JOIV : Int. J. Inform. Visualization, 6(1) - March 2022 90-95

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Tree-based Filtering in Pulse-Line Intersection Method Outputs for An Outlier-tolerant Data Processing

Cahya Damarjati [a,*], Karisma Trinanda Putra [b], Heri Wijayanto [c], Hsing-Chung Chen [d,e], Toha Ardi Nugraha [b]

[a] *Department of Information Technology, Universitas Muhammadiyah Yogyakarta, Yogyakarta 55183, Indonesia*
[b] *Department of Electrical Engineering, Universitas Muhammadiyah Yogyakarta, Yogyakarta 55183, Indonesia*
[c] *Department of Information Engineering, Universitas Mataram, Mataram 83115, Indonesia*
[d] *Department of Computer Science and Information Engineering, Asia University, Taichung 413, Taiwan*
[e] *Department of Medical Research, China Medical University Hospital, China Medical University Taichung 404, Taiwan*

*Corresponding authors: *cahya.damarjati@umy.ac.id*

***Abstract*— Pulse palpation is one of the non-invasive patient observations that identify patient conditions based on the shape of the human pulse. The observations have been practiced by Traditional Chinese Medicine (TCM) practitioners since thousands of years ago. The practitioners measure the patient's arterial pulses in three points of both patient wrists called *chun*, *guan*, and *chy*, then diagnose based on their knowledge and experience. Pulse-Line Intersection (PLI) method extract features of each pulse from the observed pulse wave sequence. PLI is performed by summing the number of intersections between the artificial line and the pulse wave. The method is proven in differentiating between hesitant with moderate pulse waves. As the method implemented in Clinical Decision Support System (CDSS) related to pulse palpation, some outlier data might emerge and affect the measurement result. Thus, outlier filtering is needed to prevent unnecessary prediction processes by machine learning (ML) models inside CDSS. This study proposed an outlier filtering model using a decision tree algorithm. This concept is designed by analyzing pulse features values and the chance of odd values combination. Then inappropriate values are excepted using several rules. Every pulse feature list that did not pass the filtering rule is categorized as outliers and were not included for further process. The proposed model works more efficiently than ML models dealing with outliers since this procedure is unsupervised learning with a small number of parameters. Overall, the proposed filtering method can be used in pulse measurement applications by eliminating outlier data that might decrease the performance of ML models.**

***Keywords*—Pulse palpation; outlier filtering; decision tree; CDSS.**

Manuscript received 8 Jan. 2022; revised 15 Feb. 2022; accepted 2 Mar. 2022. Date of publication 31 Mar. 2022.

I. Introduction

Non-invasive clinical diagnosis often entails the use of sophisticated diagnostic equipment which generates pictures of the organs, soft tissues, and bones within the patient without the need for a wound. One of the non-invasive clinical diagnoses is pulse palpation which results in signal data. It is a method of obtaining a preliminary diagnostic of a patient's health status by measuring the wrist artery pulse with the physician's finger [1]. Modern, sophisticated scientific technology and effective signal analysis techniques may be coupled to accomplish scientific, understandable, and precise extraction of the pulse signal's unique properties [2]. Furthermore, TCM contains rich knowledge about pulse waves and their relationship with the disease a patient suffered [3]. Conceptually, a pulse wave cycle consists of systolic and diastolic phases. A tiny and temporary rise in the diastolic phase usually occurs called dicrotic notch [4], as depicted in Fig. 1. The reality of measured pulse waves is more complex than Fig. 1 is illustrated. TCM physicians have classified pulse waves into 26 categories [5]. However, recent research focused on fewer categories, e.g., slippery, taut, moderate, hesitant, wiry, soft, smooth, and unsmooth [6], [7]. Some data, as shown in Fig. 2 contains the mentioned pulse types. Furthermore, some research suggests a correlation between pulse wave shapes with diseases, e.g., hypertension [8], cancers [9], and diabetes [10].

A study by Chen et al. [11] proposed CDSS for pulse palpation for differentiating between hesitant and moderate pulse. The system has PLI method component for pulses wave features extraction. Using SHapley Additive exPlanations (SHAP) for feature importance analysis, all features manifested varied nonzero results based on test data and applied machine learning models. Thus, there is no need to remove features from PLI results because all the features are usable for further process.

