

Machine Learning-Based Diabetes Detection Using Photoplethysmography Signal Features

Filipe A. C. Oliveira¹, Felipe M. Dias¹, Marcelo A. F. Toledo¹,
Diego A. C. Cardenas¹, Douglas A. Almeida¹, Estela Ribeiro¹,
Jose E. Krieger¹, Marco A. Gutierrez¹

¹Heart Institute (InCor), Clinics Hospital University of Sao Paulo Medical School
Sao Paulo - SP - Brazil

{filipe.acoliveira, f.dias, marcelo.arruda}@hc.fm.usp.br

{diego.cardona, douglas.andrade, estela.ribeiro}@hc.fm.usp.br

j.krieger@hc.fm.usp.br, marco.gutierrez@incor.usp.br

Abstract. *Diabetes is a chronic condition which prevention and control is done mostly by minimally invasive devices. In this work, we propose a non-invasive method based on photoplethysmography (PPG) for cost-effective and discomfort-free diabetes detection and prevention. We used PPG signal features and patient metadata from a public dataset for classifying subjects as Diabetic or non-Diabetic. The Logistic Regression and eXtreme Gradient Boosting algorithms were evaluated using a five-fold cross validation approach and achieved a mean AUC of 0.79 ± 0.15 and 0.73 ± 0.17 , respectively. Our results align with existing literature, supporting the use of machine learning techniques for developing non-invasive diabetes detection and prevention devices.*

1. Introduction

Diabetes is a chronic disease characterized by insufficient insulin release by the pancreas or lack of tissue response, leading to higher blood glucose levels (BGL) [DeFronzo et al. 1992]. When undetected or untreated, this higher glucose concentration may lead to several vascular illnesses, such as kidney and heart diseases. Diabetes management includes the prevention of such illnesses by constant BGL self-monitoring [Kirk and Stegner 2010].

Several devices for assessing the patient’s BGL are either invasive based on fasting plasma glucose test, which requires a blood sample collection by a healthcare professional, or minimally invasive through finger prick blood test and continuous glucose monitors (CGM) for self-monitoring at home. However, all these methods require skin perforation, and only the CGM provides continuous measurements. Due to the high cost, discomfort, and risk provided by minimally invasive devices, frequent BGL monitoring is not implemented at large. A non-invasive approach to detect diabetes would be valuable for the identification of early stages of diabetes in healthy and pre-Diabetic subjects, contributing to global efforts in preventing this disease and instrumental to assist monitor therapy interventions [LaMonte et al. 2005].

The Photoplethysmography (PPG) is one of the non-invasive approaches that has been studied for the detection of diabetes and pre-diabetes [Zanelli et al. 2022]. The PPG

is an optical technique for measuring changes in blood volume in the microvascular bed of tissues. This technique typically involves emitting light on the tissue and measuring changes in light absorption, reflection, or scattering to determine changes in blood volume [Mejía-Mejía et al. 2022]. Moreover, the quantification of certain molecules in the blood can be performed by employing light of particular molecules' peak absorption frequencies [Mukkamala et al. 2022]. The canonical case is oximetry (SpO₂). Besides that, PPG waveform can be used to measure other important cardiovascular and respiratory parameters, such as heart rate (HR), respiration rate, and blood pressure [Nitzan and Ovadia-Blechman 2022].

PPG can be applied as an indirect measure of a subject's hemodynamics information, conveying information pertaining to the condition of the blood vessel, such as the arterial stiffness, which is found to be elevated in individuals with diabetes [Pilt et al. 2013]. Other work show that the area under a PPG pulse is reduced when the level of HbA_{1c} (glycated hemoglobin) level increase [Usman et al. 2011]. The HbA_{1c} provides an estimation of a person's average blood glucose levels over the past two to three months and is used for diabetes diagnostic and management.

An advantage of PPG-based methods for assessing BGL is that they can provide continuous non-invasive measurements. These are not only important for monitoring, but also for detecting diabetes in its early stages of development. In [Reddy. et al. 2017], they applied an SVM method for predicting diabetes using features related to heart rate variability (HRV) and PPG morphology, using a fingertip pulse oximeter private dataset. In [Hettiarachchi and Chitraranjan 2019], they proposed to predict diabetes based on morphological features extracted from PPG signals from a public dataset [Liang et al. 2018] using a Linear Discriminant Analysis method. In [Moreno et al. 2017], they proposed a method for screening patients for diabetes by extracting features from the signal of a pulse oximeter. The features served as input for different classification algorithms, such as random forest and gradient boosting. [Srinivasan and Foroozan 2021] made an analysis in the frequency domain using a convolutional neural network (CNN) with 30 seconds PPG scalograms and metadata over a large dataset. More recently, [Zanelli et al. 2023] proposed a transfer learning approach to detect diabetes through 1-second raw PPG signals.

