Journal of Physics: Conference Series **1450** (2020) 012052 doi:10.1088/1742-6596/1450/1/012052

K-NN based algorithm for degree of stenosis classification using dual non-invasive photoplethysmography system

A Stephanus[1], Y C Du[2]

[1] Department of Electrical Engineering, Politeknik Negeri Ambon, Jalan Ir. M. Putuhena, Ambon, Indonesia
[2] Department of Electrical Engineering, Southern Taiwan University of Science and Technology, Nantai Street, Tainan City, Taiwan

E-mail: alphinstephanus@polnam.ac.id

Abstract. Haemodialysis (HD) patients who undergo long-term treatment are very susceptible to arterial stenosis. In this study, we propose two main features taken from patients undergoing the dialysis process, namely: Rising Slope – RS and Falling Slope - FS. These features are yielded from a photoplethysmography signal extraction on the hand that is used to create vascular access called HD hand. Eleven dialysis patients with the arteriovenous fistula (AVF) method were the object of this study. The feature data was taken twice, before and after the dialysis process. Utilizing the t-test on the variance of RS features on the HD hand showed a statistically significant value of 0.0211 ($p < 0.05$). Furthermore, these RS features are used as input for K-NN classifiers to classify degrees of stenosis in patients undergoing HD. Four patient data with RS features from the HD hand that was not previously known became this classifier test data. From the experimental results showed that K-NN with Euclidean and Minkowski distance could classify the degree of stenosis well. The percentage of misclassification of the system against unknown data was 18 percent (82 percent accurate) based on the cross-validated classification accuracy.

1. Introduction

More than 300,000 individuals in the United States rely on a vascular access to receive haemodialysis (HD) treatment [1]. HD is the most common treatment for patients with End-stage renal disease (ESRD). In the midst of this treatment the blood is purified from unused products and excess fluid is removed by using a dialyzer. To attain the blood, a vascular access, usually placed on one of the patient's forearms is used to insert the needles coming from the dialyzer. A very conventional type of vascular access called arteriovenous (AV) fistula is made through a surgical operation during which an artery and a vein in the arm are connected together. The connection point is referred to as an anastomosis and is often located near the patient's wrist or elbow. The state of the fistula may deteriorate in course of time. The most common form of fistula failure is venous stenosis [2].

As a result of repeated puncturing of this access every two days or long term use, access occlusion and failures are caused by inadequate arterial inflow or venous outflow occlusion. This causes thrombosis, resulting in intimal hyperplasia, chronic fibrin, cellular deposit, aneurysm and limb ischemia [1-3]. When it happens, the fistula must be revised and remedied. An early detection of stenosis is become important since it prevent total occlusion and thereby prolongs the life of the fistula.

PPG is an optical measurement technique that can be used non-invasively to sense blood volume changes in the microvascular bed of tissue. A PPG waveform conveys some specific physiological information such as cardiac synchronous changes in the blood volume with each heartbeat, thermoregulation, respiration, and vasomotor activity [4]. PPG signals have used multi-channel data measurement at the ear lobes, index fingers, thumbs and great toe sites. These measurements can be used in computer-based digital signal processing (DSP) and pulse waveform analysis [5]. There for, this optical technique is a promising solution for detecting arteriovenous fistula (AVS) stenosis monitoring. Angiography can be used to evaluate the level of clinical vascular stenosis [6]. The disadvantage of angiography is that it is invasive and requires surgery. Doppler ultrasound is also used to evaluate the diameters of an AVF [7]. However, both methods and their equipment are expensive and require experienced and advanced operators.

In recent times, some research attempted to quantify and evaluate AVF stenosis by numerous ways. In 2009, Vasquez *et al*. used wavelet and support vector machines to detect AVF stenosis by the blood flow sounds [2]. In 2014, Hsien Yi Wang *et al.* attempted to extract varying feature of blood flow sound of AVF stenosis using time and frequency domain analysis [8]. Yi Chun Du *et al.* used Sprott system to design a self-synchronization error formulation (SSEF) to quantify the differences in changes of blood volume between the right-side and left-side PPG signals. Bilateral PPGs have significant differences in rise time (RT) and amplitudes (AMPs), proportional to the degree of stenosis (DoS) [9]. When vascular access becomes obstruction on one side, bilateral PPG pulses gradually become asynchronous with each heartbeat. In addition, an arteriovenous access (AVA) has significant inflow or outflow stenosis and has the intra-flow decreases [10]. There are many features have been examined, including PPG rise time, pulse transit time (PTT), amplitude, shape and width ratio. These features can be provided to detect AVF stenosis [11].