Although with a good data processing technique from data extraction using data acquisition device to displaying the information at the end of the CDSS procedures, the existence of the outlier data points (*i.e.,* unwanted information, indicates the not pulse data) might occur [12]. Due to improper wear transducer, broken equipment, wiring, or noises. Outlier filtering using a decision tree approach is proposed in this study to solve the issue above. The use of viral machine learning is not suggested because the classification type of machine learning models could not solve a related issue for two reasons. The first is that the outlier data must be generated similarly to pulse data. When the generated outlier is only a few, the ratio between two classes becomes imbalanced [13], making the machine learning model training unreliable [14]. Secondly, generating varied outliers is hard because many aspects need to be considered, e.g., noises simulation and simulating data in broken equipment. Removing outlier class makes the model training in vain since the models need a challenge in training validation step and improve their learning performance. Some machine learning presenting outlier classification, i.e., One-class SVM (OCSVM) [15] and K-Means [16], were used as performance analysis and comparison to assert the better performance of the proposed approach.

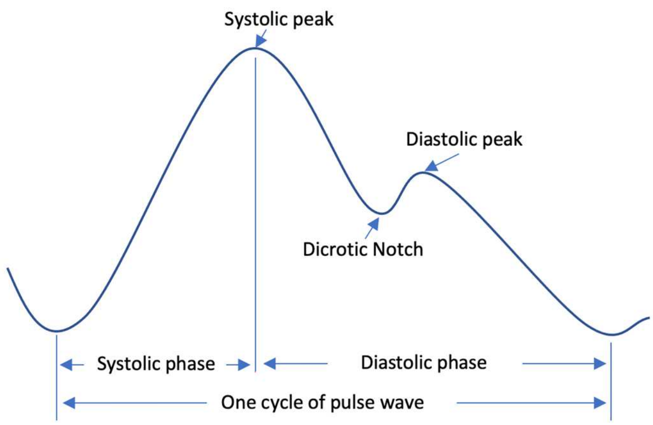

Fig. 1 Pulse wave component

The remainder of this paper is organized as follows. Pulse palpation CDSS is presented in the Methodology section, followed by a detailed description of each component. The proposed outlier filtering approach is also described in the Materials and Method section. Then, the experiment results and their performances are described in the Results and Discussion section. Finally, the conclusion of this study is written in the Conclusion section.

II. Materials and Method

Pulse palpation CDSS was proposed by Chen et al. [11] in accommodating hesitant pulse detection as specified illness patient's early diagnosis with some adjustments in proper for current study needs. In brief, there are four sub-sections presented, *i.e.,* Dataset analysis, CDSS architecture, PLI feature extraction, and proposed outlier filtering. For further reading, the rest of the contents of this section is well-written as follows.

A. Dataset Analysis and Pre-processing

The experiment in this study used pulse data in image format gathered in Chinese Medical University Hospital and reviewed by its Institutional Review Board under CMUH107-REC2-050 and CMUH109-REC2-159 records [11]. A total of 46 patients were involved in the process. One patient data contains six pulses taken from six different points, *i.e., chun*, *guan*, and *chy* [17] for both left wrist and right wrist. The measurement of each point occurs for six seconds as recommended by TCM Physicians. The experiment is under the time that is depicted in Fig. 2. The raw dataset contains 5 - 9 pulse wave cycles in one measurement according to each patient condition that additional process to split and extract the feature is needed. Some data contain only 3 pulses because the measurement was conducted for one patient's wrist.