In this work, we presented a method to classify non-Diabetic and Diabetic patients using a set of features extracted from PPG signals and metadata for training machine learning models using Logistic Regression (LR) and eXtreme Gradient Boosting (XGBoost) [Chen and Guestrin 2016b] algorithms. We aimed to correlate the PPG signals morphological characteristics and the subject's diabetes status adjusted by the individual's metadata information. The method proposed in this paper utilizes shorter PPG signals, each segment lasting 2.1 seconds. This approach enables easy integration into various wearable devices, while also potentially offering benefits such as reduced computational costs and processing times.

2. Materials and Methods

In this section, the description of the publicly available dataset used to classify the PPG signals into non-Diabetic or Diabetic patients is presented and also the data selection criteria. Next, we describe the proposed methodology for Diabetes classification, based on signal preprocessing and feature extraction of PPG signals, and the use of two algorithms

for the classification step. Fig. 1 shows the proposed methodology for the classification between non-Diabetic and Diabetic groups.

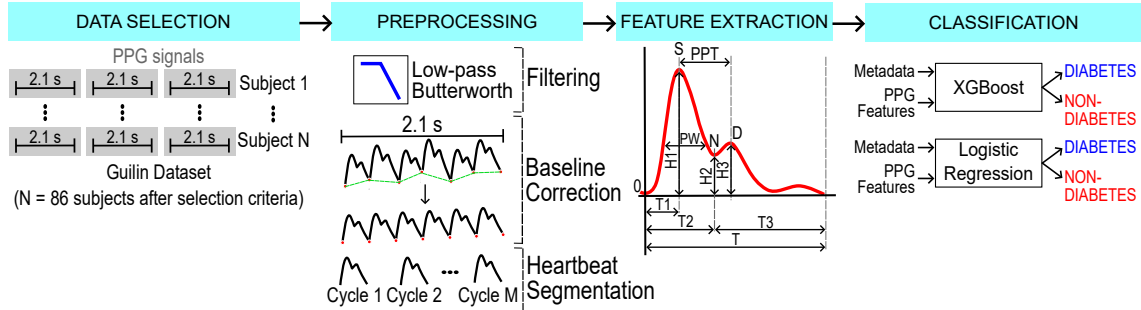


Figure 1. The proposed methodology for the classification of patients into Diabetes or non-Diabetes groups.

2.1. Data Selection

We used a publicly available dataset provided by a study that contains physiological information and PPG data from 219 patients of Guilin People’s Hospital in China [Liang et al. 2018]. The PPG signal was collected from the fingertip using the transmission method in infrared wavelength (905 nm) at a sample rate of 1 kHz. The PPG signal was collected from resting patients, where three 2.1-second PPG segments were captured from each patient.

The dataset contains information about three different classes of diseases: diabetes, hypertension, and cerebrovascular diseases, along with metadata information such as age, weight, height, body mass index (BMI), arterial blood pressure (ABP), and an unique patient ID number.

For this study, we divided the patients into two groups: non-Diabetic and Diabetic patients. We kept 59 healthy subjects in the non-Diabetic group by removing all patients with any stage of hypertension or cerebrovascular disease and all 38 patients diagnosed with diabetes were kept in the other group. It was not feasible excluding diabetic patients diagnosed with other diseases since only 9 subjects had no other disclaimed pathology, although none of the diabetic group present any cerebrovascular disease. Due to segmentation errors and noisy signals, 11 subjects were excluded from the analysis. Table 1 show a description of the data used in this study.

Table 1. Summary of the data used.

Class	Subjects	Male	Cycles	Age	Height	Weight	HR	BMI
<i>Non-Diabetic</i>	54	22	281	45 ± 16	161 ± 8	56 ± 11	73 ± 11	22 ± 4
<i>Diabetic</i>	32	14	172	59 ± 12	160 ± 8	62 ± 12	74 ± 11	24 ± 4
<i>Total</i>	86	36	453	50 ± 16	161 ± 8	59 ± 11	74 ± 11	23 ± 4

2.2. Preprocessing

To remove high-frequency artifacts in the PPG, we preprocessed each signal segment with a 6th-order Butterworth low-pass filter with a cutoff frequency of 16 Hz. During the acquisition of PPG signals, different artifacts may cause baseline oscillations. To address

this issue, we applied the Fitting-based Sliding Window (FSW) algorithm instead of a high-pass-filter to remove the baseline, as the high-pass filter may eliminate important low-frequency characteristics present in the signal [Zhang et al. 2020].

The FSW algorithm involves a sliding window that identifies the local minimums between each cycle (heartbeat) in the signal segments and uses them to fit the baseline. The fitted line was subtracted from the signal, removing the baseline fluctuations while preserving low-frequency characteristics. The valleys between the cycles were also used to segment the PPG cycles in each signal segment. The FSW algorithm details are described in [Zhang et al. 2020].

We performed a visual inspection on the cycles extracted from PPG signals of all patients in the study. During this process, six Diabetic and five non-Diabetic patients were removed from the experiment due to segmentation flaws or presence of very noisy signals.

