Development of a personalized anomaly detection model to detect motion artifacts over PPG data using *catch22* features

Andrea Valerio[1], Danilo Demarchi[1], Brendan O'Flynn[2], Salvatore Tedesco[2,*]

[1]Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Torino, Italy

[2]Tyndall National Institute, University College Cork, Lee Maltings Complex, Dyke Parade, T12R5CP Cork, Ireland

salvatore.tedesco.@tyndall.ie

Abstract— **As remote health monitoring grows, it's crucial to distinguish high-quality biomedical signals from low-quality ones. Identifying and mitigating motion artifacts (MAs) is essential for accurate data from wearable devices. Methods: In this work, a high-performing subset of time-series features denoted as catch22 (22 CAnonical Time-series CHaracteristics) was used to detect the presence of MAs in photoplethysmogram (PPG) data acquired from the brachial and digital artery of 31 healthy subjects. Three unsupervised algorithms were employed along with catch22 to detect MAs within the dataset, these were: One-Class Support Vector Machine, Isolation Forest, and Local Outlier Factor. Results: Aggregated precision, recall, and F1-score were computed per each method to assess the detection performances according to a variety of features and anomalies' distribution. These metrics resulted respectively equal to 0.5, 0.64, and 0.55 for OC- SVM, 0.91, 0.94, and 0.92 for IF, and 0.74, 0.75, and 0.74 for LOF. Conclusion: Experimental findings illustrate that by employing the catch22 feature subset, it is viable to discern the presence of MAs in beat-to-beat pulse waveforms without recurring to prior knowledge or data-driven PPG features.**

Keywords—Anomaly detection, PPG, catch22, Isolation forest.

I. INTRODUCTION

Wearable health monitoring devices incorporating plethysmography (PPG) sensors have demonstrated significant potential in advancing cardiovascular disease monitoring [1]. Signal quality degradation arises from changes in blood flow velocity induced by motion, relative movement between PPG sensors and human skin, or low tissue perfusion [2]. In artifact detection, many studies traditionally utilize features derived from waveforms, and spectral domain data. However, a notable challenge stems from the variability in data signals among participants with diverse pulse waveform morphologies [3]. To address this limitation, catch22 (22 CAnonical Time-series CHaracteristics) stands out as a promising solution [4]. This high performing feature subset captures a diverse and interpretable time series signature and eliminates the need for user-specific parameter adjustments and avoids dependence on hand-crafted features, such as peak-to-peak amplitude, heart rate, and pulse width that may not universally apply to subjects [5], [6]. Depending on the field of application, several anomaly detection algorithms have been implemented to detect outliers within an ensemble of data [7]. Supervised anomaly detection techniques create a predictive hypothesis function to discern abnormal behavior, relying on annotated instances to establish expected patterns and outlier classes [8]. On the contrary, unsupervised anomaly detection techniques address the detection of outliers in unlabeled data. These methods operate based on the assumption that regular instances are more prevalent than outliers, and normal instances tend to form dense clusters, while outliers are sparser. In this work, three unsupervised methods with different detection mechanisms were used to identify anomalies in physiological data sets, these were: One Class-SVM, Isolation Forest, and Local Outlier Factor. In this application, our contributions are as follows: (1) We compared three anomaly detection algorithms, assessing their performance when customized for specific subjects versus traditional training on the entire dataset; (2) We observed a significant improvement of up to 41% in motion artifact detection by fine-tuning one of the tested methods; and (3) We demonstrated the feasibility of detecting MAs in multi-site data using a subset of features that had not been previously employed in this context.

This publication has emanated from research supported by a research grant from the Disruptive Technologies Innovation Fund (DTIF) project HOLISTICS funded by Enterprise Ireland (EI). Aspects of this research have emanated from research conducted with the financial support of Science Foundation Ireland under Grants 12/RC/2289-P2-INSIGHT, 16/RC/3835-VISTAMILK and 16/RC/3918-CONFIRM, which are co-funded under the European Regional Development Fund.

II. MATERIAL AND METHODS

A. Data Collection Protocol

In this study, a custom acquisition system [9] was utilized to detect arterial pulse waveforms between the brachial artery and the digital artery[10], [11]. The data used in this research were initially gathered to examine blood pressure variations during cognitive and physical tasks [12]. A pre-clinical trial was conducted at the Tyndall National Institute in University College Cork (UCC), Ireland, with approval from the UCC Clinical Research Ethics Committee of the Cork Teaching Hospitals. The trial involved 31 healthy volunteers (20 males, 11 females) aged 21 to 34 years. Before starting each data collection, the operator identified the optimal location for the brachial artery using tactile arterial palpation, marking the spot for consistency throughout the session. Twelve acquisitions

