Clustering of Photoplethysmography Data Signals for Developing Noise Filters

1st Rifqi Abdillah
Department of Informatics. Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
6025211012@mhs.its.ac.id

2nd Riyanarto Sarno
Department of Informatics. Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

3rd Taufiq Choirul Amri
Department of Informatics. Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
60252110009@mhs.its.ac.id

4th Faris Atoil Haq
Department of Informatics. Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
faris.4haq@gmail.com

5th Kelly Rossa Sungkono
Department of Informatics. Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
kelly@its.ac.id

6th Dwi Sunaryono
Department of Informatics. Faculty of Intelligent Electrical and Informatics Technology
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
dwis@if.its.ac.id

2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) | 978-1-6654-5645-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAIIC57133.2023.10066966

***Abstract*—This paper aims to evaluate Photoplethysmography signals taken using fingertip pulse waves on human fingers which are generally not all in good condition. In common devices, the data obtained does not only contain photoplethysmography signals, but some noise also contaminates it. Noise is an unwanted ripple-shaped signal existing in signal transmission. Noise will interfere the desired quality of the received signal and ultimately change the information contained in the signal. This situation requires improvements to the photoplethysmography signal to make the signals are in the best condition so machine learning produces a more optimal output. Noise filters cannot be done with the same treatment because noise level in each data is different and must have different filter weights. This paper proposes a method to filter noise based on the level of noise in the signal. The approach taken in this study uses two stages, clustering and noise filtering. The first approach is clustering using the K-means clustering method by utilizing the coefficient of variation and slope features to group signals based on their noise level. The second approach uses exponential filtering, which performs by weighting the filter based on the cluster so that the data have different adjustments ratio of the level of smoothing. The result of the signal-to-noise ratio on Non-filtered Data is 181.49. Signal to noise ratio on the Constant Weighted Filter is 183.79 and increases to 187.48 after using the Clustered and Weighted Filter method.**

Keywords—photoplethysmography, signal processing, clustering, noise filtering, k-means, exponential filter

I. INTRODUCTION

Using Photoplethysmography signal as a medium to determine someone's cardiovascular system is commonly known in medical technology. The condition of people' cardiovascular can be checked by measuring changes of blood volume in skin tissues [1]. Photoplethysmography signal measurements generally use two components: the emitters and receivers/detectors. The emitter emits light through the human tissue layer, to then be reflected and recaptured by the detector, as shown in Fig. 1. Whereas, The detector will provide a signal in the form of an electric voltage.

Photoplethysmography signals are used to measure physiological parameters of human body, such as heart rate, oxygen saturation, respiration rate, blood pressure, and arterial stiffness [2]. Photoplethysmography signals are usually used in medical devices, especially in practical medical devices such as wearable health equipment [3]. The utilization of photoplethysmography in any device needs to be supported with good quality of signals to gain better result. That good signals will however be significantly impactful for improving monitoring, monitoring, and diagnostic techniques for medical devices.

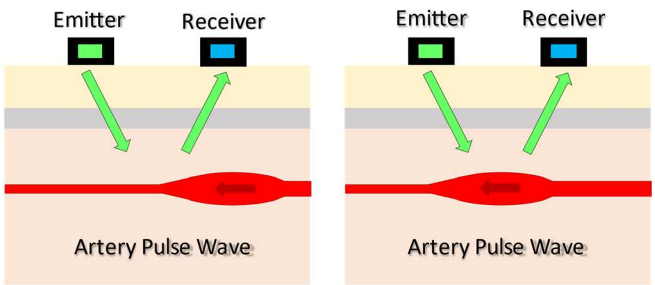


Fig. 1 Photoplethysmography signal data acquisitions

In contrary, not all signals are obtained in good condition in direct implementation in the field using fingertip pulse waves. Some are already disturbed by the interferences, in which it makes the signals have some noise [4]. Noise is unwanted signals in a communication or information system. Noise can interfere the received signal's quality and affect the received signal's sensitivity. It can even result a reduction in the bandwidth of a system [5]. Moreover, some signals contain very high noise caused by human error, for example, a finger moving during the data retrieval process using a fingertip pulse wave. This condition requires several filters that can select the signal and select the one with the best conditions so that when it is processed with machine learning, it can produce better accuracy [6].

