A Machine Learning Approach to Predict Diabetes Using Short Recorded Photoplethysmography & Physiological Characteristics.

Chirath Hettiarachchi[1], Charith Chitraranjan[1]

[1] Department of Computer Science & Engineering, University of Moratuwa, Sri Lanka
chirathyh@gmail.com, charithc@cse.mrt.ac.lk

Abstract. Diabetes is a global epidemic, which leads to severe complications such as heart disease, limb amputations and blindness, mainly occurring due to the inability of early detection. Photoplethysmography (PPG) signals have been used as a non-invasive approach to predict diabetes. However, current methods use long, continuous signals collected in a clinical setting. This study focuses on predicting Type 2 Diabetes from short (~2.1s) PPG signals extracted from smart devices, and readily available physiological data such as age, gender, weight and height. As this kind of PPG signals can be easily extracted using mobile phones or smart wearable technology, the user can get an initial prediction without entering a medical facility.

Through the analysis of morphological features related to the PPG waveform and its derivatives, we identify features related to Type 2 Diabetes and establish the feasibility of predicting Type 2 Diabetes from short PPG signals. We cross validated several classification models based on the selected set of features to predict Type 2 Diabetes, where Linear Discriminant Analysis (LDA) achieved the highest area under the ROC curve of 79%. The successful practical implementation of the proposed system would enable people to screen themselves conveniently using their smart devices to identify the potential risk of Type 2 Diabetes and thus avoid austere complications of late detection.

Keywords: Machine Learning, Diabetes Type II, Photoplethysmography, Feature Selection.

1 Introduction

The symptoms of a disease are usually detected at a later stage once it is visible to human perception. This could lead towards great risks to humans and possible severe complications related to the disease. Hence there is immense importance and value of early disease prediction which ensures proper management and early diagnosis reducing the potential risk to patients.

In the recent past there has been an increase of interest towards the application of Machine Learning techniques to identify novel relationships among variables, which are not easily detectable. This has become a key motivator towards the development of early disease prediction systems focusing on early trivial symptoms of disease. Such early disease screening systems are practically useful towards the end user if these systems could operate using data streams readily available in routine life. After early screening, users would then be able to undertake established medical tests and procedures as required.

Our research focuses on developing a system capable of predicting Type 2 Diabetes using readily available data of general users. We use physiological characteristics and short(~2.1s) recoded Photoplethysmography (PPG) signal measurements of

the users towards Diabetes prediction. Photoplethysmography signals have been used in previous work as a non-invasive approach to predict diabetes. However, these methods use long continuous signals collected in a clinical setting.

Diabetes is a severe global phenomenon which has been the root cause of millions of deaths worldwide. The study focuses on detecting Type 2 Diabetes which is caused by the body's inefficient use of insulin, resulting in abnormally high levels of sugar in the blood. The symptoms of the disease are less marked and is often detected several years after the onset through complications which can result in premature heart disease, blindness, limb amputations and kidney failure. Hence the utmost importance of early detection and continuous surveillance. [1]

Photoplethysmography can be identified as a non-invasive, inexpensive, optic technique which measures the blood volume changes in blood vessels through which oxygen saturation, blood pressure, cardiac output could be measured [2] [3]. In recent researches it has been identified that PPG is a promising technique towards early screening of diseases as the PPG waveform possess significant information embedded within [4]. The recent advancement in hardware has enabled the integration of high-quality PPG sensors within mobile phones & smart watches, which has brought a non-invasive health screening device to the fingertip of everyday users [5] [6]. Such hardware enhancement has prompted towards the development of early disease prediction systems through the utilization of rich data available. The rapid development of such smart wearable devices has thus enabled the easy access towards regular measurement of required health parameters which is recommended for diseases such as Diabetes.