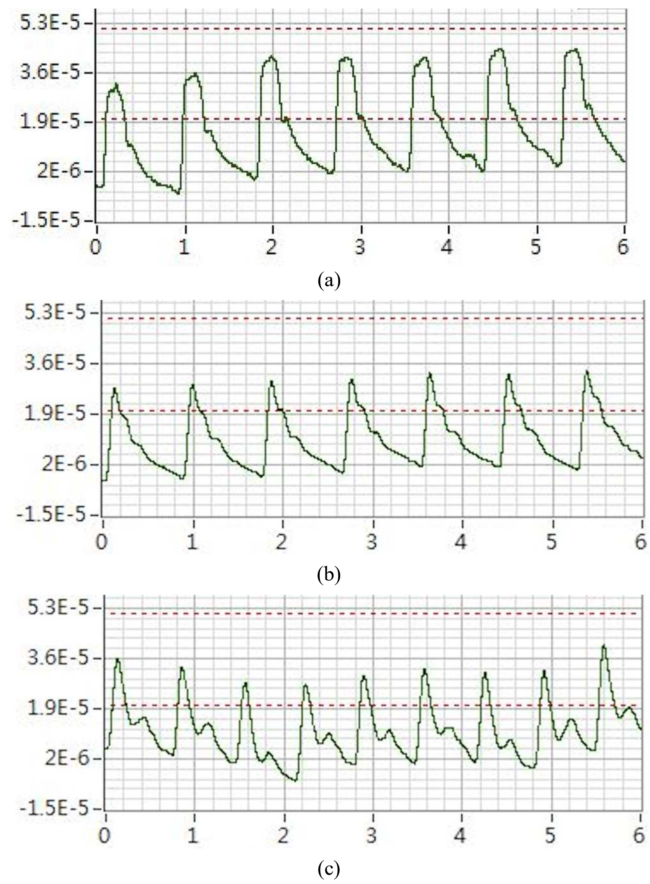

Fig. 2 Variation of pulses that is gathered from different patients. (a) Hesitant, (b) Slippery, (c) moderate. Vertical numbers represent the amplitude of the pulse, while horizontal numbers represent the measurement time in seconds.

Since the dataset for this study was in image format, some processes must be conducted to convert it into time-series format. It is beneficial for further application since the measured pulse wave is in a time-series format. The process consists of three steps, i.e., image extraction, image binarization, and pulse generation. Image extraction is gathering the specific measurement area from raw pulse images. The phase begins by specifying the six top-left measurement area coordinates g, as well as the height h and width w, because the raw image content is fixed in size and location. Then it proceeds to cut the raw image segment with the predetermined g and save it to new pictures I' as specified in Eq. (1), where g_x signifies the X axis and g_y signifies Y axis.

$$I' = M_{(g_x,g_y)}: M_{(g_x+w,g_y+h)} \tag{1}$$

The image binarization step requires a threshold t color value to convert all grids to black and the pulse wave I'' to blue, as stated in Eq. (2).

$$I''_{(g_x,g_y)} = \begin{cases} 1, & I'_{(g_x,g_y)} \geq t \\ 0, & I'_{(g_x,g_y)} < t \end{cases} \tag{2}$$

The pulse generation step is converting binarized images into serialized data. The time attribute that is used here is based on the image width pixel value. The data D is the y coordinate of the pulse wave I'' in every x value, as stated in Eq. (3). It works by multiplying the value with h as the location of y coordinate. The h would be the D_x value since every h is multiplied by either 0 or 1 of I''.

$$D_x = \sum_{i=1}^{h} I''_{(g_x,g_i)} * h \tag{3}$$

B. CDSS Architecture

Over the last decades, the advanced evolution of data mining and machine learning techniques has accelerated the revolution of human healthcare by providing more accurate predictions, which can convert medical record data into reliable information and diagnostics. This information is well valuable for medical practitioners improving decision support in the healthcare environment. A new system in the healthcare environment should be defined to provide a faster and much easy way to explain the diagnosis results. During patient assessments, similar symptoms or diagnoses may have occurred. This system (*i.e.,* CDSS) can assist clinicians in diagnosing patient diseases with similar symptoms, making appropriate treatments can be accurately defined. The CDSS with machine learning (ML) has offered many benefits over the conventional healthcare systems. However, it still lacks understanding capability and managing the information, particularly during the decision phase. This study improves CDSS architecture [11] by involving an additional filtering technique during the data collection phase.

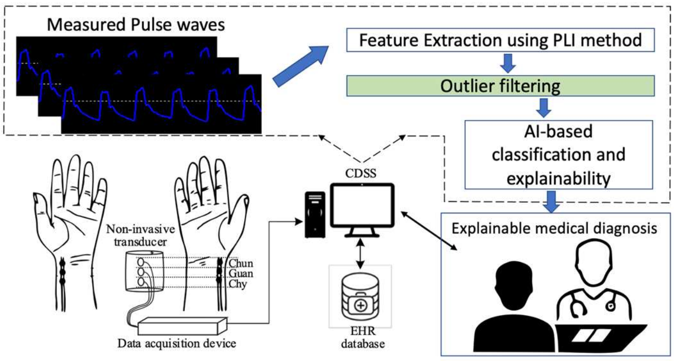

Fig. 3 The proposed CDSS architecture includes outlier filtering to improve the robustness of the prediction model by eliminating failure data points during patient observations.