2.3. Feature extraction

Each PPG cycle segmented by the FSW served as input for the feature extraction algorithm. We used the same features presented in [Costa et al. 2023], resulting in 104 features. The features were computed from each signal cycle (heartbeat) and its corresponding first and second derivatives. The features included a variety of values such as pulse width, area, intervals, peak-to-peak interval, and systolic amplitude. A detailed description of the features used here can be found in [Costa et al. 2023], [Chowdhury et al. 2020], [El-Hajj and Kyriacou 2021], and [Lin et al. 2020].

The metadata (sex, age, height, weight, heart rate, and BMI) was used as features in the classification algorithm along with the PPG signal features. However, the data related to blood pressure was not included due to its strong relation to hypertensive and pre-hypertensive patients. This could create a potential bias in the classification algorithm as high blood pressure could be falsely associated with the presence of Diabetes, since most Diabetic patients included in the experiment have a degree of hypertension.

This feature extraction step resulted in 453 sets of features that were used to perform the detection of Diabetes on PPG signals, being 6 features in each set related to the metadata information and 104 related to features extracted from the PPG signal cycles.

2.4. Classification

We employed two well-known algorithms commonly used in machine learning classification tasks, the Logistic Regression and the XGBoost [Chen and Guestrin 2016a] algorithms. XGBoost is known for its high accuracy, ability to handle complex data, and scalability, but it's less interpretable. Logistic Regression, on the other hand, is a simpler model with better interpretability, but may not perform as well in complex datasets.

L_1 penalty and *LIBLINEAR* solver were used as hyperparameters for the LR algorithm. For the XGBoost algorithm we used the tree booster with learning rate of 0.1. Since the input data is unbalanced and their proportion may vary within each fold, the weights for each class were adjusted according to the number of samples representing each class for both classifiers.

The performance of the LR and the XGBoost models was evaluated using statistical metrics: Accuracy (Acc), Sensitivity (Se), Specificity (Sp), F1-score, Positive Predictive Value (PPV), and the Area Under the ROC Curve (AUC).

The “Diabetes” column of the dataset was used as the label, considering one for Diabetics and zero for non-Diabetics. The classification was performed and evaluated using a 5-fold cross-validation strategy, which involves splitting the dataset into five equal parts or folds. In each iteration, four of the folds are used for training and one fold is used for testing. This process is repeated five times with each fold serving as the testing set once. The performance of the classifiers (LR and XGBoost) is then averaged across the five iterations.

Features sets obtained from PPG cycles were split in each fold patient-wise. This way, features extracted from signals of the same patient are placed in training folds or testing fold, never in both. By allocating each patient in training and testing sets by their IDs, we avoided the classification bias caused by mixing information of the same patient in both training and testing folds in each iteration [Costa et al. 2023].

The overall performance of the classifiers is obtained by evaluating the mean and standard deviation of all statistical metrics obtained in each iteration. Moreover, we used the *permutation_importance* algorithm from Scikit-learn to determine feature importance for both classifiers. Since we have a 5-fold strategy, the mean feature importance for all 5-folds was computed.

3. Results

This study used data from 86 subjects, with 453 cycles extracted using the FSW method. Out of these cycles, 172 belong to Diabetic patients. A total of 110 features were used as input for the classifiers, with 104 of these being morphological features extracted from each PPG cycle and the remaining 6 features comprising metadata of the patients.

The goal of the study was to classify non-Diabetic and Diabetic patients by using LR and XGBoost classifiers on features extracted from PPG cycles. The 5-fold cross-validation strategy permitted the evaluation of the classifiers’ performance on multiple iterations and reduced the impact of any random fluctuations in the data.

Table 2 summarizes the overall performance of the proposed methods for the binary classification between Diabetic/non-Diabetic. The metrics are presented as the mean and standard deviation of the five iterations of the cross-validation strategy. In addition, Figs. 2 and 3 show the AUC for the LR and XGBoost methods, respectively.

Table 2 shows the classification metrics of this study, which achieved a mean AUC of 0.79 ± 0.15 for LR and 0.74 ± 0.17 for XGBoost, respectively. The best achieved AUC with LR model is comparable to the AUC of 0.79 ± 0.15 , reported in [Hettiarachchi and Chitraranjan 2019], which used the same dataset for diabetes detection.

Additionally, we present the feature importance measures, displaying the 7 most significant features for the two proposed models’ ability to predict the outcomes. The importance and name of each feature is shown in Fig. 4 for Logistic Regression and in Fig. 5 for XGBoost. Based on the feature importance results obtained from both algorithms, we have compiled a summary of the six most significant PPG features in Table 3. We

Table 2. Comparison of the performance results of our two proposed algorithms for Diabetic/Non-Diabetic classification.