Another significant issue dealing with the photoplethysmography signal obtained using the fingertip pulse wave is the different noise levels in each collected data. Some data have a slight noise, but some others have moderate to high noise levels. To overcome this problem, there have been several studies performed to reduce noise in PPG signal data. Some popular methods to use are an Adaptive Scaled Fourier Linear Combiner (SLFC) to reduce noise in the PPG signal [7], which contains noise from user movement, Dynamic Monitoring [8], and Particle filter-based PPG signal estimation [9].

However, in reality, the level of noise in it is different. The selection of noise filters is generally made by observing the noise ratio at a minor level. The problem is how to optimize the existing data group by selecting the most optimal filter variable to produce a clean signal. Instead of giving a signal with a good signal-to-noise ratio, a mistake in variable determination to optimize noise filtering will provide data with a poor signal-to-noise ratio or change the information in the data [10].

Each data in a signal group has unique information or features. One of the features contained in it is the noise level in the data. The noise level provides information on the noise ratio in the data. These data can be grouped based on similarities in the noise ratio. Clustering is a method of grouping data into several clusters so that the data in a cluster have a maximum similarity [11]. By grouping, each group with different noise levels can be treated using different noise filtering according to the noise ratio level.

This study proposes a noise filter that weights according to the noise level in each data. K-means clustering is used to group each data based on similarities in the noise ratio level. This clustering process uses the coefficient of variation and slope to cluster the Photoplethysmography signal data set. The coefficient of variation is calculated using the standard deviation and sample mean for each data. The slope is calculated by comparing changes in the nth data with the previous data. Furthermore, each data group will be given a noise filter according to the noise ratio. The noise filter is an exponential filter that uses weighting to set the fineness to be applied to the data. The test is carried out by comparing the noise ratio in the data with three categories of groups. The first category is data without noise filtering, the second is data with noise filtering with constant weighting, and the third is data with noise filtering with a different weighting according to the clustering group.

II. Literature Review

The photoplethysmography signal data collection has generally been taken using the fingers and earlobes [12]. A fingertip is usually used on the finger, with an emitter and receiver in the device to pick up the photoplethysmography signal. An ear probe is usually used to process the PPG signal data on the auricle. There are several differences in the measurement results of the two measurement methods. The fingertip used on the finger has better measurement results when compared to the measurement results using an ear probe [13]. The use of a fingertip has an error of +0.71% when compared to the ABL90 Flex Analyzer, while the use of an ear probe has an error of +4.20% when compared to the ABL90 Flex Analyzer.

Several factors influence the process of taking photoplethysmography signal data. This factor is critical because it will determine the quality of the photoplethysmography signal. Internal influencing factors include clinical settings for pulse oximetry measurement, PPG configuration mode, and PPG sensing for other physiological measurements [14]. External factors also affect the process of taking photoplethysmography signal data. One of the main factors that affect the performance of PPG sensing is its vulnerability to interference, especially from user movements during data collection [15]. Other factors that also affect the results of the photoplethysmography signal are the amount of blood flowing into the peripheral blood vessels, different types of skin and blood, ambient light, and the wavelength of the emitter used [16].

Research related to photoplethysmography signals for the medical field has been increasing. The value of heart rate (HR), heart rate variability (HRV), respiratory rate (RR), blood oxygen saturation, and HbA1c level can be measured using the PPG signal [17]. The data acquisition of the signals needs special treatment. Besides, the data acquisition process of photoplethysmography signals uses non-invasive methods using a fingertip pulse wave, as shown in Fig. 2, which has many disturbances that need to resolve to obtain optimal signal results [18]. Several factors, as described above, can be resolved by taking precautions before the data acquisition process by clarifying inclusion and exclusion and reasonable standard operating procedures. However, several other factors, such as noise, must be resolved immediately before further data processing.