2024 IEEE SENSORS | 979-8-3503-6351-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/SENSORS60989.2024.10785142

Table 1 Computed evaluation metrics per each anomaly detection model

Training Strategy	MLA	$Precision_{AD}$	$Recall_{AD}$	$F1\text{-}Score_{AD}$	Precision	Recall	F1-Score
Entire Dataset	OC-SVM	0.39	0.35	0.37	0.69	0.68	0.68
	IF	0.53	0.49	0.51	0.77	0.74	0.75
	LOF	0.43	0.39	0.41	0.71	0.69	0.70
Aggregated*	OC-SVM	0.50 ± 0.17	0.64 ± 0.31	0.55 ± 0.20	0.75 ± 0.09	0.82 ± 0.16	0.77 ± 0.10
	IF	0.91 ± 0.14	0.94 ± 0.11	0.92 ± 0.12	0.95 ± 0.07	0.97 ± 0.05	0.96 ± 0.06
	LOF	0.74 ± 0.28	0.75 ± 0.26	0.74 ± 0.27	0.87 ± 0.14	0.87 ± 0.14	0.87 ± 0.14

Abbreviations: MLA, machine learning algorithm, $Precision_{AD}$, $Recall_{AD}$ and $F1\text{-}Score_{AD}$ refer to the scoring criteria computed for the anomaly class.
* Evaluation metrics computed averaging the results achieved across subjects within dataset

were obtained from both the brachial and digital artery for every subject. The data collection protocol included three phases: the first three acquisitions were taken with the subject at rest, acquisitions four to nine were recorded during two cognitive tasks, and the final three acquisitions were measured after a 10-minute treadmill session.

B. *Data Processing*

The processing steps were applied to extract features for detecting motion artifacts. All pulse waveforms were processed uniformly, with feature extraction performed using the catch22 library. As the first step, the duration of all records was standardized to retrieve the same amount of data samples. In particular, the duration window for each dataset was set at 60 seconds. To remove the presence of the DC offset and high-frequency noise, each acquisition then underwent band-pass filtering between 0.3 Hz and 15 Hz, implemented with a 4th-order Chebyshev II filter [2]. Subsequently, the signal was divided into non-overlapping 3-second windows for a total of 20 epochs per each record. This segmentation served the purpose of partitioning the signals into segments of equal length. Moreover, it allowed the identification of any pulse waveforms corrupted by the presence of MAs. All detected epochs were annotated and subsequently utilized as labels for assessing the performance of the tested algorithms. Finally, the catch22 library was used to extract the 22 features from all windows along with mean value, standard deviation, skewness, kurtosis, and the pairwise correlation coefficient between two epochs sharing the same ID and derived from the two mentioned electrode placements. Among the extensive dataset, only 85 epochs (approx. 0.57%) were identified as corrupted by MAs. These instances were distributed across 18 out of the total 31 participants

C. *Anomaly Detection Algorithms*

This study performs a comparative analysis of anomaly detection algorithms, assessing their efficacy when customized for specific subjects w.r.t the conventional training process applied across the entire dataset. Three algorithms representing the current state-of-the-art were selected: One Class-SVM (OC-SVM), Isolation Forest (IF), and Local Outlier Factor (LOF). Each approach was selected based on the diverse methodology used for detecting anomalies within datasets [13], [14], [15]. We compared the outlier detection performance of the mentioned methods when applied directly to the entire dataset and singularly on each subject. In both cases, hyperparameter optimization was conducted for the adopted algorithms. In the first case, optimization was performed on the entire population, while in the second case, tuning was customized for each individual subject. Consequently, models were trained and tested for each subject in the dataset, with the optimization process involving various parameters. Firstly, the number of features ranged from 1 to 27. Highly correlated features were removed first, followed by sequential removal until the optimum subset was identified. The contamination parameter, which affects the model's ability to detect anomalies by defining the proportion of outliers in the dataset, was then tuned, ranging from 0.001 to 0.1. We quantified the performances of the mentioned methods by employing metrics such as precision, recall, and F1-score.
Finally, the performance of each model were assessed by computing the mean value and standard deviation of the mentioned metrics derived for all subjects in the dataset.

III. RESULTS

A. *Model Evaluation*

Table I compares the aggregated prediction performance of three anomaly detection algorithms when personalized for specific subjects versus results obtained using the entire dataset. The evaluation of motion artifact recognition algorithms on the complete dataset, as depicted in Table I, reveals a general challenge in achieving robust performance in identifying MAs in PPG data. Specifically, the IF algorithm demonstrates the highest performance among the tested methods, with an F1-score of 0.51, though deemed unsatisfactory. LOF shows a score of 0.41, while the OC-SVM algorithm performs least favorably with a score of 0.37. A similar trend is observed in the second scenario, where aggregated metrics indicate IF as the best-performing algorithm. It achieves an F1 score of 0.94, whereas LOF and OC-SVM score 0.73 and 0.53, respectively. Fig. 1 illustrates the F1-score performance for each subject across the three algorithms. In all comparisons, IF emerges as the top-performing anomaly detection algorithm, maintaining a score of at least 0.8 for all subjects except subject #17, where performance degradation is observed across all methods. Conversely, OC-SVM consistently ranks as the least effective method, yielding an F1-score of 0.3. Although LOF matches IF's performance in half of the subjects, its effectiveness decreases significantly in the remaining cases. The higher consistency of IF compared to LOF is reflected in the