Metrics	Logistic Regression	XGBoost
<i>Se</i>	0.66 ± 0.27	0.56 ± 0.22
<i>Spe</i>	0.75 ± 0.19	0.70 ± 0.16
<i>F1-score</i>	0.59 ± 0.20	0.52 ± 0.16
<i>Acc</i>	0.70 ± 0.12	0.64 ± 0.10
<i>PPV</i>	0.61 ± 0.24	0.54 ± 0.21
<i>AUC</i>	0.79 ± 0.15	0.75 ± 0.17

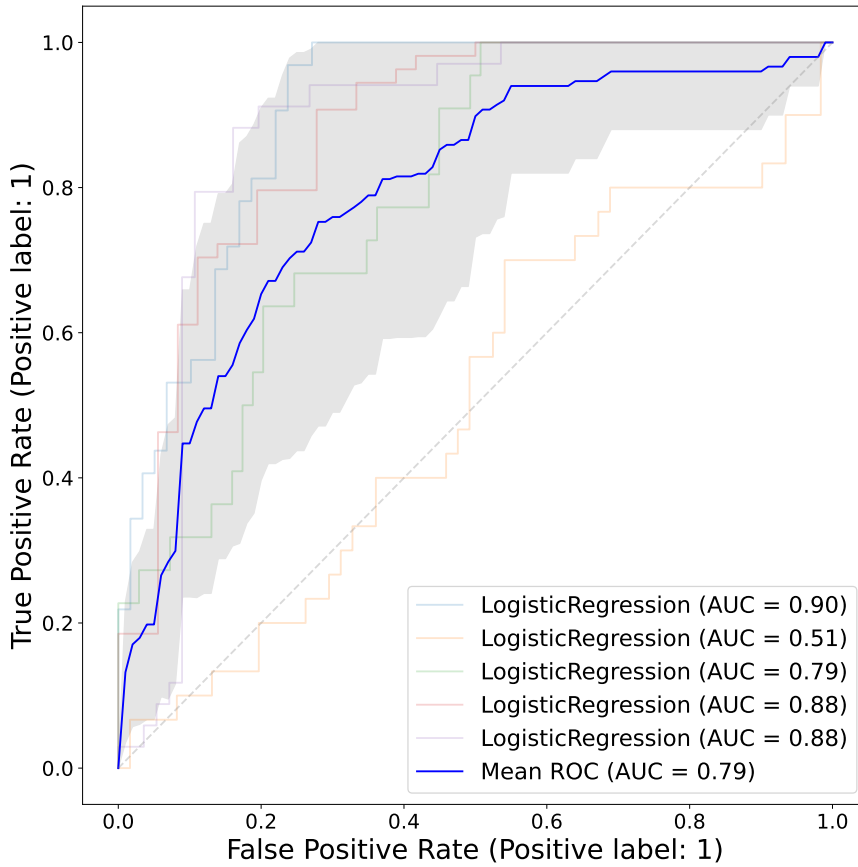


Figure 2. Receiver Operating Curve (ROC) and AUC for each fold of the Logistic Regression method.

have excluded from this table the features associated with metadata, namely age, weight, height, and BMI, as they are self-explanatory.

4. Discussion

In this study, we proposed a method aimed at aiding in the diagnosis of diabetes based on PPG signals. This is an exploratory study that correlates the morphological features extracted from PPG signals with metadata information from the subjects.

We show that it is possible to achieve results comparable to the state-of-the-art