2 Related Work

The objective of this study is to predict Diabetes using a single instance of a short-recorded PPG signal combined with physiological characteristics, which is uncommon. However, there have been studies that have focused on analyzing the PPG signals and other features of the users towards the prediction of Type 2 Diabetes. Studies have been conducted to predict diabetes through the use of Electronic Medical Records (EMR) which mainly focus on prescriptions and diagnoses as features [7]. Also there have been researches which have been focused on developing systems to assess the Blood Glucose levels using the PPG signal [8]. Similar research based on PPG to directly predict Diabetes, has focused upon the analysis of continuous monitoring of the signal to extract features related to Heart Rate Variability (HRV), which is used towards the predictions. Reddy, V. Ramu, et al [9] have focused on analyzing features related to HRV and shape information of the PPG waveform towards disease prediction. Ballinger, Brandon, et al [10] focused on predicting a range of diseases including Diabetes using features such as the user medical history, step count and continuous optical heart rate measured using PPG.

In contrast the core focus of this research relies on developing a system capable of using basic user known physiological characteristics and a single instance of a PPG signal recorded for at least 2 - 3 seconds to be used towards Diabetes prediction. The successful research in this domain would ensure great value addition in the fields of smart wearables and healthcare.

3 Methodology

The research was conducted based on a de-identified open clinical trial dataset for non-invasive detection of cardiovascular diseases by Liang, Yongbo, et al [11], which contains physiological characteristics, short recorded PPG signals and information related to the presence of Diabetes and Hypertension of patients. The Normal, Diabetes and Hypertension related data was extracted from the database, where signal processing and feature extraction algorithms were developed to obtain a target set of features. The extracted target feature set was analyzed in order to identify features related to Diabetes and was tested against multiple Machine Learning algorithms, and hyper parameters tuned to obtain the best results. The evaluation of the different models was conducted based on the area under the ROC curve which considers both the true positive rate and the false positive rate of the binary classification problem.

3.1 Data Description

The final extracted data for evaluation purposes eliminating the erroneous signals, contained a total of 150 subjects comprising fully normal (51 subjects), Prehypertension (39 subjects), Hypertension (28 subjects) and Diabetes (32 subjects). Subjects with cerebral infarction and cerebrovascular disease were excluded from the study. It should be noted that the number of subjects in the target group Diabetes was comparatively less and there were diabetes subjects who were also suffering from both Hypertension and Prehypertension. Hypertension and Diabetes are strongly interconnected diseases, both affecting the cardiovascular system of the human body. Diabetes is also considered as a risk factor for Hypertension, where both diseases stem from the Metabolic Syndrome. The breakdown of the subjects with respective to their category is presented in Table 1.

Table 1. Dataset Subject Disease Distribution

Subject Description	Number of Subjects
Normal	51
Prehypertension Only	39
Hypertension Only	28
Diabetes Only	9
Diabetes & Prehypertension	16
Diabetes & Hypertension	7

The morphological features of the PPG signals analyzed in this research are related to the cardiovascular system. Due to the strong inter-connection between the two diseases, there is an inherent overlap between the effect of the possible features towards disease prediction. Thus, the research focusses on diabetes prediction at the presence of prehypertension & hypertension, which can be identified as the other major disease affecting the cardiovascular system.

Hence, considering the practical disease distribution in the environment, it is important to clearly identify unique features to develop a system which is able to clearly

predict diabetes when fully normal, and subjects with a form of hypertension are present. The system can be further enhanced through the analysis of other diseases affecting the cardiovascular system which is not within the scope of this study. The age, gender, height, weight and BMI information were obtained as a set of features representing the physiological characteristics of the patients. In addition, the Body Fat percentage was calculated using the formula given in [12] and used towards the analysis.

3.2 Data Preprocessing & Feature Extraction

The PPG signals related to the above identified subjects were analyzed in order to extract features related to the cardiovascular system. The signals in the aforementioned dataset was captured with a sampling rate of 1KHz using a 12-bit ADC, and a hardware filter design of 0.5 - 12 Hz bandpass. For each patient three segments of short recorded PPG waveforms were captured and the best signal was selected based on the Skewness Signal Quality Index (SSQI) [13] for the study. The selected signal was applied a 4th Order Chebyshev II filter in order to eliminate the noise [14]. The final processed signal and its second derivative the Accelerated Photoplethysmography (APG) signal, was used towards extracting identified features related to the cardiovascular system using the MATLAB software. The list of extracted features is presented in Table 2 with a brief description to understand the effect of the features. The extracted data points from the signals are presented in Fig 1.

It should be noted that the development of robust algorithms to extract the PPG features is a tedious task due to the motion artifacts and noise components present within the signals. This can be identified as a practical challenge when incorporating such prediction systems in practice.