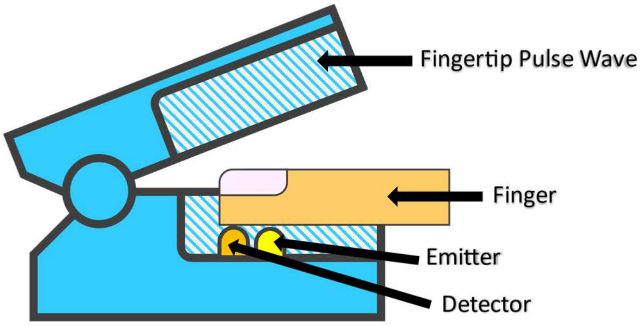


Fig. 2 Signal acquisition using fingertip pulse wave

The information in the photoplethysmography signal can be extracted from the features. This process requires the photoplethysmography signal to be in good condition. However, feature extraction on the photoplethysmography signal is disturbed due to the mixing of noise and clean photoplethysmography signal in a non-linear way, as shown in Fig. 3. This problem can be solved by robust de-noising of photoplethysmography signals.

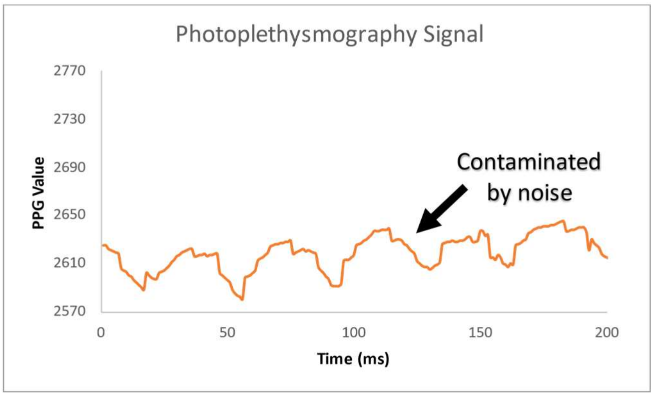


Fig. 3 Photoplethysmography signal contaminated by noise

Generally, de-noising is accomplished by applying the same weight to each data, even though the noise mixed in each signal has a different ratio. This difference is due to differences in user conditions during the data acquisition process. A grouping process is needed to separate each level of contaminating noise so that special handling can be done for each signal.

The process of data grouping is processed using the clustering method. Clustering methods are divided into two types, hierarchical clustering and partitional clustering. The main differences between the two clustering methods are time consumption, assumptions, input parameters, and the resulting clusters. The partitional clustering method is faster than the hierarchical clustering method. Hierarchical clustering only requires similarity measures, whereas partitional clustering requires more robust assumptions such as the number of clusters and initial center. Hierarchical clustering does not require input parameters, while partitional clustering requires the number of clusters to start running. Hierarchical clustering produces a much more meaningful and subjective division of clusters, but partitional clustering produces distinct k clusters.

Noise filtering on photoplethysmography signals has come a long way. This shows that this problem is still a concern for many researchers who use photoplethysmography signals. Some simple filters, such as moving averages and low pass filters, are used to process signal de-noising. More advanced filters are also used to overcome this problem.

III. Research Method

This study uses 677 Photoplethysmography signal data taken using Fingertip Pulse Wave. The data obtained is data from patients with an age range of 17 years to 70 years. This research includes three steps: k-means clustering, exponential filter, and evaluation using signal-to-noise ratio.

A. K-Means Clustering

The approach to be taken in this study is to use the clustering method, as shown in Fig. 4. The clustering method used is K-Means clustering. K-Means clustering is relatively simple, so it will be easier to implement. K-Means clustering offers convergence in the clustering process for small and large data sizes. K-Means clustering can adapt to the addition of data. The generalization of cluster forms offered is also very diverse, in terms of size and shape, for example, clusters in the form of an eclipse.