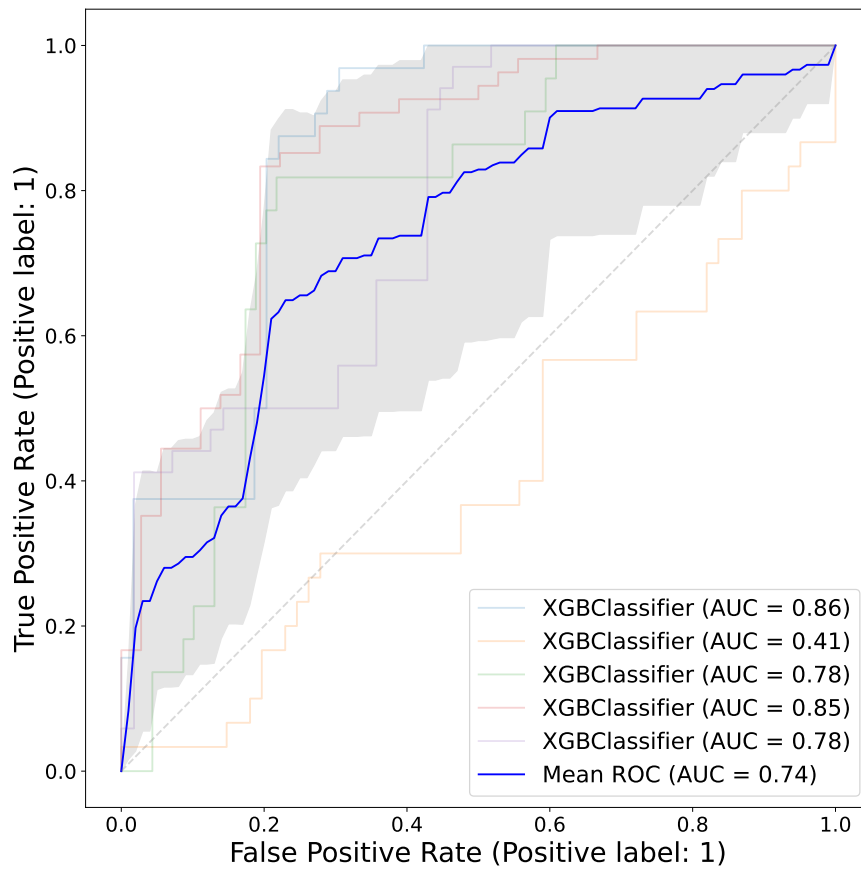


Figure 3. Receiver Operating Curve (ROC) and AUC for each fold of the XGBoost method.

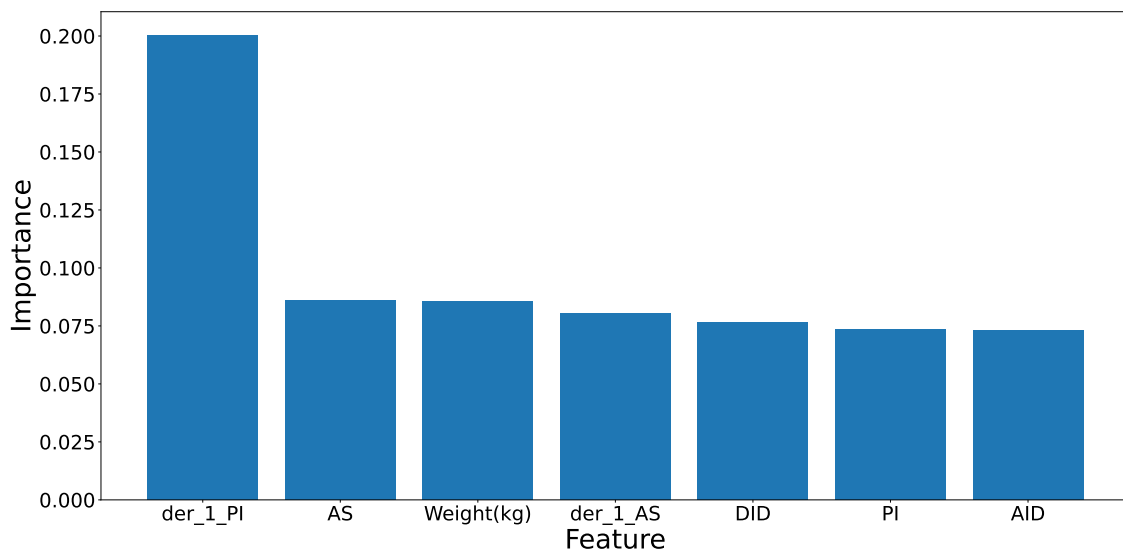


Figure 4. Importance of the 7 most significant features for Logistic Regression. The values represent the the mean of the five folds.

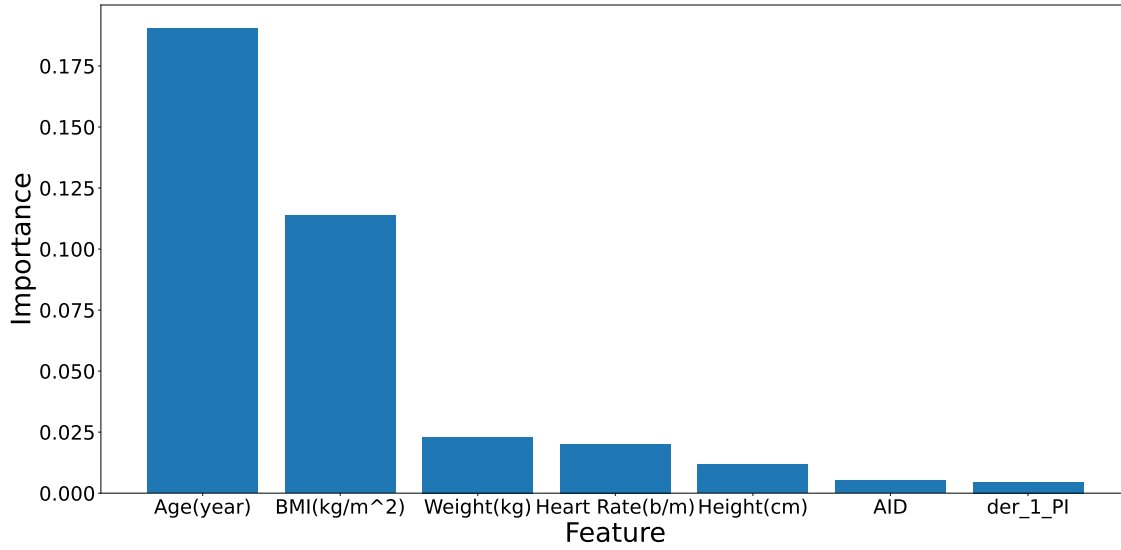


Figure 5. Importance of the 7 most significant features for XGBoost. The values represent the the mean of the five folds.