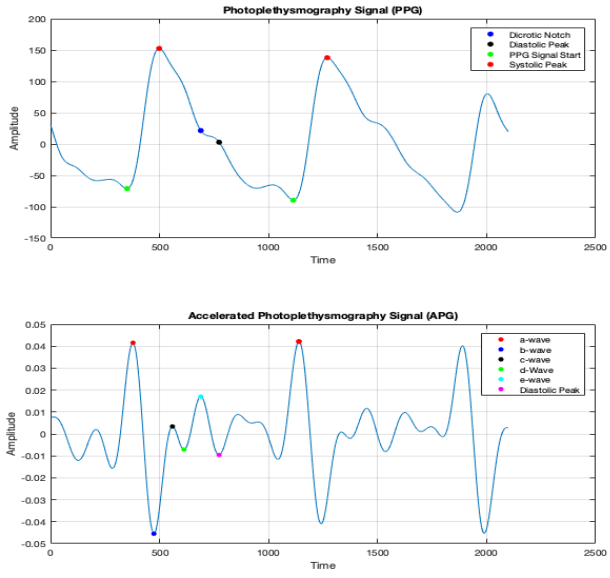

Fig. 1. Extracted data points from Photoplethysmography (PPG) and Accelerated Photoplethysmography (APG) Signals.

Table 2. Features Selected using ANOVA Tests

Feature	Description	ANOVA Test (Diabetes)		ANOVA Test (Hypertension)	
Physiological Features		F Value	P Value	F Value	P Value
Gender		1.313	0.256	1.929	0.15
Age		11.170	**0.001**	20.158	3.13E-08
Height	Centimeters(cm)	0.267	0.607	0.078	0.925
Weight	Kilograms(kg)	1.125	0.293	3.040	0.0517
BMI	Weight(kg) / Height^2 (m^2)	0.569	0.454	3.989	0.0211
Body Fat (%)	(1.2*BMI) + (0.23*Age) - (10.8*Gender) - 5.4	1.268	0.265	3.943	0.022
APG Signal Features					
b/a Ratio	Increase Arterial Stiffness	0.845	0.362	4.414	0.0142
c/a Ratio	Decreased Arterial Stiffness	1.957	0.167	1.824	0.166
d/a Ratio	Decreased Arterial Stiffness	0.164	0.687	5.859	3.78E-03
e/a Ratio	Decreased Arterial Stiffness	5.642	**0.021**	0.571	**0.566**
(b-c-d-e)/a Ratio	Vascular Aging	0.053	0.818	6.429	2.26E-03
(b-e)/a Ratio	Vascular Aging	0.201	0.656	2.969	0.0553
(b-c-d)/a Ratio	Sensation of Coldness Treatment	0.827	0.367	6.602	1.93E-03
(c + d - b)/a Ratio	Vascular Aging	0.827	0.367	6.602	1.93E-03
a-a Interval	Complete Heart Cycle	0.205	0.652	0.498	0.609
(-d/a) Ratio	Index Left Ventricular Overload	0.164	0.687	5.859	3.78E-03
PPG Signal Features					
Systolic Amplitude	Stroke Volume / Local Vascular Distensibility	0.166	0.685	4.622	1.17E-02
Pulse Area	Total Area Under the PPG Curve	0.935	0.338	0.702	0.498
Inflexion Point Area (IPA) Ratio	Total Peripheral Resistance	0.309	0.580	0.186	0.831
Pulse Interval (PI)	Complete Heart Cycle	0.137	0.712	0.449	0.639
PI / Systolic Amplitude Ratio	Cardiovascular System Properties	6.291	**0.015**	2.305	**0.104**
Augmentation Index (AI)	Vascular Tone / Endothelial Dysfunction	5.793	**0.019**	0.225	**0.799**
Adjusted AI	Vascular Tone / Endothelial Dysfunction	5.793	**0.019**	0.225	**0.799**
Large Artery Stiffness Index	Arterial Stiffness	0.368	0.547	4.567	1.23E-02
Rise Time	Cardiovascular Disease Classification	0.016	0.899	5.102	7.53E-03