In this paper, the Rising Slope - RS feature extracted from PPG signals is proposed as an input to the K-NN classifier to classify the degree of stenosis of an arteriovenous fistula (AVF). Eleven patients who underwent the HD process were the subject of this study. Feature data is taken before the subject undergoes and after the HD process.

2. Materials and methods

2.1. Materials

2.1.1. Protocol of experiment. The subjects in this study were eleven patients who underwent HD treatment and were recruited through the Institutional Review Board (IRB): VGHKS17-CT3-11 of Kaohsiung Veterans General Hospital (KVGH) with mean ± SD age of 77 ± 10.8. In conducting the research each subject was asked to relax in an HD treatment room with a temperature of 25 ± 1◦C.

2.1.2. Signal acquisition. The PPG probe was mounted within the clip at the right and left index finger to acquire PPG signals, as shown in Figure 1(a). The dual PPG signal was obtained using two channels of PPG signals were synchronously captured at a sampling rate of 1 kHz from the measurement system to an embedded system (MSP430, Texas Instruments) which provided an analog-to-digital (A/D) data transfer to a laptop PC with MATLAB software for further analysis. The MSP430 also locate each pulse foot (PF)-pulse foot (PF) interval to acquire sampling data within a sampling window, as shown in Figure 1(b) and 1(c). It can provide a promising continuous measurement during hemodialysis treatment.

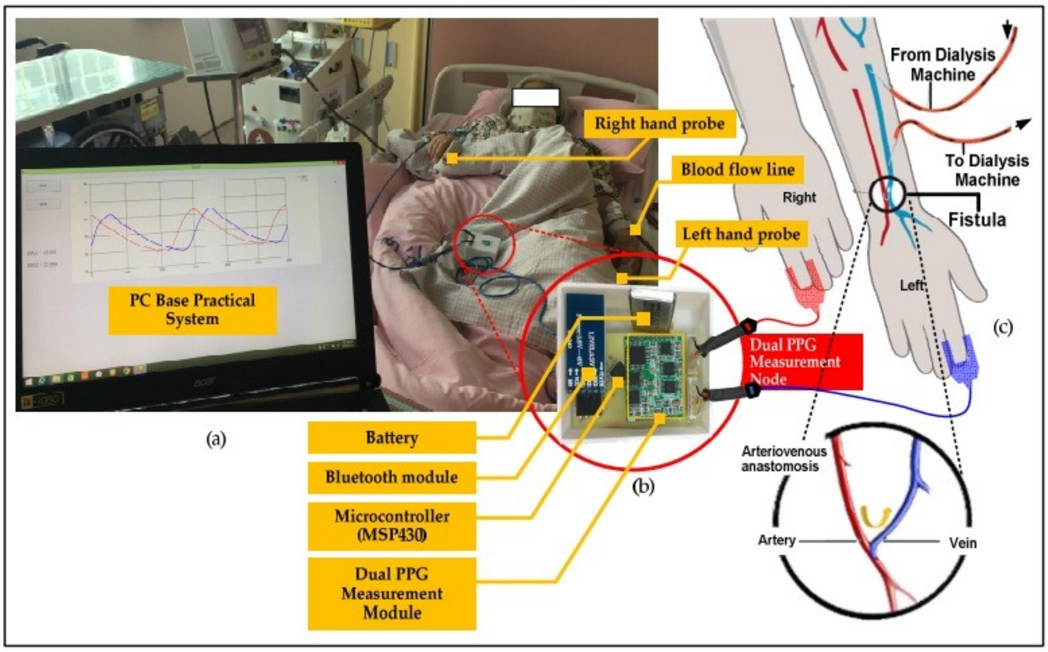

Figure 1. Dual noninvasive photoplethysmography system: (a) PC based system and probe placement; (b) Dual PPG module; (c) HD hand with probe.

2.2. Methods
As seen in Figure 2, the method used in this study consists of four stages: preprocess, feature generation, feature selection and classification.

2.2.1. Preprocessing. The method in this study began with the preprocessing stage where the PPG signal was obtained from eleven patients via a dual PPG module for 1 minute for analyzing purposes. To filter PPG signals from the surrounding environment noise a moving average low pass filter is implemented.

