Neural Network-based Diabetic Type II High-Risk Prediction using Photoplethysmogram Waveform Analysis

Yousef K Qawqzeh

Department of Computer Science and Information
College of Science in Zulfi, Majmaah University
Al-Majmaah 11952, Saudi Arabia

Abstract—This work aims to predict and classify patients into diabetic and nondiabetic subjects based on age and four independent variables extracted from the analysis of photoplethysmogram (PPG) morphology in time domain. The study has two main stages, the first one was the analysis of PPG waveform to extract b/a, RI, DiP, and SPt indices. These parameters contribute by some means to the prediction of diabetes. They were statistically significant and correlated with the HbA1C test. The second stage was building a neural network based classifier to predict diabetes. The model showed an accuracy of 90.2% in training phase and an accuracy of 85.5% in testing phase. The findings of this research work may contribute towards the prediction of diabetes in early stages. Also, the proposed classifier showed a high accuracy in predicting the existence of diabetes in Saudi people population.

Keywords—Diabetes; prediction; classification; photoplethysmogram; neural networks; diagnosis

I. INTRODUCTION

The diabetes is considered as the most common chronic diseases in the world [1]. Normally, people having type 2 diabetes are exposed to different types of complications, which in turns leads to death [2]. In fact, the lack of insulin causes diabetes since it regulates sugar. Type 2 diabetes occurs when the insulin produced by pancreas is not effectively used by the body [3]. Changes in the lifestyle habits and foods worldwide increase the chance of having Type 2 diabetes among all societies [4-5]. Its expected by 2030 that diabetic patients will be increased by more than 100% [6]. This study utilized the photoplethysmogram (PPG) waveform analysis to extract features based on volumetric changes in PPG's contour that could by some means contribute to the prediction of type 2 diabetes. PPG signal represents the measurement of volumetric changes in blood tissues and vessels [7-8]. As there is no agreed definition about the origin of PPG waveform, it thought to represent the plethysmograph obtained optically [9]. Detecting of changes in blood volume in tissue microvascular bed can be achieved by PPG [10]. PPG's pulsatile components are related to blood volume changes inside arteries [11]. The well-known pulse oximeter is used to illuminate the skin (index finger, ear, or forehead) and then measure light absorption changes [12]. Different existed clinical applications utilized PPG technique in commercially available medical devices, such as pulse oximeters, blood pressure measurement systems,

vascular diagnostics systems [13], and HR monitoring programs [14]. However, signal processing and features extraction will be elaborated in depth in research methods section.

Furthermore, the study implements an artificial neural network (ANN) classifier to predict and diagnose diabetic cases from non-diabetic ones. ANN represents a layered network structure that consists of an input layer, hidden layer(s), and an output layer. Each layer contains several nodes called neurons, were each neuron receives, processes its input and passes the output to the next layer [15]. ANN learning process consists of a training and testing phases in which the weights values are identified during the training phase [16]. Details on the input data (independent variables) and the target (the dependent variable) are demonstrated in research methods section.

II. LITERATURE REVIEWS

Diabetes is considered as a global endemic that has a rapid increase in prevalence globally [17]. A good research work by [18] has proposed a PPG-based research work to assist the clinicians in diabetes screening and for the adopting of a suitable treatment plan towards the preventing of end-organ damage. In seeking the development of good and reliable classifiers for the prediction of diabetes, several classifiers were developed [19- 21]. A study by [22] claimed that they can use a deep learning algorithm for the detection of a prevalent diabetes by the analysis of PPG signal with a plausible difference.

The classification process is one of the most important topics in artificial intelligence area. Normally, the matter of classification happens when we want to assign an object to a predefined class depending on several attributes related to that object [23]. Classification plays a vital role in clinical diagnosis problems in which its outcomes will be used for further treatment [1].

The use of ANN in classification is considered very important in research and application areas [24]. One of the most common neural networks architecture that has been used for diseases prediction and diagnosis systems is the multi-layer neural networks architecture [25]. Additionally, a study by [26] used a deep convolutional neural networks (DCNN) to classify several types of cancer disease. To develop and evaluate a deep

neural network learning classification and prediction models has been conducted for the diagnosis of patients with coronary heart disease [27]. A new convolutional neural network for breast cancer histopathology images classification has been designed by [28].