As depicted in Fig. 3, data were extracted from the patients during the observation phase. The data is in the form of a patient's pulse signal. Furthermore, the PLI method is used to represent these biomedical signals into data subsets based on their wave cycles. Visually, these signals are difficult to distinguish even though signal processing techniques such as the Fourier transform are applied. PLI can extract biomedical signals based on the representation defined by the TCM practitioner. However, sometimes these signals have similar patterns even though they belong to different classes. This data subset includes outliers that can degrade the prediction results. Therefore, a filtering technique is offered to eliminate signals with high ambiguity in this phase.

Furthermore, the results are predicted and explained by AI models combined with the SHAP technique. Finally, clinicians can use medical records and prediction results to help improve the accuracy of decision-making and patient care. Furthermore, these data are stored in an electronic health record (EHR) to increase the reliability of future healthcare.

C. Feature Extraction using PLI

As depicted in Fig. 4, PLI initiates by creating line L with specific angle a with peak point p_k to the end of signal x-axis *i.e.,* g_x of v_k and v_{k+1}. The lines L denotes angle $i \in \{0°, 15°, 30°, 45°, 60°, 75°, 105°, 120°, 135°, 150°, 165°\}$. The subset data generation *i.e.,* L_i^k is calculated in Eq. (4) where b is y-axis of p_k.

$$L_i^k = \sin(-a)\,(x - g_x) + b \tag{4}$$

For every L_i^k, each intersection t with I_k is a PLI feature x_i^k as denoted in Eq. (5).

$$x_i^k = \sum gxt \begin{cases} 1: \left(y - axis\ of L_i^k(g_x)\right) = \left(y - axis\ of I_k(g_x)\right) \\ 0: otherwise \end{cases} \tag{5}$$

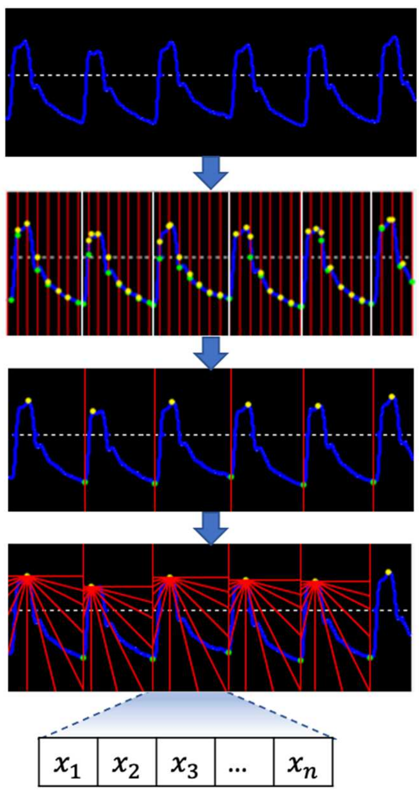

Fig. 4 PLI feature extraction steps resulting PLI's features.

The baseline characteristics of features, *i.e.,* extracted from each subset data, are stated in Fig. 5 and Table 1. Based on patient observations using TCM methodology, 369 pulses were identified as the hesitant type, 683 pulses were in the moderate type, and the rest were not included in both categories. In Table 1, data is in mean ± standard deviation form. Although each feature's mean and standard deviation is small, it has slightly different values for both types, which indicates that each class has unique characteristics. Higher values are occurred for hesitant pulse than moderate pulse for x = {*Int_0, Int_15, Int_30, Int_45, Int_105, Int_120, Int_135, Int_150, Int_165*} with (P<0.001). This might be the characteristic of TCM scholars for a hesitant pulse, *i.e.,* sharp and choppy. On the other hand, for moderate pulse, higher measurement values only in x = {*Int_60, Int_75*} with (P<0.001). This corresponds to the nature of moderate pulse, which commonly has skewness around 60° to 75°. PLI's line commonly intersects with the observed signal between this intersection angle.