3.3 Feature Selection

The above extracted features have been clearly identified and described in Elgendi et all [2] and Allen, John et all [3]. The identified features were analyzed in previous researchers mainly towards the understanding and interpretation of the cardiovascular system. The cardiovascular system is affected by various diseases such as Hypertension, Diabetes and Renal failure. Hence it is important to uniquely identify the features

related to diabetes which affects the vascular system. An ANOVA test with a 95% confidence interval was carried out focusing on the normal subjects (51 subjects) and the subjects with only diabetes (9 subjects). Through the test it was identified that the Age, Augmentation Index (AI), Adjusted AI, e/a Ratio and the ratio between the Pulse Interval to the Systolic Amplitude are suitable features towards the prediction of type 2 diabetes.

Previous research studies [15] [16] [17], have also identified that the Augmentation Index (also known as the Reflection Index) can be used towards detecting Endothelial Dysfunction in diabetes patients. Endothelial Dysfunction [18] occurs due to damages in the vascular endothelium which affects the operations of the vascular system. This can also lead to Atherosclerosis, and is considered a major risk factor for cardiovascular diseases. Thus, AI and Adjusted AI can be utilized towards diabetes prediction, as Hyperglycemia [19] is a major risk factor towards endothelial dysfunction. It was also identified that the ratio of Pulse Interval to its Systolic Amplitude can be used towards diabetes prediction. This ratio was previously identified by Poon et al [20] to understand the properties of the cardiovascular system. The e/a ratio can be used towards the prediction of Diabetes, where Takazawa et all [21] demonstrated that the increase of the e/a ratio results in decrease of arterial stiffness, and that e/a decreases with age. The Age was also identified as a key feature towards diabetes prediction, which is due to the higher probability of diabetes among older subjects.

It is important to note that other features analyzed related to vascular ageing and arterial stiffness were not identified as predictors towards diabetes, even though they represent the vascular system. A second ANOVA test with a 95% confidence interval was conducted focusing on the normal (51 subjects), prehypertension (39 subjects) and hypertension (28 subjects) subjects. The results of the earlier test were justified as the e/a ratio, AI, Adjusted AI and the ratio of pulse interval to its systolic amplitude were not identified as prominent features towards the prediction of hypertension. Hence it can be clearly identified that the e/a ratio, AI, Adjusted AI and the ratio of pulse interval to its systolic amplitude are unique features towards the prediction of type 2 diabetes. However Age was also identified as a predictor towards hypertension.

3.4 Machine Learning Models

The main objective of the study focused on predicting Diabetes based on the extracted signal and physiological features. Hence supervised Machine Learning techniques [22] were used to carry out binary classification where the evaluation metric was set to the area under the ROC curve. The study evaluated the Naive Bayes classifier, Linear Discriminant Analysis (LDA) which are statistical algorithms, logic-based algorithms such as Decision Trees and its variants focusing on ensemble and boosting methods such as Random Forest and AdaBoost Classifiers. Logistic Regression and Support Vector Machine (SVM) models were also evaluated towards the classification of Diabetes.

The selected models were initially tuned to identify the optimum hyper parameters for each model using random search and stratified 10-fold cross validation, optimizing the target evaluation metric. All the classification models were tuned ensuring the selection of suitable hyper parameters to avoid overfitting due to the relatively small number of data samples. Upon successful tuning of the algorithms a final 10-fold cross validation was run on the entire dataset in order to select the best performing model. The classification models were implemented and evaluated using the scikit-sklearn python library [23].

A binary classification ensuring equal samples for each class was carried out towards the prediction of diabetes. The evaluation focused upon three cases. Normal versus only diabetes, normal versus diabetes subjects with prehypertension, normal versus diabetes subjects with prehypertension and hypertension. A final test was carried out using only the selected PPG signal features to evaluate the robustness of using PPG for diabetes prediction. The subjects below 30 years were excluded from the test ensuring an even distribution of normal and diabetes subjects across different age groups. The age distribution of the subjects is presented in Fig 2.