Table 3. Description of the six most significant PPG signal features of the experiment.

Name	Description
<i>der_1_PI</i>	Intensity of the 1st inflection point in the 1st derivative of the PPG cycle
<i>AS</i>	Slope of the systolic portion in the PPG cycle (from the beginning of the cycle to the systolic peak)
<i>der_1_AS</i>	Slope of the line up to the 1st inflection point (from the beginning of the 1st derivative to the 1st inflection point)
<i>DID</i>	Difference in intensity between the systolic peak and the end of the PPG cycle
<i>AID</i>	Difference in intensity between the beginning of the PPG cycle and the systolic peak
<i>PI</i>	Intensity of the systolic peak in the PPG cycle normalized by the average value (baseline)

using a limited dataset, with features extracted from single heartbeats (cycles) of the PPG signals. Table 4 displays a comparison of works that propose to detect diabetes through PPG signals. [Hettiarachchi and Chitraranjan 2019] used the same dataset as our study; however, they used features extracted from the whole 2.1 s segment available, reducing the input size of their models. [Moreno et al. 2017], on the other hand, used a private dataset with one-minute long segments registered from 1,170 subjects and achieved results in the same range as ours (AUC 0.69, Se 0.65 Spe 0.64) using different machine learning techniques. [Zanelli et al. 2023] used 1 s segments of PPG raw signals from 100 subjects to predict diabetes. Even though they proposed a transfer learning strategy, their best-performing model was the one without transfer learning, combining age and sex metadata

information.

Different from the works on Table 4, [Nirala et al. 2019] achieved Acc 0.98, Se 0.99 and Spe 0.97, using data from 141 subjects. However, they randomly partitioned the data into k-equal folds on a k-fold strategy, implicating that they mixed the same subject data on the training and test partitions. [Gupta et al. 2022] also did the same approach, not clearly specifying the dataset division. Previous research has shown that utilizing the same subject for both the training and test sets can yield better outcomes, as data leakage may occur due to the high interdependence among intra-subject heartbeats [Costa et al. 2023]. [Reddy. et al. 2017], on the other hand, achieved Acc 0.89, Se 0.90 and Spe 0.88, using data from 100 subjects. Nevertheless, they didn't make it clear if during partition of the data on the k-fold strategy they ensure that the same subject signals wasn't on both training and test sets.

Table 4. Comparison of the performance results of our two proposed algorithms for Diabetes detection through PPG signal and the related state-of-the-art works.

Reference	F1-score	AUC	Accuracy	Sensitivity	Specificity	Dataset	Method
[Moreno et al. 2017]	–	0.70	–	0.80	0.48	Private (s = 1,170)	RF and GB - PPG features and metadata
[Reddy. et al. 2017]	–	–	0.89	0.90	0.88	Private (s = 100)	SVM - PPG features
[Hettiarachchi and Chitraranjan 2019]	0.71 ± 0.15	0.79 ± 0.15	0.71	–	–	[Liang et al. 2018] (s = 64)	LDA - PPG features and metadata
[Hettiarachchi and Chitraranjan 2019]	0.69 ± 0.10	0.74 ± 0.17	0.79	–	–	[Liang et al. 2018] (s = 64)	SVM - PPG features and metadata
[Nirala et al. 2019]	0.98	0.98	0.98	0.99	0.97	Private (s = 141)	SVM - 37 PPG features
[Avram et al. 2020]	–	0.77	–	0.75	0.65	Private (s = 55,433)	CNN - PPG raw signal and metadata
[Srinivasan and Foroozan 2021]	–	0.83	0.76	0.77	0.76	MIMIC III (s = 808)	2D CNN - PPG scalogram and metadata
[Gupta et al. 2022]	0.99	–	0.98	0.99	0.96	[Liang et al. 2018] (s = 219)	RF - PPG features
[Zanelli et al. 2023]	0.46	0.56	–	0.75	0.76	Private (s = 100)	CNN - PPG raw signal, age and sex
[Zanelli et al. 2023]	0.31	0.61	–	0.50	0.72	Private (s = 100)	CNN and Transfer Learning - PPG raw signal, age and sex
Our method: XGBoost	0.52 ± 0.16	0.74 ± 0.17	0.64 ± 0.10	0.56 ± 0.22	0.70 ± 0.16	[Liang et al. 2018] (s = 86)	XGBoost - 104 PPG features and 6 metadata
Our method: LR	0.59 ± 0.20	0.79 ± 0.15	0.70 ± 0.12	0.66 ± 0.27	0.75 ± 0.19	[Liang et al. 2018] (s = 86)	LR - 104 PPG features and 6 metadata

[Nirala et al. 2019] suggest in their work that the absence of a dicrotic notch on PPG waves can be observed in Diabetic subjects. [Zanelli et al. 2023] also compares a waveform from a Diabetic and a non-Diabetic subjects, highlighting the absence of a dicrotic notch in the former. However, they imply that these differences are not always readily apparent. On the contrary, [Srinivasan and Foroozan 2021] indicates that despite using a much larger dataset than the aforementioned studies, it is not possible to observe this dicrotic notch difference between the two groups. Our work aligns with Srinivasan's findings, where no discernible difference in regarding the presence of dicrotic notch was found. Moreover, the absence of dicrotic notch may be related to aging or other factors [MILLASSEAU et al. 2002].