III. RESEARCH METHODS

A. Subjects and Methods

This study aimed to study type II diabetic subjects, Saudi people at Zulfi General hospital, in a one-year time period (from Aug 2018 to Aug 2019). Normally, the patient will be diagnosed as type II diabetes if his HbA1C exceeds 6.5% [29]. The study hired a control group which has no diabetic history, and this health status is double checked by the analysis of a blood sample from each participant from the control group. A questionnaire-based data collection form is used to collect detailed information about participants, both diabetic and control subjects. Participants were advised to fast at least 6 hours before performing the HbA1C test and before the recording of PPG signal as well.

B. Data Collection

HbA1C tests are performed by analyzing a blood sample taken from participants by a specialized laboratory technician. The recorded HbA1C results are documented and labeled for each subject. The patient is then advised to move to an equipped research room to record his PPG signal using a customized PPG waveform recording system. A special instrument with a data acquisition board (NI cDAQ-9172) is used to digitize PPG signal locally and to transmit its digital data to a personal computer with a sampling rate of 5500 Hz. The recording of each PPG signal from each patient lasts for 1 minute approximately. The recording is repeated twice for each patient to allow for the selection of best PPG pulses. Subjects are requested to rest in a supine position in a quiet and controlled room temperature $(25 \pm 1C^{o})$ for 5 minutes to allow for cardiovascular stabilization. All participants received a brief orientation about the purpose of this study and about the expected outcomes of these recordings. Each subject showed his willingness to participate in this study by signing a written consent that declared his agreement. A customized algorithm is developed in MATLAB environment to analyze the PPG signals in time domain and then to extract features of interest from its morphology.

C. Measurements of Parameters

- HbA1C test: A blood sample for each subject has been taken and tested with Diazyme, enzymatic assay method by a specialized laboratory technician at Zulfi Hospital.

- The RI index: The RI index is calculated as the ratio of PPG's second wave peak to its first wave peak amplitudes. Fig. 1(a) also illustrates the calculation of RI index. The analysis of RI index variations could assist in the prediction of diseases and disorders associated with aging. A study by [7] reviewed the importance of PPG's RI index in the evaluation of atherosclerosis risk. Several studies focused on utilizing the RI index as an independent factor for the prediction

of arterial stiffness and cardiovascular diseases. Its noteworthy to state that the RI index is declared to be highly related to age, as we age, the RI index will be increased.

- The b/a index: The b/a index is calculated as the ratio of "b" valley to the "a" peak, both a and b are extracted from PPG's second derivative morphology as shown in Fig. 1(b). The b/a index gained great attention from several researchers where they were looking for facilitating the interpretation of the original PPG's waveform using its derivatives. Early work by [30] has been conducted in which it concludes that the b/a index reflects arterial stiffness and atherosclerosis. Also, a study by [31] concluded that the b/a index is associated with peripheral arterial distensibility. Thereby, the b/a index is considered to be a very important factor in the examination of arterial wall elasticity. Other studies like [7; 32-33] stated that the b/a index is highly related to atherosclerosis and aging.

- The diastolic pulse peak (DiP): The Dpp represents the amplitude of PPG's second peak (the diastolic peak). The DiP tends to be clearly pronounced in healthy young subjects, while it tends to be less seen in old and diabetic subjects [34].

- The systolic peak time (SPt): The SPt is calculated as the time the PPG's systolic wave takes to arrive to its highest peak in milliseconds. As we age, the SPt index tends to increase in which it may reflect arterial stiffness.
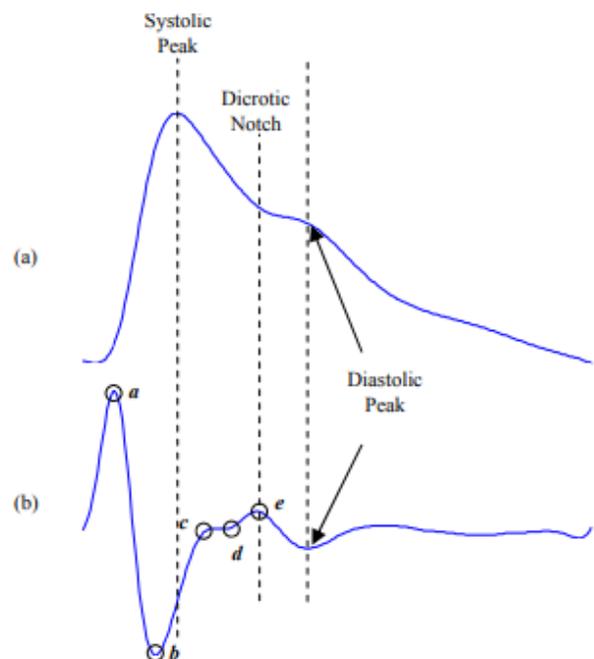

Fig. 1. The PPG Waveform were RI can be Calculated using the Systolic and Diastolic Peaks, (b) The Second Derivative of the PPG Waveform where the b/a Index can be Calculated. Source: [35].