D. Outlier Filtering

This section describes our proposed and suggested outlier filtering method for the applied pulse palpation CDSS. The approach used a simple decision tree with specific filter conditions. This concept also has been recently applied for detecting outliers in sequential data streams [18]. Furthermore, it carried an explainability feature in the process [19]. This study used the decision tree concept because the data is unique, and common machine learning for detection outlier could not achieve their maximum performance.

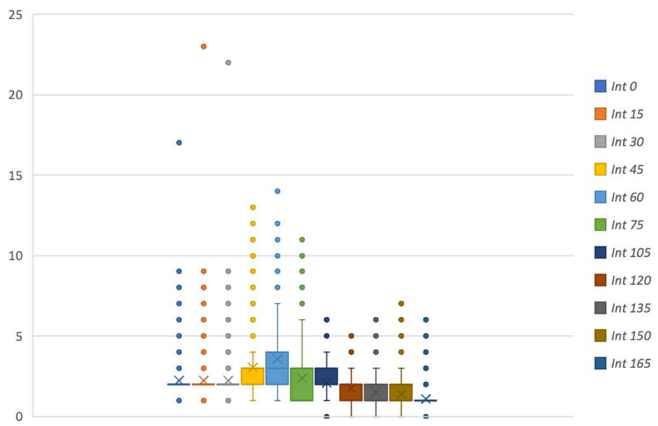

Fig. 5 PLI's results features distribution

TABLE I
PLI'S FEATURES DISTRIBUTION BETWEEN HESITANT AND MODERATE TYPE PULSES

Features (x)	**Hesitant (n=369)**	**Moderate (n=683)**	**P-Value**
Int_0	2.66 ± 1.44	1.99 ± 1.00	< 0.001
Int_15	2.63 ± 1.44	2.00 ± 1.15	< 0.001
Int_30	2.63 ± 1.44	2.00 ± 1.12	< 0.001
Int_45	3.65 ± 1.97	2.69 ± 1.42	< 0.001
Int_60	3.30 ± 2.16	3.71 ± 1.67	< 0.001
Int_75	1.42 ± 1.09	2.85 ± 1.94	< 0.001
Int_105	2.24 ± 0.88	2.03 ± 0.89	< 0.001
Int_120	1.89 ± 0.85	1.66 ± 0.76	< 0.001
Int_135	1.73 ± 1.05	1.37 ± 0.60	< 0.001
Int_150	1.50 ± 0.92	1.27 ± 0.59	< 0.001
Int_165	1.14 ± 0.60	1.03 ± 0.34	< 0.001

ALGORITHM 1. OUTLIER FILTERING FOR A SERIES OF x FEATURE

```
Input: array x
Output: bool isOutlier
isOutlier = False
angles_1={0,15,30,45,60,75}
if x_angles_1 contains 0 then
    isOutlier = True
end
if count(3 < x ≤ 10) > 5 then
    isOutlier = True
end
if count(x > 10) > 2 then
    isOutlier = True
end
return isOutlier
```

As formulated in Algorithm 1, three conditions, *i.e.,* lower, middle, and upper bound filtering, are defined. The lower bound filter is used to prevent outlier that has features with a small number value. PLI's line always crosses the diastolic phase of a pulse wave since the phase duration is longer than the systolic phase. In the systolic phase, some waves might not intersect with PLI's line because of their short duration. Thus, the filter condition is *Int_0, Int_ 15, Int_30, Int_45, Int_60,* and *Int_75* must not return 0 value since its crosses the diastolic phase of the wave.

Middle bound filtering is defined as preventing features with common values but is considered an outlier. PLI's features common values range from 1 to 3 and rarely four or higher because of the same direction between PLI's line and part of pulse wave. The rare condition is not happening for every feature. Thus, the condition for this filter is that features could have a value ranging from 3 to 10 but is limited to 5 features. For upper bound filtering, the condition is that features could have a value higher than 10 but are limited to 2 features. The reason is the same direction between pulse and PLI's line condition.