Heart Rate Variability (HRV) is broadly used as a feature in diabetes detection and BGL estimation studies [Reddy. et al. 2017]. However, this feature may not be calculated with confidence in short segments of data such as the available in our dataset. For this reason, the HRV was not estimated for this experiment.

The small number of patients in the dataset raises concerns about potential bias in the classifier's performance. There is a possibility that the classifier may learn to estimate other clinical parameters, such as blood pressure or other clinical conditions, instead of

focusing solely on Diabetes classification. In addition, several studies involving Diabetes classification do not explicitly consider the bias caused by mixing PPG signals from the same patient in the training and testing steps [Panwar et al. 2020], which may result in better metrics due to overfitting. We addressed this issue by adopting the 5-fold cross-validation strategy considering the patient’s identification number while splitting the data.

Comparing the significance of each feature shown in Figs. 4 and 5, it is seen that the Logistic Regression algorithm considered mostly the PPG-related features as important for predicting diabetes. On the other hand, the XGBoost algorithm made predictions considering almost exclusively the patient’s metadata information. The better results provided by LR algorithm suggest that the PPG cycles may indeed carry diabetes-related information. Both classification algorithms consider as important two specific PPG features: *der_1_PI* and *AID*. These features use information from the systolic wave between its onset and peak, indicating a potential value of this portion of the PPG wave for further analysis.

One of the reasons for the lack of studies using this dataset may be related to its reduced number of patients and its short time signals, considering that each patient has only six seconds of PPG signal divided into three non-continuous segments. This aspect may impair the performance of segmentation algorithms, such as the FSW. Such impairment seen in this work led to the exclusion of eleven patients from the experiment due to segmentation errors and noisy signals. The FSW algorithm was first introduced in [Zhang et al. 2020] for removing the baseline and segmenting PPG signals. However, when applied to signals with a small number of cycles, it tends to discard the edges, resulting in a reduction of the data passed to the feature extraction algorithm and, consequently, a lower number of computed features, which negatively affects the performance and generalization of the classifiers.

The clinical use of PPG-based diabetes detection algorithms demands higher accuracy and reliability on different patients with varying medical conditions. Therefore, in order to enhance the effectiveness and reliability of the findings, ensuring their generalizability and applicability, future studies on the detection of diabetes using wearable devices should be conducted on a larger and more diverse dataset, including patients with different medical conditions, ages, and ethnicities.

5. Conclusion

Our study proposed a method for diagnosing diabetes based on PPG signals, utilizing morphological features extracted from single beats of the PPG signals. We achieved comparable results to state-of-the-art studies using a limited dataset. Comparison with other studies showed that our approach yielded promising accuracy, sensitivity, and specificity values. However, the small number of patients in our dataset and the short duration of the PPG signals available for the analysis may have impacted our results. Additionally, publicly available well-annotated datasets in this field are needed for further advance research in this area.

Acknowledgments

This work was supported by Foxconn Brazil and Zerbini Foundation, as part of the research project “Remote Patient Monitoring System”.