D. Statistical Analysis

PPG parameters were expressed by their median ± SD values. Statistical analysis was implemented in the statistical package SPSS (version: IBM SPSS Statistics 25.0 for windows). Pearson correlation was used to examine the associations between the dependent (HbA1C) and the independent parameters. The value of P parameter <0.05 was considered statistically significant. Feedforward neural networks classifier using SPSS was used to predict the existence of HbA1C using the independent variables (IVs).

The neural network consists of one input layer, one hidden layer and one output layer. The input layer contains five independent variables namely b/a index, Age, RI, SPt, and DiP. Each one of them contributes by a certain amount towards the prediction of the dependent variable (HbA1C). The hidden layer utilized a Sigmund function that processes the inputs and classifies them into a diabetic or non-diabetic patient. The output layer falls into two binary outcomes, diabetic or non-diabetic.

IV. RESULTS AND DISCUSSIONS

This section presents the multilayer perceptron. The case processing summary for the data is illustrated in Table I which shows the percent of the training sample (77%) and the percent of the testing sample (23%).

The network's information is explained in Table II which describes the input layer, hidden layer, and output layer elements. The input layer contains the following covariates as inputs parameters which are b/a index, Age, RI index, SPt index, and DiP index respectively. This network has one hidden layer that has three units with a Hyperbolic tangent as an activation function. Also, the output layer used the Softmax as an activation function to classify the dependent variable (the A1C index).

The network architecture is shown in Fig. 2. The architecture consists of three layers where the first column from the figure's left side represents the independent variables (inputs). The five indices in the input layer along with the Bias are used for classification. The output of each neuron (each input) is passed to all next neurons in the hidden layer which are multiplied by some weight and their product will be used to classify the samples into one binary class either diabetic or non-diabetic patient using a Softmax activation function.

In Table III, the model summary is elaborated which shows the performance of the model by looking to percent incorrect predictions error percentage in the training phase (9.8%) and in the testing phase (14.5%). The results claimed an accuracy of 90.2% in predicting the correct class in training time, and an accuracy of 85.5% in predicting the correct class during the testing time. As the error percent is approximately small in this model, we assume that this model performs well in predicting and classifying the patient into a diabetic or non-diabetic subject.

Parameter estimates in Table IV has two sections, the first left-upper part represents the result from the input layer to the hidden layer (predictors multiplied by weights), while the second right-down part represents the result from the hidden layer to the output layer. However, the classification results listed in Table V describe the overall percentage of model performance. In the training phase, the model classified 68 subjects correctly while 9 subjects were incorrectly classified as nondiabetic with an overall percentage of 90.2%. Besides, during the testing phase, the model performs well in diagnosing the subjects into 20 diabetic patients correctly while 3 subjects were classified incorrectly as nondiabetic patients with an overall percentage of 85.5%. Moreover, the model showed a better ability to predict the nondiabetic subjects by a percentage of 91.6% during the training phase where 9 subjects only have been classified incorrectly as diabetic patients.

TABLE. I.   CASE PROCESSING SUMMARY

		N	Percent
Sample	Training	184	77.0%
	Testing	55	23.0%
Valid		239	100.0%
Excluded		0	
Total		239	

TABLE. II.   NETWORK INFORMATION

Input Layer	Covariates	1	ba
		2	Age
		3	RI
		4	SPt
		5	DiP
	Number of Units[a]		5
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		3
Output Layer	Dependent Variables	1	A1C
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

a. Excluding the bias unit
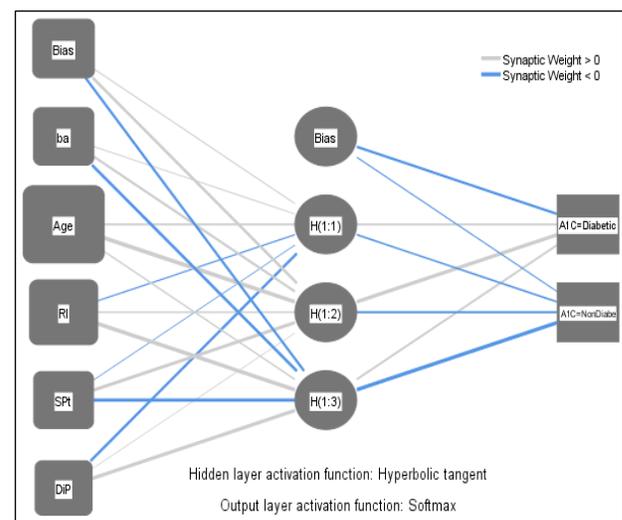
Fig. 2.   The Architecture of the Developed NN Classifier.