III. RESULTS AND DISCUSSION

In this section, the performance of proposed approach and applied machine learning *e.g.,* OCSVM and K-Means are presented. OCSVM is Support Vector Machine (SVM) variance that optimizing accuracy for one label/class data. It could be used for outlier filtering since SVM's hyperplane (ρ) contributed to distinguished data's class [16] based on its tuned threshold. It used convex optimization formula in Eq. (6) where ξ acts as training's error, and n as the amount of data [20].

$$\min_{\omega,\xi,\rho} \frac{\|\omega\|^2}{2} - \rho + \frac{1}{vn}\sum_{i=1}^{n} \xi_i \quad (6)$$

OCSVM has been used in many studies in detecting outliers. The authors [21] used OCSVM and Particle Swarm Optimization (PSO) to classify anomaly data in a power system. The purpose was to ensure that the system's stable and safe condition was constantly maintained. In other areas, the authors in [22] used OCSVM to detect any anomaly activity in the industrial control system. The authors used Tennessee Eastman Process for the simulation. It could perform high accuracy in detection besides the data stream type dataset. However, the dataset nature in this study is different. Sometimes, outliers are data with normal values but do not shape pulse data. Therefore, in this experiment, OCSVM accuracy is 80.13% when predicting outliers. It was expected since OCSVM is the type of unsupervised machine learning.

K-Means is also an unsupervised machine learning type [23], [24]. It is used for clustering data by automatically positioning centroids and gathering nearby points as a cluster member [25], [26]. It could judge a data point as an outlier as the position is the farthest from centroid [27], [28]. K-Means make clusters by taking two steps, i.e., assignment and update [16]. The assignment step is assigning nearby datapoints x_p as the member of a cluster S by selecting the minimum range between a point with all centroids m. Eq. (7) formulated the step where i and j represents the index of cluster and k represents the clusters total number. Update step is applied for repositioning all centroids based on current cluster members as formulated in Eq. (8) where t represents routine numbers until all centroids stop moving.

$$S_i^{(t)} = \left\{x_p : \left\|x_p - m_i^{(t)}\right\|^2 \le \left\|x_p - m_j^{(t)}\right\|^2 \forall j, 1 \le j \le k\right\} \quad (7)$$

$$m_i^{(t+1)} = \frac{1}{\left|S_i^{(t)}\right|}\sum_{x_j \in S_i^{(t)}} x_j \quad (8)$$

Even though K-Means has been used broadly in detecting outliers [29]–[31], the outlier is defined based on the distance between centroid-data and its defined threshold percentage value. If the percentage value is 95%, the 5% farthest data would be counted as outliers. It is ineffective since the threshold is not well stated, and an outlier case with a normal value is not detected. Furthermore, our middle-bound filter condition could solve the issue. Thus, our approach reaches 100% accuracy currently because the filter conditions are defined after analyzing current pulse data.

This study provides a new perspective for outlier filtering and improvements to health monitoring data, *i.e.,* biomedical signals. However, this research evaluates a small number of patients' data due to the lack of participants and the complex procedures of collecting medical data records. In future work, the proposed method will be enhanced and optimized to detect outliers using more varied data silos through a federated learning scheme [32]. This scheme offers a greater number of valuable data from several medical institutions that can collaborate to improve the prediction system reliability.

IV. Conclusion

This article presents the use of filters in the generalization of machine learning for classification tasks in TCM. The filter acts as a method for discarding the measurement identified as outlier data points. The original classification models, proposed in [11], did not consider a single filtering technique for the pulse palpation observations. Therefore, the entire data observations are included, although a single data point is faulty due to measurement errors. As a result, the performance of the generalization models is insufficient. The present contribution improves the PLI method to have an outlier rejection filter suitable for pulse palpation observations. Practical solutions are provided to simultaneously avoid outlier data points that degrade the model accuracy and improve system effectivity by guaranteeing parameters convergency during training procedures. The proposed method is compared with the current procedures without additional filters. The results show the benefits of the filter compared to the naive approaches for all ML models. These advances could be useful to improve the development of CDSS, aiding the diagnostics procedures made by medical practitioners more reliable.