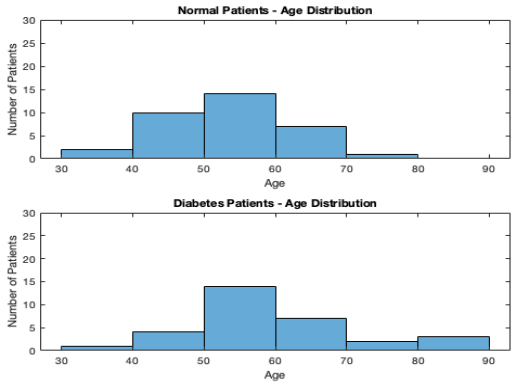

Fig. 2. Age Distribution of Normal and Diabetes Patients.

4 Evaluation & results

The classification results for the above identified four cases are presented in Table 3. The Naïve Bayes classifier performed worst in all the cases as indicated by low F1, precision and recall scores. Decision Trees and LDA achieved accuracies of 83% and 82% respectively in classifying diabetes only subjects from normal subjects. However, in order to enhance the confidence of this finding the test should be carried out with additional samples from diabetes only subjects. In contrast diabetes prediction for subjects with prehypertension and hypertension provided more confident results due to the relatively larger number of subjects.

SVM performed well predicting diabetes with the presence of prehypertension providing an accuracy of 71%, whereas LDA was able to obtain an accuracy of 79% for predicting diabetes with the presence of hypertension and prehypertension. It should be noted that this test case portrays the real-world scenario of a robust diabetes prediction system encompassing all possible diabetes and hypertension combinations.

The evaluation of the robustness of using PPG for the diabetes prediction is of utmost importance to ensure that the predictions are not biased from physiological features such as the age which is used in this study. The conducted control test achieved an accuracy of 70% for the Decision Trees classifier which establishes the suitability of using the PPG signal for diabetes prediction.

Table 3. Diabetes Prediction Results

Experiment	Classification Algorithm	ROC	F1	Precision	Recall
Normal (n = 9) vs Diabetes Only (n = 9) (3-fold stratified cross validation)	Naïve Bayes	0.54 ± 0.18	0.24 ± 0.17	0.28 ± 0.21	0.22 ± 0.16
	Logistic Regression	0.33 ± 0.16	0.5 ± 0.14	0.41 ± 0.07	0.67 ± 0.27
	AdaBoost Classifier	0.78 ± 0.16	0.67 ± 0.24	1.0 ± 0.0	0.56 ± 0.31
	Random Forest	0.78 ± 0.18	0.67 ± 0.24	1.0 ± 0.0	0.56 ± 0.31
	Decision Trees	**0.83 ± 0.14**	**0.79 ± 0.21**	**0.92 ± 0.12**	**0.78 ± 0.31**
	SVM (Linear Kernel)	0.63 ± 0.21	0.61 ± 0.2	0.51 ± 0.13	0.78 ± 0.31
	LDA	**0.82 ± 0.14**	**0.76 ± 0.13**	**0.67 ± 0.12**	**0.89 ± 0.16**
Normal (n = 25) vs Diabetes with Prehypertension (n = 25) (10-fold stratified cross validation)	Naïve Bayes	0.54 ± 0.23	0.03 ± 0.09	0.03 ± 0.08	0.03 ± 0.1
	Logistic Regression	0.66 ± 0.26	0.69 ± 0.04	0.53 ± 0.06	1.0 ± 0.0
	AdaBoost Classifier	0.61 ± 0.25	0.52 ± 0.19	0.56 ± 0.27	0.57 ± 0.28
	Random Forest	0.60 ± 0.27	0.52 ± 0.3	0.53 ± 0.32	0.57 ± 0.36
	Decision Trees	0.53 ± 0.28	0.38 ± 0.32	0.44 ± 0.39	0.37 ± 0.34
	SVM (Linear Kernel)	**0.71 ± 0.2**	**0.55 ± 0.3**	**0.49 ± 0.29**	**0.65 ± 0.39**
	LDA	**0.69 ± 0.21**	**0.68 ± 0.16**	**0.73 ± 0.23**	**0.72 ± 0.25**
Normal (n = 32) vs Diabetes with Prehypertension & Hypertension (n = 32) (10-fold stratified cross validation)	Naïve Bayes	0.69 ± 0.21	0.03 ± 0.09	0.025 ± 0.08	0.03 ± 0.1
	Logistic Regression	0.67 ± 0.23	0.67 ± 0.02	0.51 ± 0.02	1.0 ± 0.0
	AdaBoost Classifier	0.68 ± 0.11	0.52 ± 0.12	0.56 ± 0.17	0.53 ± 0.17
	Random Forest	0.70 ± 0.21	0.64 ± 0.16	0.63 ± 0.19	0.69 ± 0.19
	Decision Trees	0.66 ± 0.16	0.60 ± 0.25	0.58 ± 0.25	0.68 ± 0.32
	SVM (Linear Kernel)	0.74 ± 0.17	0.69 ± 0.10	0.65 ± 0.15	0.78 ± 0.18
	LDA	**0.79 ± 0.15**	**0.71 ± 0.15**	**0.74 ± 0.19**	**0.75 ± 0.23**
Only PPG Signal Features Target subjects with age >= 30. Normal (n = 31) vs Diabetes with Prehypertension & Hypertension (n = 31) (10-fold stratified cross validation)	Naïve Bayes	0.45 ± 0.2	0.05 ± 0.15	0.04 ± 0.12	0.07 ± 0.20
	Logistic Regression	0.43 ± 0.24	0.40 ± 0.25	0.35 ± 0.22	0.47 ± 0.31
	AdaBoost Classifier	0.59 ± 0.25	0.52 ± 0.32	0.52 ± 0.34	0.60 ± 0.39
	Random Forest	0.52 ± 0.21	0.44 ± 0.30	0.47 ± 0.36	0.48 ± 0.35
	Decision Trees	**0.70 ± 0.20**	**0.56 ± 0.28**	**0.66 ± 0.30**	**0.56 ± 0.34**
	SVM (Linear Kernel)	0.42 ± 0.25	0.45 ± 0.30	0.40 ± 0.30	0.63 ± 0.46
	LDA	0.64 ± 0.15	0.56 ± 0.15	0.56 ± 0.18	0.58 ± 0.16