TABLE. III. MODEL SUMMARY

Training	Cross Entropy Error	49.419
	Percent Incorrect Predictions	9.8%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.05
Testing	Cross Entropy Error	18.105
	Percent Incorrect Predictions	14.5%
Dependent Variable: A1C		

a. Error computations are based on the testing sample

TABLE. IV. PARAMETER ESTIMATES

(Predicted)					
Predictor	H(1:1)	H(1:2)	H(1:3)	[Diabetic]	[Non diabetic]
Input Layer	(Bias)	.136	.819		
	ba	.063	.626		
	Age	.581	1.555		
	RI	-.259-	.431		
	SPt	-.042-	.957		
	DiP	-.738-	.013		
Hidden Layer 1	(Bias)			-.651-	-.244-
	H(1:1)			.480	-.455-
	H(1:2)			1.032	-.621-
	H(1:3)			.582	-1.077-

TABLE. V. CLASSIFICATION

Sample	Observed	Predicted		
		Diabetic	NonDiabetic	Percent Correct
Training	Diabetic	68	9	88.3%
	NonDiabetic	9	98	91.6%
	Overall Percent	41.8%	58.2%	90.2%
Testing	Diabetic	20	3	87.0%
	NonDiabetic	5	27	84.4%
	Overall Percent	45.5%	54.5%	85.5%

The classification of diabetic and nondiabetic patients is shown in Fig. 3. The boxplot demonstrates the Tukey's five numbers summary which are the minimum, first quartile (Q1), median, third quartile (Q3), and the maximum. For the diabetic group, the median value of those who are diagnosed correctly as a diabetic patient was 0.9. In the other hand, the median value of those who are diagnosed correctly as a nondiabetic patient was almost 0.95.

Moreover, a very interesting output from the developed model to look at is the independent variable importance which is illustrated in Table VI and in Fig. 4. Both Table VI and Fig. 4 showed that Age index (0.338) was the dominant index among the independent variables that contribute by some means to the prediction of diabetes as we age. The second important factor was RI index (0.225) in which it scored the second place of importance after Age. Finally, b/a index, SPt index and DiP index were contributed by a less amount towards the prediction of diabetes by 0.56, 0.152, and 0.129

percent, respectively. Note that the summation of all independent variables importance percent equals to 1.
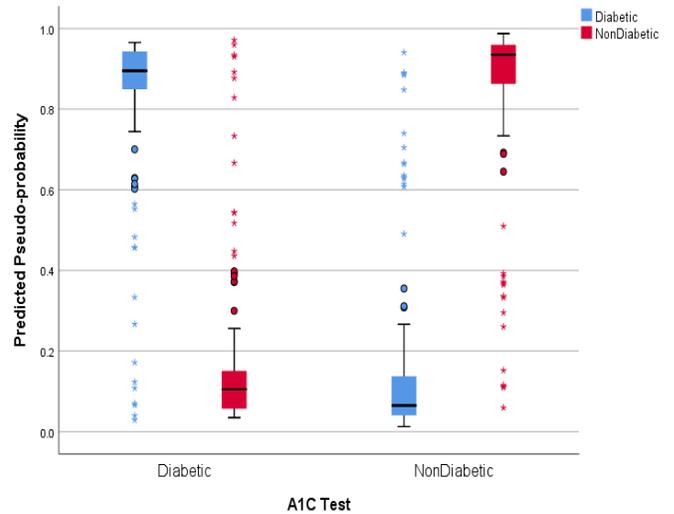

Fig. 3. Boxplot Diagram for the Classification of Diabetic and Nondiabetic Patients.