Acknowledgment

This study was funded by Asia University, Taiwan, and China Medical University Hospital, China Medical University, Taiwan, under Grant No. ASIA-110-CMUH-22, ASIA-108-CMUH-05, ASIA-106-CMUH-04 and ASIA-105-CMUH-04. This study was supported by the Ministry of Science and Technology (MOST), Taiwan, under MOST Grant No. 110-2218-E-468-001-MBK. This study was also supported by MOST under Grant No. 110-2221-E-468-007 and Grant No. 110-2218-E-002-044. Universitas Muhammadiyah Yogyakarta, Indonesia, also funded this study.

References

[1] H.-C. Chen, S.-S. Kuo, S.-C. Sun, and C.-H. Chang, “A Distinguishing Arterial Pulse Waves Approach by Using Image Processing and Feature Extraction Technique,” *J Med Syst*, vol. 40, no. 10, p. 215, Oct. 2016, doi: 10.1007/s10916-016-0568-4.

[2] Z. Chen, A. Huang, and X. Qiang, “Research on Pulse Classification Based on Multiple Factors,” in *Bio-inspired Computing: Theories and Applications*, vol. 952, J. Qiao, X. Zhao, L. Pan, X. Zuo, X. Zhang, Q. Zhang, and S. Huang, Eds. Singapore: Springer Singapore, 2018, pp. 63–71. doi: 10.1007/978-981-13-2829-9_7.

[3] A. de Sá Ferreira and A. J. Lopes, “Pulse waveform analysis as a bridge between pulse examination in Chinese medicine and cardiology,” *Chin. J. Integr. Med.*, vol. 19, no. 4, pp. 307–314, Apr. 2013, doi: 10.1007/s11655-013-1412-z.

[4] M. T. Politi *et al.*, “The dicrotic notch analyzed by a numerical model,” *Computers in Biology and Medicine*, vol. 72, pp. 54–64, May 2016, doi: 10.1016/j.compbiomed.2016.03.005.

[5] H. Chang, J. Chen, and Y. Liu, “Micro-piezoelectric pulse diagnoser and frequency domain analysis of human pulse signals,” *Journal of Traditional Chinese Medical Sciences*, vol. 5, no. 1, pp. 35–42, Jan. 2018, doi: 10.1016/j.jtcms.2018.02.002.

[6] L. S. Xu, M. Q.-H. Meng, and K. Q. Wang, “Pulse image recognition using fuzzy neural network,” in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 3148–3151.

[7] B. Peng, C.-H. Luo, N. Sinha, C.-C. Tai, X. Xie, and H. Xie, “Fourier series analysis for novel spatiotemporal pulse waves: Normal, taut, and slippery pulse images,” *Evidence-Based Complementary and Alternative Medicine*, vol. 2019, 2019.

[8] L. Qiao *et al.*, “The Association of Radial Artery Pulse Wave Variables with the Pulse Wave Velocity and Echocardiographic Parameters in Hypertension,” *Evidence-Based Complementary and Alternative Medicine*, vol. 2018, pp. 1–11, Dec. 2018, doi: 10.1155/2018/5291759.

[9] Y. Shi *et al.*, "Study on the Syndrome Characteristics and Classification Model of Non-Small Cell Lung Cancer Based on Tongue and Pulse Data," 2021.

[10] Y. Hao *et al.*, "A noninvasive, economical, and instant-result method to diagnose and monitor type 2 diabetes using pulse wave: case-control study," *JMIR mHealth and uHealth*, vol. 7, no. 4, p. e11959, 2019.

[11] H.-C. Chen *et al.*, "Pulse-line intersection method with unboxed artificial intelligence for hesitant pulse wave classification," *Information Processing & Management*, vol. 59, no. 2, p. 102855, Mar. 2022, doi: 10.1016/j.ipm.2021.102855.

[12] M. Zhang, X. Li, and L. Wang, "An Adaptive Outlier Detection and Processing Approach Towards Time Series Sensor Data," *IEEE Access*, vol. 7, pp. 175192–175212, 2019, doi: 10.1109/ACCESS.2019.2957602.

[13] F. Thabtah, S. Hammoud, F. Kamalov, and A. Gonsalves, "Data imbalance in classification: Experimental evaluation," *Information Sciences*, vol. 513, pp. 429–441, 2020.

[14] H.-C. Chen, E. Prasetyo, S. S. Kusumawardani, S.-S. Tseng, T.-L. Kung, and K.-Y. Wang, "Learning Performance Prediction with Imbalanced Virtual Learning Environment Students' Interactions Data," in *International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, 2021, pp. 330–340.