In the raw spectral photoplethysmography signal, three features are extracted, standard deviation, samples mean, and slope. The standard deviation or standard deviation is the distribution of data in a sample to see how far or how close the data values are to the average. As shown in Equation (1), The standard deviation is the square root value of the variance and indicates the standard deviation of the data about its mean value. If the standard deviation value is small, it is getting closer to the average. However, the data variation is more comprehensive if the standard deviation value is significant.

$$S = \frac{\sqrt{\sum_{i=1}^{n} (x_i-\bar{x})^2}}{n-1} \quad (1)$$

where
S : Standard deviation
x_i : each of the values of the data
x_i : mean of x_i
n : number of samples

as shown in Equation (2), the mean value can be determined by dividing the amount of data by the amount of data. The mean is a measure of the center of the data. The mean is a representative data set or average value that is considered a value that is closest to the actual measurement results.

$$\bar{X} = \frac{\sum x}{n} \quad (2)$$

where
$\bar{X}$: sample mean
$\sum x$: the sum of all samples in one data
n : number of samples

these two variables are used to calculate the coefficient of variation, as shown in Equation 3, which combines the slope for the clustering process.

$$CV = \frac{S}{\bar{X}} \quad (3)$$

where
CV : coefficient of variation
S : Standard deviation
$\bar{X}$: sample mean

There are five steps taken in the k-means clustering process. The first step is determining the number of clusters by the user. The second step is to determine the centroid value. The third step is determining the distance between the centroid and object points. The fourth step is grouping objects based on centroids. The last step is repetition to the second stage until the resulting constant centroid value, and the cluster members do not move to another cluster.

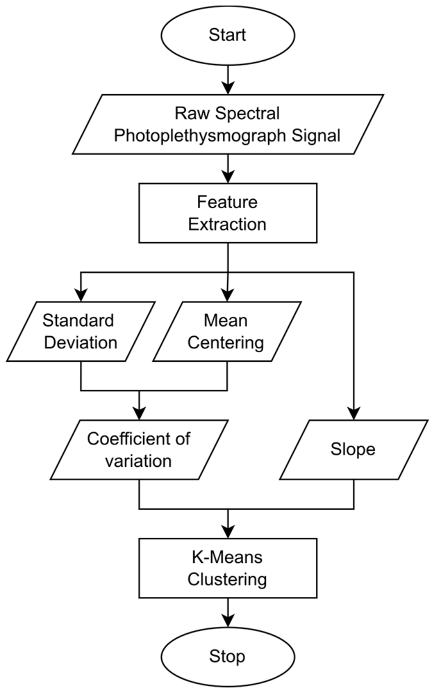
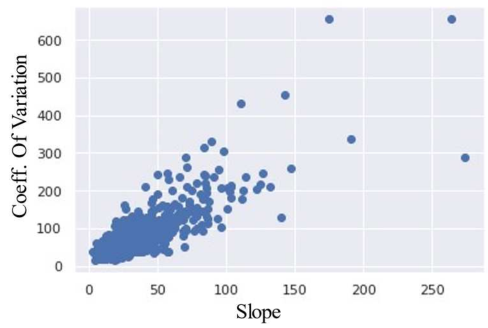
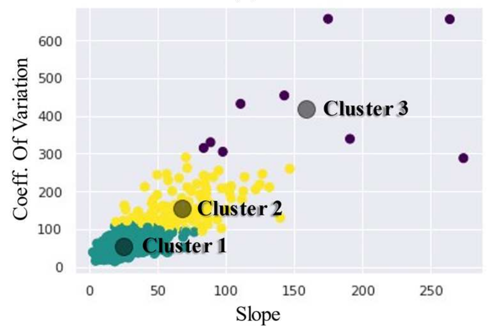


Fig. 4 Process of k-means clustering on photoplethysmography signals

In this step, three centroids are used to cluster the data. Fig. 5 (a) shows the results of plotting data based on the Coefficient of Variation and Slope for each data sample. The graph shows that many of the data is in good condition because it has a small Coefficient of variation and Slope. However, there are some scattered data with a more significant Coefficient of Variation and Slope compared to other data samples.



(a)



(b)

Fig. 5 (a) data distribution (b) clustering result

The clustering process will produce three groups of data as shown in Fig. 5 (b). The first group is data samples with a small Coefficient of Variation and Slope, the second group is data samples with a moderate Coefficient of Variation and Slope, and the third group is data samples with a significant Coefficient of Variation and Slope. These three data groups will use different treatments to reduce noise so that each group will have its filter weight.