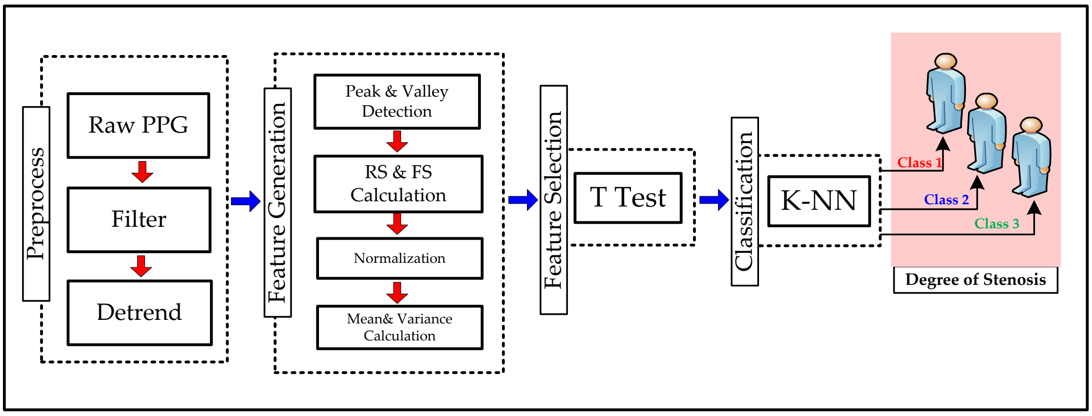

Figure 2. The method used in the study.

2.2.2. Feature generation. At this stage, the PPG signal is detected peaks and valleys through the calculation of the location of the local maxima and then de-trended to eliminate outliers, drifts, and offsets. With this detection data, the Rising slope - Rs and Falling slope - FS values are then calculated with Equations (1) and (2) respectively.

$$\text{Rising Slope (RS)}=\frac{V_p - V_n}{T_{pn} - T_n} \tag{1}$$

$$\text{Falling Slope (FS)}=\frac{V_{n+1} - V_p}{T_{n+1} - T_{pn}} \tag{2}$$

RS and FS feature calculations are calculated for both patients' hands, before and after dialysis treatment as depicted in Figure 3. Normalization of the data obtained is then calculated scaling for: 0 and 1 with Equation (3):

$$x_{R,F} = \frac{x_{ik} - \min(x_k)}{\max(x_k) - \min(x_k)} \tag{3}$$

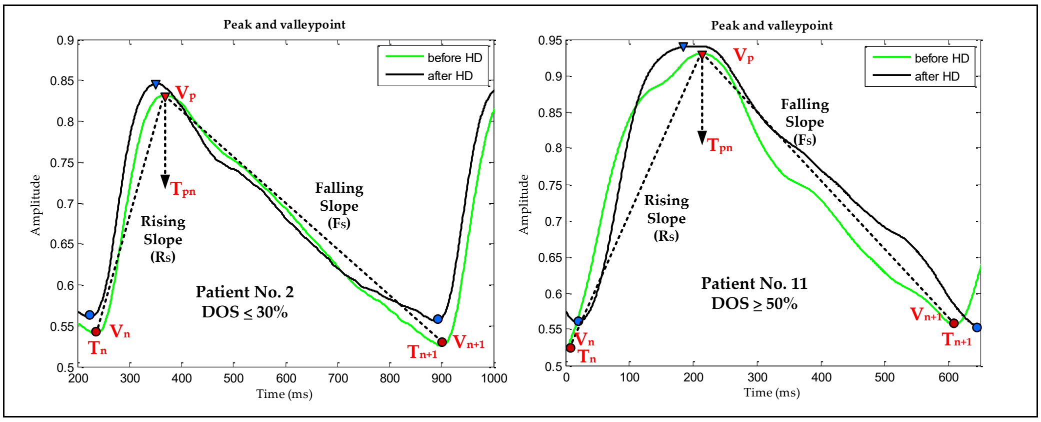

Figure 3. PPG signal taken from two patients for RS and FS calculation.

2.2.3. Feature selection. To produce a good classifier, the selected features are needed. This stage tried to get the appropriate feature and was done by calculating the mean value, the variance in sequence, and then the statistical significance is calculated using the t-test. The equation used is as follows:

$$\bar{X}_{b,a} = \sum \frac{x_i}{n} \tag{4}$$

$$S^2_{b,a} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n - 1} \tag{5}$$

$$t_{b,a} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}} \tag{6}$$

2.2.4. Classification

2.2.4.1. Degree of Stenosis (DoS). The DoS, in clinical research terms mean a degree index of narrowing normal vessel of AVF subjects and it measured by B-Mode ultrasound or Angiography images. These experiments followed IRB approval procedure. Refer to previous research [12-14], the DoS calculated using Equation (7):

$$\text{DoS } \% = \left(1 - \frac{d^2}{D^2}\right) \times 100\% \tag{7}$$

which D is the diameter of the normal vessel in direction of blood flow while d is the diameter of the stenosis lesion. If DoS is 100% means total occlusion, over 50% usually means need to be surgical treatment and between 30% to 50% means it may have effect upon the efficiency during hemodialysis. Thus, we followed clinical comment to design three different classes/grades for classification as shown in Table 1.