TABLE. VI. INDEPENDENT VARIABLE IMPORTANCE

|	Importance	Normalized Importance
ba	.156	46.2%
Age	.338	100.0%
RI	.225	66.5%
SPt	.152	45.0%
DiP	.129	38.0%
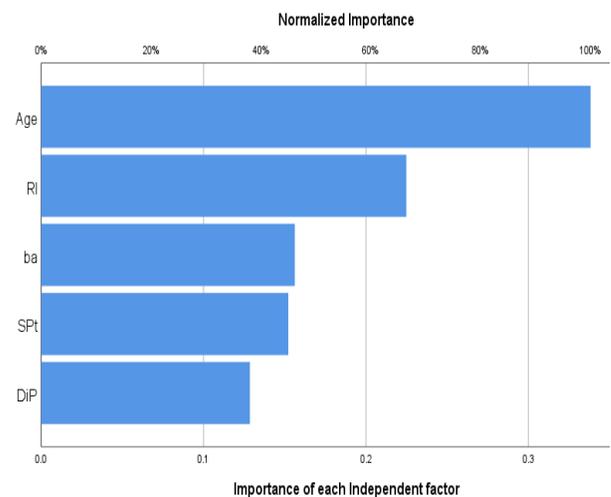
Fig. 4. The Normalized Importance of Each of the Dependent Variables in the Prediction of the Dependent Variable.

V. CONCLUSIONS

In this work, we recorded a PPG signals from two participant groups. The first one with established and clinical confirmed diabetic, while the second one with no history proof of diabetes. Their recorded PPG waveforms were analyzed in a customized algorithm to extract the parameters of interest which are RI, B/a, DiP, and SPt indices, respectively. After that

an HbA1C test was conducted for all diabetic and nondiabetic groups and their measurement were recorded and documented. A Pearson correlation test was fired to examine the associations between the independent variables and the dependent variable. The study then performed a neural network classification model to investigate and examine the amounts of contribution that each independent variable did towards the prediction of diabetes. The developed classifier showed a high degree of accuracy in both training phase (90.2%) and in testing phase (85.5%) in which it makes the developed classifier a reliable model for diabetes evaluation and early diabetes stages prediction. The developed model may aid in the prevention of diabetes by predicting the high-risk diabetes in its early stages, which in turn might deliver a great benefit to both people and doctors in terms of prediction and classification.

REFERENCES

[1] K. Kannadasan, E. Damodar, and K. Venkatanareshbabu, "Type 2 diabetes data classification using stacked autoencoders in deep neural networks," Clinical Epidemiology and Global Health, 2018.

[2] O Abdulfatai, O. Olusegun, and O. Lateefat, "Type 2 Diabetes Mellitus: A Review of Current Trends," Oman Medical Journal, vol. 27, no. 4, 2012, pp269-273

[3] H. Chen, C. Tan, Z. Lin, and T. Wu, "The diagnostics of diabetes mellitus based on ensemble modeling and hair/urine element level analysis," Comput Biol Med, vol. 50, 2014, pp70–75

[4] EI. Mohamed, R. Linder, G. Perriello, ND. Di, S. Poppl, and AL. De, "Predicting type 2 diabetes using an electronic nose-based artificial neural network analysis," Diabetes, nutrition & metabolism, vol. 15, no. 4, 2002, pp215–221

[5] K. Polat and S. Gunes, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," Digit Signal Process, vol. 17, no. 4, 2007, pp702–710

[6] P. Manaswini and S. Ranjit, "Predict the onset of diabetes disease using Artificial Neural Network (ANN)," International Journal of Computer Science & Emerging Technologies, vol. 2, I 2, 2011, pp303-311

[7] Y. Qawqze, U. Rubins, and M. Alharbi, "Photoplethysmogram second derivative review: Analysis and applications," Scientific research and essays, vol. 10, no. 21, 2015, pp633-639

[8] A. Reisner, P. Shaltis, D. McCombie, and H. Asada, "Utility of the Photoplethysmogram in Circulatory Monitoring," Anesthesiology, vol. 108, no. 5, 2008, pp950-958

[9] J. Moraes et al, "Advances in photopletysmography signal analysis for biomedical applications," Sensors, vol. 18, no. 1894, 2018

[10] I. Challoner, "Non-Invasive Physiological Measurements," ed P Rolfe (London: Academic), 1979, pp125-151

[11] M. Elgendi, "The use of photoplethysmography for assessing hypertension," npj Digit. Med, vol. 2, no. 60, 2019, pp1-11

[12] K. Shelley and S. Shelley, "Pulse Oximeter Waveform: Photoelectric Plethysmography in Clinical Monitoring," Clinical Monitoring: Practical Applications for Anesthesia and Critical Care, pp420-428.

[13] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," Physiol. Meas, vol. 28, ppR1-R39.

