testing were conducted using data obtained from the Empatica smart wristband. Cross-validation (k = 4) was performed to find optimal model parameters and obtain more reliable results. In each iteration of cross-validation, the data was split into training (75%) and testing (25%) sets. It was ensured that data from the same patient did not appear in both the training and testing sets simultaneously. The results represent the average values from all cross-validation iterations. The functionality of the best models was verified using data measured by a different device—a smartphone.

To predict the specific blood glucose value from PPG signals, various machine learning methods were tested for the regression task, including KNN, SVM, and RF.

The training and testing procedures were again conducted using data from the Empatica smart wristband. In the training process, cross-validation (k = 4) was used, maintaining the same data split ratio for training and testing sets as used in the classification task. The mean absolute error (MAE) was determined as the average of all cross-validation iterations. Subsequently, the best model (RF) was further validated using data from smartphone.

Results

The classification task was evaluated using two metrics: accuracy (ACC) and F1 score (F1), which represents the harmonic mean of precision and recall. Table 3 summarizes the achieved results of the best classification models (SVM with RBF kernel and RF) for the training and testing datasets from the smart wristband.

Table 3: Results of blood glucose classification into two groups (msp – minimum sample leaf, nf – number of trees).

Smart wristband					
	training set		**testing set**		
Model	ACC (%)	F1 (%)	ACC (%)	F1 (%)	AUC
SVM RBF	86.4	86.3	75.6	75.6	0.76
RF *nf* = 90 *msp* = 9	89.2	89.7	74.9	74.1	0.76
Smartphone					
	testing set				
Model	ACC (%)	F1 (%)	AUC		
SVM RBF	74.0	73.1	0.71		
RF *nf* = 90 *msp* = 9	74.5	76.7	0.71		

Additionally, Table 3 shows the optimal hyperparameters used in RF model: *msp* – minimum sample leaf and *nf* – number of trees. Subsequently, the functionality of both models was validated on the dataset obtained from the smartphone. To assess the classifiers' quality, the Receiver Operating Characteristic (ROC) curve, and the Area under the Curve (AUC) were used. The average AUC value for each classifier is shown in Table 3.

Table 4 summarizes MAE of the best model (RF) with optimal parameter settings for predicting specific

Lekar a technika – Clinician and Technology 2023, vol. 53(1), pp. 19–24, DOI: 10.14311/CTJ.2023.1.04
ISSN 0301-5491 (Print), ISSN 2336-5552 (Online)

BGL on the training and testing datasets from the smart wristband. The functionality of the proposed model was validated using data measured by the smartphone.

Table 4: Results of blood glucose prediction (msp – minimum sample leaf, nf – number of trees).

Smart wristband		
	training set	**testing set**
Model	MAE (mmol/l)	MAE (mmol/l)
RF *nf* = 80 *msp* = 45	0.86	1.25
Smartphone		
	testing set	
Model	MAE (mmol/l)	
RF *nf* = 80 *msp* = 45	1.37	

Discussion

The results of the proposed methods are comparable to the findings of other authors. For instance, Zhang et al. [11] utilized KNN methods for classifying BGL from PPG signals and achieved an accuracy of 74%. As for predicting specific BGL from PPG signals, Nie et al. [12] used RF and achieved MAE of 1.72 mmol/l. Conversely, Manurung et al. [13] used deep learning, which generally yields better results, achieving a result of 0.32 mmol/l. Due to a lack of data, this approach was not feasible. The observed difference in MAE between the training and testing sets can be attributed to a patient with an exceptionally low BGL compared to others, included in the testing set. This discrepancy highlights a limitation of the study, including the number of measured subjects (16) and the range of measured BGL (4.3–13.2 mmol/l). It is important to note that there is currently no freely available database of PPG signals containing reference BGL, so all data had to be measured. Despite the limitations, this work demonstrates good results in both the classification task and the prediction of specific BGL, which are comparable to the state-of-the-art methods, highlighting the potential for using smart devices and PPG signals for non-invasive blood glucose estimation.

Conclusion

This paper focuses on non-invasive estimation of BGL from PPG signals. Two databases of PPG signals were created. One contains data measured by the Empatica smart wristband. The second database consists of PPG signals recorded by a smartphone. It is only used to validate the functionality of the proposed algorithms on data obtained from a different device. Dataset preparation involved assessing the quality of PPG signals obtained from the smartphone, which are often of lower quality. All PPG signals were normalized and preprocessed, and relevant features were extracted. Data collected from the smart wristband were used for training and testing the algorithms proposed for both classification and prediction of BGL. Various machine learning models were evaluated for classification and prediction, with the RF and SVM with an RBF kernel models achieving the best results for classifying PPG records into two groups (low and high glycemia). These models achieved Acc of 76% (SVM) and 75% (RF) on the testing dataset from the smart wristband. The functionality of the proposed models was then validated on the testing dataset from the smartphone, where both models achieved similar results: 74% (SVM) and 75% (RF). RF achieved the lowest MAE = 1.25 mmol/l on the testing dataset from the smart wristband and 1.37 mmol/l on the testing dataset from the smartphone. The obtained results for both classification and regression tasks are comparable to those of other authors, highlighting the potential for using smart devices and PPG signals for non-invasive BGL estimation.