5 Conclusion & Future Work

In this paper we have focused on integrating both physiological and short recorded PPG signal characteristics, which can be easily extracted through the help of smart devices, in order to predict Type 2 Diabetes. Through the two ANOVA tests we were able to identify the best features towards the classification of diabetes. It was identified that the e/a ratio, AI, Adjusted AI and the ratio of pulse interval to its systolic amplitude are

unique features towards the prediction of type 2 diabetes. The accuracy of predicting diabetes without the presence of hypertension or prehypertension was 83% which was achieved by the Decision Tree classifier. The confidence of the obtained results can be further improved through data collection and validation from additional subjects which is the current focus of the research. It is important to identify and incorporate additional features and samples related to diabetes in order to improve the accuracies further.

Even though biomedical signal processing approaches are present for PPG feature extraction [24] [25], they require tuning of parameters and setting thresholds which are greatly affected by noise and motion artifacts. The future research also focusses on automatic feature extraction from timeseries PPG signals, using Machine Learning to ensure the development of a robust diabetes detection system.

Ballinger et all [10] was able to achieve an accuracy of 84.51% (area under the ROC curve) incorporating medical health records, step count and continuous measurement of heart rate from PPG. Moreno, Enrique Monte, et al [26] achieved an accuracy of 69.4% using HRV and Cepstral Analysis. According to our knowledge this is the first research focusing on analyzing the easily obtainable short (~2.1s) photoplethysmography signals in order to predict diabetes in a practical setting with the presence of Hypertension & Prehypertension. The approach demonstrates great potential through the achievement of an accuracy of 79%, with precision and recall scores of 74% and 75% respectively. The potential of the PPG signal is evident through similar studies focusing on Hypertension, Arterial Fibrillation and Blood Pressure estimation. Hence, it is evident that there exists great potential in using PPG signals to develop intelligent systems for disease prediction through the utilization of smart wearable devices.