[14] Md. Shofiqul Islam, Md. Shifat-E-Rabbi, A. Dobaie, and Md. Hasan, "PREHEAT: Precision heart rate monitoring from intense motion artifact corrupted PPG signals using constrained RLS and wavelets," Biomedical Signal Processing and Control, vol. 38, 2017, pp212-223

[15] A. Al-Jarrah and H. Shatnawi, "Non-proliferative diabetic retinopathy symptoms detection and classification using neural network," Journal of medical engineering & technology, vol. 41, no. 6, 2015, pp498–505.

[16] K. Verma, L. Verma, and P. Tripathi, "Image classification using backpropagation algorithm," J Comput Sci Soft Appl, vol. 1, 2014, pp7–15.

[17] SI, Sherwani, HA. Khan, A. Ekhzaimy, A. Masood, and MK. Sakharkar, "Significance of HbA1c Test in Diagnosis and Prognosis of Diabetic Patients," Biomark Insights, vol. 11, 2016, pp95–104.

[18] N. Neelamshobha, R. Periyasamy, S. Bikesh, and K. Awanish, "Detection of type-2 diabetes using characteristics of toe photoplethysmogram by applying support vector machine," Biocybernetics and Biomedical Engineering, vol. 39, no. 1, 2019, pp38-51.

[19] E. Ubeyli, "Comparison of different classification algorithms in clinical decision-making," Expert Syst, vol. 24, 2007, pp17-31.

[20] J. Kandhasamy and S. Balamurali, "Performance analysis of classifier models to predict diabetes mellitus," Proc Comput Sci, vol. 47, 2015, pp45-51.

[21] N. Nai-Arun and R. Moungmai, "Comparison of classifiers for the risk of diabetes prediction," Proc Comput Sci, vol. 69, 2015, pp132-142.

[22] A. Robert, T. Geoffrey, K. Peter, M. Gregory, P. Mark, O. Jeffrey , and A. Kirstin, "Predicting diabetes from photoplethysmography using deep learning," J Am Coll Cardiol, vol. 73, no. I 9, s. 2, 2019.

[23] G. P. Zhang, "Neural networks for classification: a survey," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 30, no. 4, 2000, pp451-462.

[24] K. Saravanan and S. Sasithra, "Review on classification based on artificial neural networks," International Journal of Ambient Systems and Applications (IJASA), vol. 2, no. 4, 2014, pp11-18.

[25] Er. Orhan, A. Tanrikulu, A. Abakay, F. Temurtas, "An approach based on probabilistic neural network for diagnosis of Mesothelioma's disease," Computers & Electrical Engineering – CEE, vol. 38, 2015, pp75-81.

[26] H. Nogay, "Classification of Different Cancer Types by Deep Convolutional Neural Networks," Balkan journal of electrical & computer engineering, 2018, pp54-59.

[27] M. Kathleen and M. Julia, "Coronary Heart Disease Diagnosis using Deep Neural Networks," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, no. 10, 2018, pp1-8.

[28] Y. Jiang, L. Chen, H. Zhang, and X. Xiao, "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," PLoS ONE, vol. 14, no. I 3, 2019, pp2-21.

[29] MayoClinic, "Type 2 diabetes," Retrieved form [https://www.mayoclinic.org/diseases-conditions/type-2-diabetes/diagnosis-treatment/drc-20351199] on Sep 25Th 2019.

[30] TN. Takazawa, M. Fujita, O. Matsuoka, T. Saiki, M. Aikawa, S. Tamura, and C. Ibukiyama, "Assessment of vasocative agents and vascular aging by the second derivative of photoplethysmogram waveform," Hypertension, vol. 32, 1998, pp365-70.

[31] I. Imanaga, H. Hara, S. Koyanagi, and K. Tanaka, "Correlation between wave components of the second derivative of plethysmogram and arterial distensibility," Jpn Heart J, vol. 39, 1998, pp775-84.

[32] B. Peskin and R. Rowen, "Breakthrough in Clinical Cardiology: InOffice Assessment with Pulse Wave Velocity (PWV) and Digital Pulse Analysis (DPA)," Clin. Cardiol, 2010, pp80-86.

[33] A. Alberto, "The finger volume pulse and the assessment of arterial properties," J. Hypertens, vol. 20, 2002, pp2341-2343.

[34] G. Papini, P. Fonseca, X. Aubert, S. Overeem, J. Bergmans, and R. Vullings, "Photoplethysmography beat detection and pulse morphology quality assessment for signal reliability estimation," 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, 2017, pp117-120.

[35] M. Algendi, "On the Analysis of Fingertip Photoplethysmogram Signals. Current Cardiology Reviews, vol. 8, 2012, pp14-25.