[15] D. Devi, S. K. Biswas, and B. Purkayastha, "Learning in presence of class imbalance and class overlapping by using one-class SVM and undersampling technique," *Connection Science*, vol. 31, no. 2, pp. 105–142, 2019.

[16] E. H. Budiarto, A. Erna Permanasari, and S. Fauziati, "Unsupervised Anomaly Detection Using K-Means, Local Outlier Factor and One Class SVM," in *2019 5th International Conference on Science and Technology (ICST)*, Yogyakarta, Indonesia, Jul. 2019, pp. 1–5. doi: 10.1109/ICST47872.2019.9166366.

[17] J. Cui, L. Tu, J. Zhang, S. Zhang, Z. Zhang, and J. Xu, "Analysis of Pulse Signals Based on Array Pulse Volume," *Chin. J. Integr. Med.*, vol. 25, no. 2, pp. 103–107, Feb. 2019, doi: 10.1007/s11655-018-2776-y.

[18] K. Gokcesu, M. M. Neyshabouri, H. Gokcesu, and S. S. Kozat, "Sequential outlier detection based on incremental decision trees," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 993–1005, 2018.

[19] D. Cortes, "Explainable outlier detection through decision tree conditioning," *arXiv preprint arXiv:2001.00636*, 2020.

[20] S. Y. Sheikh and M. T. Jilani, "A ubiquitous wheelchair fall detection system using low-cost embedded inertial sensors and unsupervised one-class SVM," *J Ambient Intell Human Comput*, Apr. 2021, doi: 10.1007/s12652-021-03279-6.

[21] Z. Wang, Y. Fu, C. Song, P. Zeng, and L. Qiao, "Power System Anomaly Detection Based on OCSVM Optimized by Improved Particle Swarm Optimization," *IEEE Access*, vol. 7, pp. 181580–181588, 2019, doi: 10.1109/ACCESS.2019.2959699.

[22] R. B. Zhang, L. H. Xia, and Y. Lu, "Anomaly Detection of ICS based on EB-OCSVM," *J. Phys.: Conf. Ser.*, vol. 1267, no. 1, p. 012054, Jul. 2019, doi: 10.1088/1742-6596/1267/1/012054.

[23] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020.

[24] K. P. Sinaga and M.-S. Yang, "Unsupervised K-means clustering algorithm," *IEEE access*, vol. 8, pp. 80716–80727, 2020.

[25] A. F. Jahwar and A. M. Abdulazeez, "Meta-heuristic algorithms for K-means clustering: A review," *PalArch's Journal of Archaeology of Egypt/Egyptology*, vol. 17, no. 7, pp. 12002–12020, 2020.

[26] S.-H. Jung, H. Lee, and J.-H. Huh, "A novel model on reinforce K-means using location division model and outlier of initial value for lowering data cost," *Entropy*, vol. 22, no. 8, p. 902, 2020.

[27] C. Chen, Y. Wang, W. Hu, and Z. Zheng, "Robust multi-view k-means clustering with outlier removal," *Knowledge-Based Systems*, vol. 210, p. 106518, 2020.

[28] P. J. Jones *et al.*, "FilterK: A new outlier detection method for k-means clustering of physical activity," *Journal of biomedical informatics*, vol. 104, p. 103397, 2020.

[29] S. Zhao, W. Li, and J. Cao, "A user-adaptive algorithm for activity recognition based on k-means clustering, local outlier factor, and multivariate gaussian distribution," *Sensors*, vol. 18, no. 6, p. 1850, 2018.

[30] B. Angelin and A. Geetha, "Outlier Detection using Clustering Techniques–K-means and K-median," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020, pp. 373–378.

[31] S. Kanjanawattana, "A novel outlier detection applied to an adaptive k-means," *International Journal of Machine Learning and Computing*, vol. 9, no. 5, pp. 569–574, 2019.

[32] Prayitno *et al.*, "A Systematic Review of Federated Learning in the Healthcare Area: From the Perspective of Data Properties and Applications," *Applied Sciences*, vol. 11, no. 23, Art. no. 23, Jan. 2021, doi: 10.3390/app112311191.