Acknowledgment

This study was supported by project FEKT-S-23-8196 and by the Specific University Research Grants of Masaryk University MUNI/A/1343/2022 provided by the Ministry of Education, Youth and Sports of the Czech Republic.

References

[1] Deng H, Zhang L, Xie Y, Mo S. Research on Estimation of Blood Glucose Based on PPG and Deep Neural Networks. IOP conference series: Earth and Environmental Science. 2021 Mar 1;693(1):012046–6. DOI: 10.1088/1755-1315/693/1/012046

[2] IDF diabetes atlas [Internet]. Brussels: International Diabetes Federation. Available from: https://diabetesatlas.org/

[3] Castaneda D, Esparza A, Ghamari M, Soltanpur C, Nazeran H. A review on wearable photoplethysmography sensors and their potential future applications in health care. International Journal of Biosensors & Bioelectronics. 2018 Aug 6;4(4):195–202. DOI: 10.15406/ijbsbe.2018.04.00125

[4] Senbekov M, Saliev T, Bukeyeva Z, Almabayeva A, Zhanaliyeva M, Aitenova N, et al. The Recent Progress and Applications of Digital Technologies in Healthcare: A Review. International Journal of Telemedicine and Applications. 2020 Dec 4;2020:8830200. DOI: 10.1155/2020/8830200

[5] Siddiqui SA, Zhang Y, Feng Z, Kos A. A Pulse Rate Estimation Algorithm Using PPG and Smartphone Camera. Journal of Medical Systems. 2016 Apr 11;40:126. DOI: 10.1007/s10916-016-0485-6

[6] Park J, Seok HS, Kim SS, Shin H. Photoplethysmogram Analysis and Applications: An Integrative Review. Frontiers in Physiology. 2022 Mar 1;12:808451. DOI: 10.3389/fphys.2021.808451

[7] Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement. 2007 Feb 20;28(3):R1–39. DOI: 10.1088/0967-3334/28/3/R01

[8] Moraes JL, Rocha MX, Vasconcelos GG, Vasconcelos Filho JE, de Albuquerque VH, Alexandria AR. Advances in Photopletysmography Signal Analysis for Biomedical Applications. Sensors. 2018 Jun 9;18(6):1894. DOI: 10.3390/s18061894

[9] Vargová E. Stanovení kvality a odhad tepové frekvence ze signálu PPG [bakalářská práce]. [Brno]: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, Ústav biomedicínského inženýrství; 2021. 39 p.

[10] Benichou T, Pereira B, Mermillod M, Tauveron I, Pfabigan D, Maqdasy S, et al. Heart rate variability in type 2 diabetes mellitus: A systematic review and meta–analysis. PLOS ONE. 2018 Apr 2;13(4):e0195166. DOI: 10.1371/journal.pone.0195166

[11] Zhang G, Mei Z, Zhang Y, Ma X, Lo B, Chen D, et al. A Non-invasive Blood Glucose Monitoring System Based on Smartphone PPG Signal Processing and Machine Learning. IEEE Transactions on Industrial Informatics. 2020 Feb 20;16(11):7209–18.. DOI: 10.1109/TII.2020.2975222

[12] Nie Z, Rong M, Li K. Blood glucose prediction based on imagingphotoplethysmography in combination with Machine learning. Biomedical Signal Processing and Control. 2023 Jan 1;79:104179. DOI: 10.1016/j.bspc.2022.104179

[13] Manurung BE, Munggaran HR, Ramadhan GF, Koesoema AP. Non-Invasive Blood Glucose Monitoring using Near-Infrared Spectroscopy based on Internet of Things using Machine Learning. In: IEEE Region 10 Humanitarian Technology Conference (R10-HTC); 2019 Nov 12–14; Depok, Indonesia. p. 5–11. DOI: 10.1109/R10-HTC47129.2019.9042479

Ing. Enikö Vargová
Department of Biomedical Engineering
Faculty of Electrical Engineering and Communication
Brno University of Technology
Technická 3058/10, 616 00, Brno – Královo Pole

E-mail: xvargo02@vutbr.cz
Phone: +420 775 990 927

Lekar a technika – Clinician and Technology 2023, vol. 53(1), pp. 19–24, DOI: 10.14311/CTJ.2023.1.04
ISSN 0301-5491 (Print), ISSN 2336-5552 (Online)