B. Exponential Filter

The clustered data will be processed using noise filtering to remove noise that contaminates the raw photoplethysmograph signal using exponential filters. Exponential filters is a moving forecasting method that gives exponential or multilevel weight to the latest data so that the latest data will get greater weight. In other words, the newer or more recent the data, the greater the weight.

$$F_{t+1} = wA_t + (1 - w)\, F_t \tag{4}$$

where

F_{t+1} : new output value
F_t : previously predicted value
A_t : actual value
w : filter weight

Exponential filters work using weights that will set the level of filtering on noise. Equation 4 shows how this filter works to predict the new output (F_{t+1}) by using the previously predicted value (F_t) multiplied by the weight (w) and the difference between the actual value (A_t) and the previous predicted result (F_t).

C. Signal to Noise Ratio

In analog and digital data communications, the signal-to-noise ratio, often written as S/N or SNR, is a unit of the ratio of desired signal strength relative to background noise (unwanted signal). SNR is expressed as a single numeric value.

IV. Experiment Result

Table I shows the results of the process of clustering the photoplethysmography signal data into three groups. From a total of 677 sample data, there are 545 sample data (80.50%) with small noise, 123 sample data (18.16%) with moderate noise, and 9 sample data (1.34%) with significant noise. Most of the sample data was included in the group with small noise, and a small portion of the sample data with a larger noise.

TABLE I. Clustering Result

Clusters number	*Noise level*	*Number of samples*	*Percentage (%)*
1	Small noise	545	80.50
2	Moderate noise	123	18.16
3	Significant noise	9	1.34

The higher the weighting value will give a low level of smoothness, while the lower one will give a higher level of smoothness accompanied by lagging. Table II shows the weighting value for each cluster. Weighting is given based on the noise level in each cluster.

TABLE II. Exponential Filter Weighting

Clusters number	*Noise level*	*Weighting (%)*
1	Small noise	70
2	Moderate noise	50
3	Significant noise	20

For small noise, a weighting of 70% is given to the exponential filter used. The Fig. 6 shows that the output has almost the same characteristics as the actual data sample.

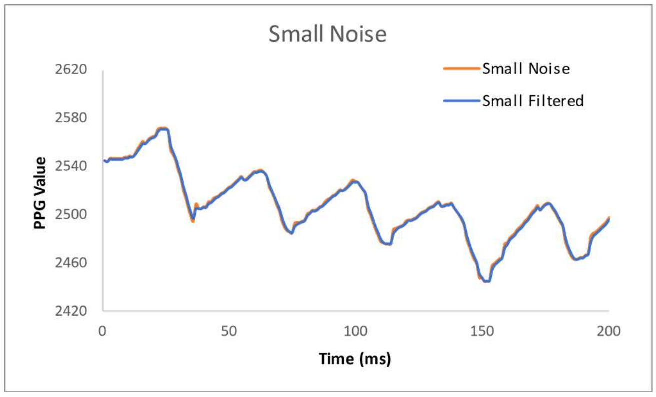


Fig. 6 Filter result on small noise data

Data in the moderate noise category are given a weighting of 50%, meaning that the resulting output is partly the actual data and the previous output. It can be seen that the noise is reduced so that the data becomes smoother, as shown in Fig. 7.

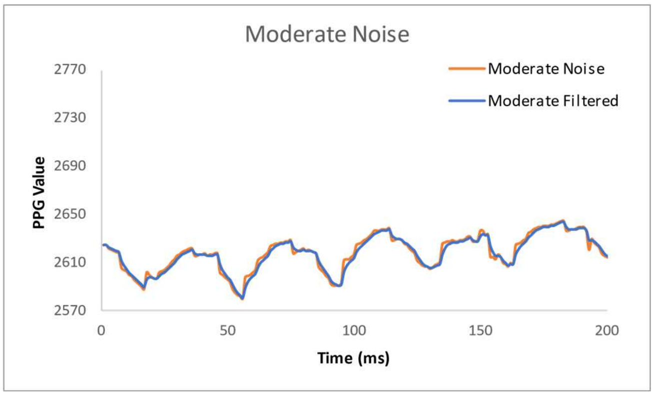


Fig. 7 Filter result on moderate noise data

Data in the third category totalled 9, with very high spikes that occurred suddenly. The most common causative factor is that the subject gives movement to his finger during the data acquisition process. This category is given a weighting of 20%, which makes the filtered data results finer, as shown in Fig. 8.