Table 1. Class partition based on *DoS*

DoS	Class/Grades
$DoS \leq 30\%$	1
$30\% \leq DoS \leq 50\%$	2
$DoS \geq 50\%$	3

2.2.4.2. K-NN Classification. K-NN is short for k-Nearest Neighbors which is a machine learning algorithm for classifying new objects based on a number of k closest neighbors. The purpose of using K-NN is to predict the object whether it belongs to one particular group or another group [15]. The formal K-NN algorithm used in this study is as follows:

Algorithm 1. Formal definition of K-NN

Input:

- K (the value of the selected nearest neighbor, typically small positive integer)
- Set of training with N samples dan M known cluster (N >> M): X_i, c_i i = 1,2,…,N. Where ci ∈ 1,2, …, M.

Algorithm :

For a given new labelled sample,

1. calculate the distance for each set of training.
2. choose the closest sample k based on distance calculation.
3. count the vote for the chosen cluster
4. Assign a client to the new unlabeled sample using the majority vote.

As a comparison of accuracy in K-NN classifiers, in this study three distance calculation are used, i.e., Euclidean distance, Manhattan distance and Cosine coefficient.

The Euclidean distance compares the minimum distance of the testing image with the training image database. The Euclidean distance of the two vectors x and y is calculated by Equation (8):

$$d(x,y) = \left(\sum_i (x_i - y_i)^2\right)^{\frac{1}{2}} \tag{8}$$

Manhattan distance is one of the most widely used measurements including replacing quadratic differences by adding up absolute differences from variables. This procedure is called absolute block or better known as city block distance which is shown as Equation (9):

$$d(x,y) = L_p = i(x\,y) = \sum_{i}^{n} \|x_i - y_i\| \tag{9}$$

The cosine coefficient is the most popular method of measuring similarity, calculating the angle between the vector and the query vector. If a vector is a unit of length, cosine of the angle between them is only the dot product of the vector, the equation is as follows:

$$cosine\ (X,Y) = \frac{\sum_{i=1}^{n} X_i \times Y_i}{\sqrt{\sum_{i=1}^{n}(X_i)^2} \times \sqrt{\sum_{i=1}^{n}(Y_i)^2}} \tag{10}$$

3. Results and discussion

3.1. Feature generation and selection

As explained earlier, there are 4 stages in this study, as shown in Figure 2. At the feature generation and feature selection stages the RS and FS values are generated from the local maxima calculation and then normalized. Table 2 follows outlining eleven patients who were subjected to the training.

Table 2. The data of HD patients.

Patient No.	Age	Gender		HD Hand		D	d	DOS (%)	Class (Training)
		Male	Female	Right	Left				
1	81	√			√	1.04	0.94	18.31	[1 0 0]
2	78	√			√	1.18	1.09	14.67	[1 0 0]
3	87	√			√	1.36	1.17	25.99	[1 0 0]
4	86	√			√	0.80	0.70	23.30	[1 0 0]
5	65	√			√	1.71	1.45	28.10	[1 0 0]
6	81	√			√	0.73	0.59	34.68	[0 1 0]
7	54	√			√	1.16	0.83	48.80	[0 1 0]
8	67	√			√	1.08	0.81	44.30	[0 1 0]
9	87	√			√	0.78	0.58	44.38	[0 1 0]
10	75		√		√	0.88	0.44	74.72	[0 0 1]
11	86	√			√	0.98	0.23	94.50	[0 0 1]

T-test is a technique in statistical science to determine whether the two groups tested are statistically different from one another. From the experiments it was showed that the RS features before and after on HD hand were statistically significant ($p < 0.05$, t-test = 0.0211) and could be used as a reliable feature set. This condition shown in Table 3.

Table 3. RS feature on HD hand.