References

- Avram, R., Olgin, J. E., Kuhar, P., Hughes, J. W., Marcus, G. M., Pletcher, M. J., Aschbacher, K., and Tison, G. H. (2020). A digital biomarker of diabetes from smartphone-based vascular signals. *Nature medicine*, 26(10):1576—1582.
- Chen, T. and Guestrin, C. (2016a). XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.
- Chen, T. and Guestrin, C. (2016b). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794. Association for Computing Machinery.
- Chowdhury, M. H., Shuzan, M. N. I., Chowdhury, M. E., Mahbub, Z. B., Uddin, M. M., Khandakar, A., and Reaz, M. B. I. (2020). Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques. *Sensors*, 20(11).
- Costa, T. B. D. S., Dias, F. M., Cardenas, D. A. C., Toledo, M. A. F. D., Lima, D. M. D., Krieger, J. E., and Gutierrez, M. A. (2023). Blood pressure estimation from photoplethysmography by considering intra- and inter-subject variabilities: Guidelines for a fair assessment. *IEEE Access*, 11:57934–57950.
- DeFronzo, R. A., Bonadonna, R. C., and Ferrannini, E. (1992). Pathogenesis of NIDDM: A Balanced Overview. *Diabetes Care*, 15(3):318–368.
- El-Hajj, C. and Kyriacou, P. (2021). Cuffless blood pressure estimation from ppg signals and its derivatives using deep learning models. *Biomedical Signal Processing and Control*, 70:102984.
- Gupta, S., Singh, A., Sharma, A., and Tripathy, R. K. (2022). dsvri: A ppg-based novel feature for early diagnosis of type-ii diabetes mellitus. *IEEE Sensors Letters*, 6(9):1–4.
- Hettiarachchi, C. and Chitraranjan, C. (2019). A machine learning approach to predict diabetes using short recorded photoplethysmography and physiological characteristics. In *Artificial Intelligence in Medicine*, volume 11526, pages 322–327. Springer International Publishing.
- Kirk, J. K. and Stegner, J. (2010). Self-monitoring of blood glucose: Practical aspects. *Journal of Diabetes Science and Technology*, 4(2):435–439.
- LaMonte, M. J., Blair, S. N., and Church, T. S. (2005). Physical activity and diabetes prevention. *Journal of Applied Physiology*, 99(3):1205–1213.
- Liang, Y., Chen, Z., Liu, G., and Elgendi, M. (2018). A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in china. *Scientific Data*, 8:180020.
- Lin, W.-H., Li, X., Li, Y., Li, G., and Chen, F. (2020). Investigating the physiological mechanisms of the photoplethysmogram features for blood pressure estimation. *Physiological Measurement*, 41(4):044003.
- Mejía-Mejía, E., Allen, J., Budidha, K., El-Hajj, C., Kyriacou, P. A., and Charlton, P. H. (2022). 4 - photoplethysmography signal processing and synthesis. In Allen, J. and Kyriacou, P., editors, *Photoplethysmography*, pages 69–146. Academic Press.

- MILLASSEAU, S., KELLY, R., RITTER, J., and CHOWIENCZYK, P. (2002). Determination of age-related increases in large artery stiffness by digital pulse contour analysis. *Clinical Science*, 103(4):371–377.
- Moreno, E. M., Lujan, M. J. A., Rusinol, M. T., Fernandez, P. J., Manrique, P. N., Trivino, C. A., Miquel, M. P., Rodriguez, M. A., and Burguillos, M. J. G. (2017). Type 2 diabetes screening test by means of a pulse oximeter. *IEEE Transactions on Bio-Medical Engineering*, 64(2):341–351.
- Mukkamala, R., Hahn, J.-O., and Chandrasekhar, A. (2022). 11 - photoplethysmography in noninvasive blood pressure monitoring. In Allen, J. and Kyriacou, P., editors, *Photoplethysmography*, pages 359–400. Academic Press.
- Nirala, N., Periyasamy, R., Singh, B. K., and Kumar, A. (2019). Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine. *Biocybernetics and Biomedical Engineering*, 39(1):38–51.
- Nitzan, M. and Ovadia-Blechman, Z. (2022). 9 - physical and physiological interpretations of the ppg signal. In Allen, J. and Kyriacou, P., editors, *Photoplethysmography*, pages 319–340. Academic Press.
- Panwar, M., Gautam, A., Dutt, R., and Acharyya, A. (2020). CardioNet: Deep learning framework for prediction of CVD risk factors. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5. IEEE.
- Pilt, K., Ferenets, R., Meigas, K., Lindberg, L.-G., Temitski, K., and Viigimaa, M. (2013). New photoplethysmographic signal analysis algorithm for arterial stiffness estimation. *The Scientific World Journal*, 2013:1–9.
- Reddy., V. R., Dutta Choudhury., A., Jayaraman., S., Kumar Thokala., N., Deshpande., P., and Kaliaperumal., V. (2017). Perdmcs: Weighted fusion of ppg signal features for robust and efficient diabetes mellitus classification. In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies - SmartMedDev*, pages 553–560. SciTePress.
- Srinivasan, V. B. and Foroozan, F. (2021). Deep learning based non-invasive diabetes predictor using photoplethysmography signals. In *2021 29th European Signal Processing Conference (EUSIPCO)*, pages 1256–1260.
- Usman, S., Reaz, M., and Ali, M. (2011). Repeated measurement analysis of the area under the curve of photoplethysmogram among diabetic patients. *Life Sci. J*, 10:532–539.
- Zanelli, S., Ammi, M., Hallab, M., and El Yacoubi, M. A. (2022). Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review. *Sensors*, 22(13):4890.
- Zanelli, S., Yacoubi, M. A. E., Hallab, M., and Ammi, M. (2023). Type 2 diabetes detection with light cnn from single raw ppg wave. *IEEE Access*, 11:57652–57665.
- Zhang, G., Mei, Z., Zhang, Y., Ma, X., Lo, B., Chen, D., and Zhang, Y. (2020). A noninvasive blood glucose monitoring system based on smartphone ppg signal processing and machine learning. *IEEE Transactions on Industrial Informatics*, 16(11):7209–7218.