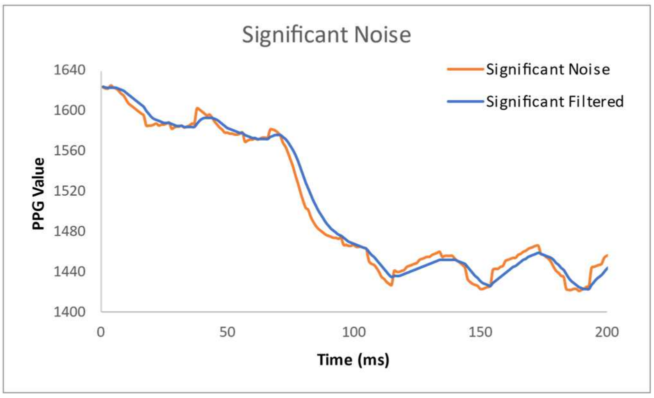


Fig. 8 Filter result on significant noise data

The next stage involves a comparison and evaluation process using the signal-to-noise ratio. The signal-to-noise ratio shows how dominant the original signal is to the noise contaminating it.

Comparisons between several categories of data were chosen as a test process. Several categories are used, signal-to-noise ratio on data that has not been given a filter, signal-to-noise ratio on data that is given a constant weighting of 70%, and data that is given filter weight according to each cluster. From the experiments result on Table III, it is known that non-filtered data has a signal-to-noise ratio of 181.49, and data with a constant weighted filter has an increase in the signal-to-noise ratio of 2.30, which indicates that there has been an improvement in the data.

TABLE III. EXPONENTIAL FILTER RESULT

Category	***SNR***
Non-filtered Data	181.49
Constant Weighted Filter	183.79
Clustered and Weighted Filter	187.48

The usage of a clustered and weighted filter using k-means clustering and an exponential filter produces a signal-to-noise ratio of 187.48. This indicates that there is a better improvement when compared to using a constant weighted filter.

V. CONCLUSION

This paper shows the advantages of using clustered and weighted filter using k-means and exponential filters to reduce noise that contaminates photoplethysmography signals. This paper describes the results of k-means clustering, the simulation results of using exponential filters, and the results of comparisons of targeted filters compared to the use of constant weighted filters. We evaluate the experiment by comparing the signal-to-noise ratio in each filter category. We find an increase in the signal-to-noise ratio with the proposed research method. This shows that the use of targeted filters provides data with minimal noise. Optimimalization is needed to improve performance in selecting the number of clusters and the weighting on the exponential filter. Future work will involve further preprocessing to evaluate data with significant noise categories so that the resulting data has a steady character or conditions that are better for processing.

ACKNOWLEDGMENT

This research was funded by Lembaga Pengelola Dana Pendidikan (LPDP) under Riset Inovatif-Produktif (RISPRO) Program; the Indonesian Ministry of Education and Culture under Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) Program; and Institut Teknologi Sepuluh Nopember (ITS) under project scheme of the Publication Writing and IPR Incentive Program (PPHKI).

REFERENCES

[1] M. Selvam, S. Johnson, J. Mikael, and E. Smieee, *A Review of Photoplethysmography-based Physiological Measurement and Estimation, Part 1: Single Input Methods**. 2020. doi: 10.0/Linux-x86_64.

[2] Z. Tian and Y. Jia, “Application of Photoplethysmographic Pulse Signal for Human Physiological Information Acquisition,” in *2021 IEEE 3rd International Conference on Frontiers Technology of Information and Computer (ICFTIC)*, Nov. 2021, pp. 382–385. doi: 10.1109/ICFTIC54370.2021.9647271.

[3] Y. Liang, M. Elgendi, Z. Chen, and R. Ward, “An optimal filter for short photoplethysmogram signals,” *Sci Data*, vol. 5, no. 1, p. 180076, May 2018, doi: 10.1038/sdata.2018.76.