Patient No.	Feature				Class
	Rising Slope – RS		Falling Slope - FS		
	BH	AH	BH	AH	
1	0.0792	0.1139	0.1233	0.1348	1
2	0.0724	0.0917	0.0920	0.1254	1
3	0.0867	0.0865	0.0847	0.0714	1
4	0.0789	0.0778	0.1037	0.0930	1
5	0.0854	0.1023	0.1572	0.0937	1
6	0.1058	0.1058	0.0950	0.0950	2
7	0.0751	0.0958	0.0871	0.0757	2
8	0.1194	0.1307	0.0686	0.1255	2
9	0.1194	0.1243	0.0686	0.1254	2
10	0.1020	0.1035	0.0770	0.1254	3
11	0.0874	0.0867	0.1146	0.0790	3
T-test	0.0211		0.5893		
Statistical Hypothesis	Significant		Not Significant		

Note: BH = Before HD, AH= After HD

3.2. Classification

The system developed in this study was trained by using data from 11 patients whose statistical significance. From Table 2, there are three degrees of stenosis classes which forms the classification model. The final stage carried out in the development of this system is performance testing. This is done by creating a cross-validated classifier from the model and testing it using unknown data. Table 4. shows that with the values k = 1, k = 2 and k = 3, the system built could predict the class correctly except by using the Cosine coefficient calculation (all classes predicted as class 001). However, the Cross-Validation Loss Check shows that the value of k = 3 indicates that the system built will predict incorrectly by 18% or correctly prediction accuracy of 82%.

Table 4. Performance of the system test.

Distance	k	Class				Cross Validation Loss Check	Accuracy
		patient A	patient B	patient C	patient D		
		[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]		
		Predicted Class					
Euclidean	1	[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]	0.64	0.36
Manhattan	1	[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]		
Cosine	1	[1 0 0]	[1 0 0]	[1 0 0]	[1 0 0]		
Euclidean	2	[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]	0.36	0.64
Manhattan	2	[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]		
Cosine	2	[1 0 0]	[1 0 0]	[1 0 0]	[1 0 0]		
Euclidean	3	[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]	0.18	0.82
Manhattan	3	[1 0 0]	[1 0 0]	[0 1 0]	[0 1 0]		
Cosine	3	[1 0 0]	[1 0 0]	[1 0 0]	[1 0 0]		

4. Conclusions

The RS feature on HD hands could significantly be a good feature set in K-NN classifier. This is evidenced by the accuracy value of more than 80 percent. To get better results, there are several things that become recommendations increase the number and variation of research subjects in this case are patients with HD treatment, compared with several methods of extraction and feature selection, and increase the number of data sets for training. The proposed system is a low-cost device and simple, therefore it has the potential to be developed as a medical diagnostic support system in a developing country.

5. References

[1] National Kidney Foundation 2006 *Updates Clinical Practice Guidelines and Recommendations* (United States: National Kidney Foundation)
[2] Vesquez P O, Marco M and Mandersson B 2009 *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*
[3] Wu J, Lin C, Du Y, Chen P, Shih C and Chen T 2015 *IEEE International Ultrasonics Symposium*
[4] Wu J, Lin C, Du Y and Chen T 2012 *IET Science, Measurement & Technology* **6** 533-540.
[5] John A 2007 *Physiological Measurement* **28**
[6] Bash S, Villablanca J P, Jahan R, Duckwiler G, Tillis M, Kidwell C, Saver J and Sayre J 2005 *American Journal of Neuroradiology* **26** 1012-1021.
[7] Tessitore N, Bedogna V, Gammaro L, Lipari G, Poli A, Baggio E, Firpo M, Morana G, Mansueto G and Maschio G *American Journal of Kidney Diseases* **42** 331-341.
[8] Wang H Y, Wu C H, Chen C Y and Lin B S 2014 *IEEE Transactions on Biomedical Engineering* **61** 1851-1857
[9] Wu J X, Lin C H, Wu M J, Li C M, Lim B Y and Du Y C 2015 *Healthcare Technology Letters* **2** 64-69
[10] Leon C and Asif A 2007 *Clinical Journal of the American Society of Nephrology* **2** 175-183
[11] Wu J X, Li C m, Ho Y R, Wu M J, Huang P T and Lin C H 2015 *IEEE Sensors Journal* **16** 2691-2700
[12] Wu J X, Lin C H, Du Y C, Chen P J, Shih C and Chen T 2015 *IEEE International Ultrasonics Symposium (IUS)*
[13] Du Y C and Stephanus A 2016 *Micromachines* **7**
[14] Du Y C and Stephanus A 2018 *Sensors* **18**
[15] Primartha R 2018 *Belajar Machine Learning Teori dan Praktik* (Bandung: Informatika)

Acknowledgments

The authors gratefully acknowledge DR. Ming-Jui Wu for providing the dataset used in this study.