References

1. “Diabetes: Fact Sheet”, World Health Organization, 2017. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs312/en/ [Ac-cessed: 12- Apr- 2018].
2. Elgendi, Mohamed. "On the analysis of fingertip photoplethysmogram signals." Current cardiology reviews 8.1 (2012): 14-25.
3. Allen, John. "Photoplethysmography and its application in clinical physiological measurement." Physiological measurement 28.3 (2007): R1.
4. Spigulis, Janis, et al. "Potential of advanced photoplethysmography sensing for noninvasive vascular diagnostics and early screening." Clinical Diagnostic Systems: Technologies and Instrumentation. Vol. 4625. International Society for Optics and Photonics, 2002.
5. Kurylyak, Yuriy, Francesco Lamonaca, and Domenico Grimaldi. "Smartphone based photoplethysmogram measurement." Digital image and signal processing for measurement systems (2012): 135-164.
6. Matsumura, Kenta, et al. "Cuffless blood pressure estimation using only a smartphone." Scientific reports 8.1 (2018): 7298.
7. Mani, Subramani, et al. "Type 2 diabetes risk forecasting from EMR data using machine learning." *AMIA annual symposium proceedings*. Vol. 2012. American Medical Informatics Association, 2012.
8. Monte-Moreno, Enric. "Non-invasive estimate of blood glucose and blood pressure from a photoplethysmograph by means of machine learning techniques." Artificial intelligence in medicine 53.2 (2011): 127-138.

9. Reddy, V. Ramu, et al. "DMSense: A non-invasive Diabetes Mellitus Classification System using Photoplethysmogram signal." Pervasive Computing and Communications Workshops (PerCom Workshops), 2017 IEEE International Conference on. IEEE, 2017.
10. Ballinger, Brandon, et al. "DeepHeart: Semi-Supervised Sequence Learning for Cardiovascular Risk Prediction." arXiv preprint arXiv:1802.02511 (2018).
11. Liang, Yongbo, et al. "A new, short-recorded photoplethysmogram dataset for blood pressure monitoring in China." *Scientific data* 5 (2018): 180020.
12. Deurenberg, Paul, Jan A. Weststrate, and Jaap C. Seidell. "Body mass index as a measure of body fatness: age-and sex-specific prediction formulas." *British journal of nutrition*65.2 (1991): 105-114.
13. Elgendi, Mohamed. "Optimal Signal Quality Index for Photoplethysmogram Signals." *Bioengineering* 3.4 (2016):21.
14. Liang, Yongbo, et al. "An optimal filter for short photoplethysmogram signals." *Scientific data* 5 (2018): 180076.
15. Gopaul, N. K., et al. "Oxidative stress could precede endothelial dysfunction and insulin resistance in Indian Mauritians with impaired glucose metabolism." *Diabetologia*44.6 (2001): 706-712.
16. Chowienczyk, Philip J., et al. "Photoplethysmographic assessment of pulse wave reflection: blunted response to endothelium-dependent beta2-adrenergic vasodilation in type II diabetes mellitus." *Journal of the American College of Cardiology* 34.7 (1999): 2007-2014.
17. Millasseau, S. C., et al. "Determination of age-related increases in large artery stiffness by digital pulse contour analysis." *Clinical science* 103.4 (2002): 371-377.
18. Deanfield, John E., Julian P. Halcox, and Ton J. Rabelink. "Endothelial function and dysfunction: testing and clinical relevance." *Circulation* 115.10 (2007): 1285-1295.
19. Hadi, Hadi AR, Cornelia S. Carr, and Jassim Al Suwaidi. "Endothelial dysfunction: cardiovascular risk factors, therapy, and outcome." *Vascular health and risk management* 1.3 (2005): 183.
20. Poon, C. C. Y., et al. "Changes in the photoplethysmogram waveform after exercise." *Computer Architectures for Machine Perception, 2003 IEEE International Workshop on*. IEEE, 2004.
21. Takazawa, Kenji, et al. "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform." *Hypertension* 32.2 (1998): 365-370.
22. Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." *Emerging artificial intelligence applications in computer engineering* 160 (2007): 3-24.
23. Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.
24. Elgendi, Mohamed, et al. "Detection of a and b waves in the acceleration photoplethysmogram." *Biomedical engineering online* 13.1 (2014): 139.
25. Elgendi, Mohamed. "Detection of c, d, and e waves in the acceleration photoplethysmogram." *Computer methods and programs in biomedicine* 117.2 (2014): 125-136.
26. Moreno, Enrique Monte, et al. "Type 2 diabetes screening test by means of a pulse oximeter." *IEEE Transactions on Biomedical Engineering* 64.2 (2017): 341-351.