[4] D. P. Purbawa *et al.*, “Adaptive filter for detection outlier data on electronic nose signal,” *Sens Biosensing Res*, vol. 36, p. 100492, Jun. 2022, doi: 10.1016/j.sbsr.2022.100492.

[5] A. John, J. Sadasivan, and C. S. Seelamantula, “Adaptive Savitzky-Golay Filtering in Non-Gaussian Noise,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 5021–5036, 2021, doi: 10.1109/TSP.2021.3106450.

[6] M. S. H. Ardani *et al.*, “A new approach to signal filtering method using K-means clustering and distance-based Kalman filtering,” *Sens Biosensing Res*, vol. 38, p. 100539, Dec. 2022, doi: 10.1016/j.sbsr.2022.100539.

[7] S. C. Kim, E. J. Hwang, and D. W. Kim, “Noise reduction of PPG signal during Free Movements Using Adaptive SFLC (scaled Fourier linear

combiner)," in *World Congress on Medical Physics and Biomedical Engineering 2006*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1191–1194. doi: 10.1007/978-3-540-36841-0_287.

[8] Y. Chen, D. Li, Y. Li, X. Ma, and J. Wei, "Use Moving Average Filter to Reduce Noises in Wearable PPG During Continuous Monitoring," 2017, pp. 193–203. doi: 10.1007/978-3-319-49655-9_26.

[9] Y.-K. Lee, O.-W. Kwon, H. S. Shin, J. Jo, and Y. Lee, "Noise reduction of PPG signals using a particle filter for robust emotion recognition," in *2011 IEEE International Conference on Consumer Electronics - Berlin (ICCE-Berlin)*, Sep. 2011, pp. 202–205. doi: 10.1109/ICCE-Berlin.2011.6031807.

[10] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, "A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots," *Applied Sciences*, vol. 11, no. 23, p. 11202, Nov. 2021, doi: 10.3390/app112311202.

[11] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," *IEEE Access*, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[12] Y. Yamakoshi *et al.*, "Side-scattered finger-photoplethysmography: experimental investigations toward practical noninvasive measurement of blood glucose," *J Biomed Opt*, vol. 22, no. 6, p. 067001, Jun. 2017, doi: 10.1117/1.JBO.22.6.067001.

[13] S. Olive, O. Twentyman, and C. Ramsay, "Comparison of fingertip and earlobe pulse oximetry with arterial blood gas results," in *1.13 Clinical Problems - Other*, Sep. 2016, p. PA3702. doi: 10.1183/13993003.congress-2016.PA3702.

[14] D. Ray, T. Collins, S. Woolley, and P. Ponnapalli, "A Review of Wearable Multi-wavelength Photoplethysmography," *IEEE Rev Biomed Eng*, pp. 1–1, 2021, doi: 10.1109/RBME.2021.3121476.

[15] C. Bruser, C. H. Antink, T. Wartzek, M. Walter, and S. Leonhardt, "Ambient and Unobtrusive Cardiorespiratory Monitoring Techniques," *IEEE Rev Biomed Eng*, vol. 8, pp. 30–43, 2015, doi: 10.1109/RBME.2015.2414661.

[16] D. Biswas, N. Simoes-Capela, C. van Hoof, and N. van Helleputte, "Heart Rate Estimation From Wrist-Worn Photoplethysmography: A Review," *IEEE Sens J*, vol. 19, no. 16, pp. 6560–6570, Aug. 2019, doi: 10.1109/JSEN.2019.2914166.

[17] S. Usman, N. Harun, R. A. Dziyauddin, and N. A. Bani, "Estimation of HbA1c level among diabetic patients using second derivative of Photoplethysmography," in *2017 IEEE 15th Student Conference on Research and Development (SCOReD)*, Dec. 2017, pp. 89–92. doi: 10.1109/SCORED.2017.8305415.

[18] A. E. Awodeyi, S. R. Alty, and M. Ghavami, "On the Filtering of Photoplethysmography Signals," in *2014 IEEE International Conference on Bioinformatics and Bioengineering*, Nov. 2014, pp. 175–178. doi: 10.1109/BIBE.2014.76.