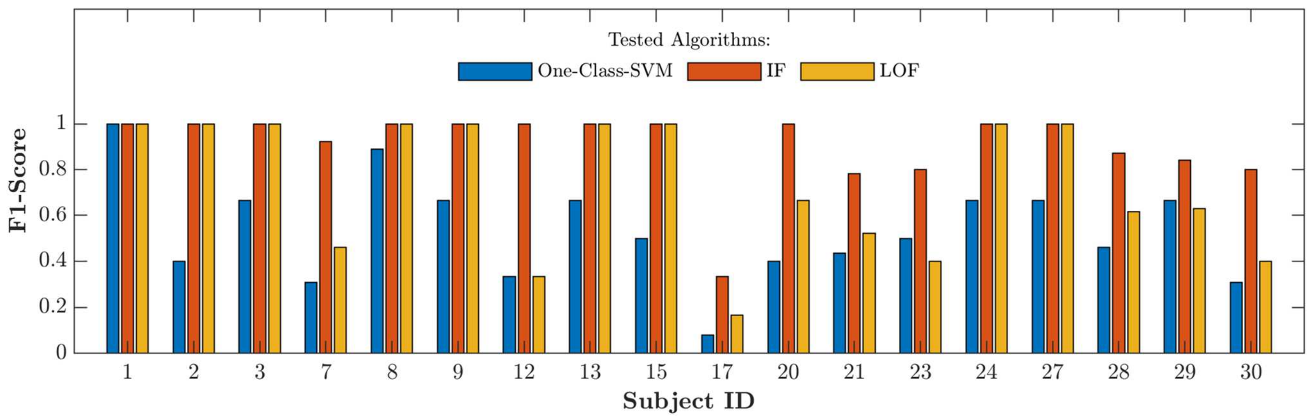

Figure 2: F1- score computed for target class across subjects displying the presence of motion artifacts.

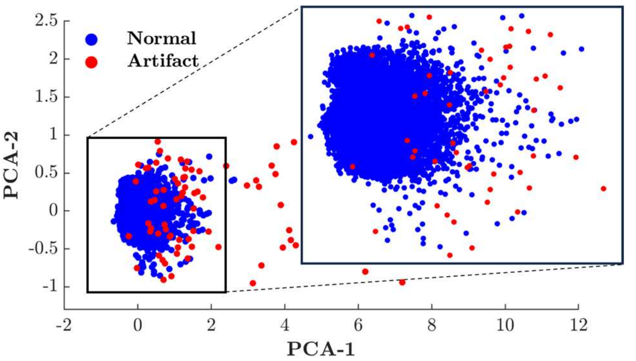

Figure 1: Principal component analysis employed to highlight the intrinsic discrepancies between normal data belonging to different subjects.

lower standard deviation across all scoring criteria. Notably, for subject #12, LOF exhibits a 65% reduction in performance compared to IF, highlighting variability in efficacy across different subjects.

IV. DISCUSSION

This work investigated the identification of MAs in PPG signals collected from diverse locations through a high-performing feature subset known as *catch22* along with anomaly detection algorithms.

In artifact detection, numerous studies have traditionally employed waveform-derived features, such as heart rate, amplitude, pulse morphology, and spectral features [3], [6]. However, a significant limitation arises from the inter-subject variability in participants with diverse pulse waveform morphologies. In addressing this drawback, *catch22* emerges as a promising solution as it prevents the need for user-specific parameter adjustments and avoids reliance on pre-defined features that may not be universally applicable across subjects. In addition, this study examines the disparities in performance between customizing the methods for individual subjects and employing a more generalized approach. In the context of anomaly detection, three specific algorithms were thoughtfully chosen: OC-SVM, IF, and LOF [16], [17], [18]. By analyzing the motion artifact recognition algorithms across the entire dataset, as outlined in Table I, unveils a pervasive challenge in achieving consistent and robust performance in the identification of MAs within PPG data. These results can be clarified by examining the underlying mechanisms of the tested anomaly detection algorithms. These unsupervised models employ distinct methodologies for identifying anomalies, leading to diverse performances on the same dataset. PCA was employed to visualize the distribution in space available in the dataset. Fig. 2 provides a representation based on the first two principal components (explained variance equal to 98%) highlighting epochs affected by artifacts in red and normal instances depicted in blue. Examining the relative position of artifacts in close proximity to the blue cluster offers valuable insights. This observation could clarify the challenges faced by OC-SVM and LOF, both reliant on the spatial distribution of normal points for anomaly detection. To address this, we explored a personalized solution by tuning each model specifically for individuals exhibiting one or more MAs. The results demonstrated that machine learning models personalized to subjects' physiological characteristics significantly enhanced predictive performance compared to a traditional generalized approach. This approach aligns with the broader trend in healthcare studies, where various studies aim to monitor physiological parameters such as blood pressure [19], diabetes management [20], and stress [21]. In our work, we adopted an unsupervised approach to overcome the drawbacks associated with supervised methods, including the need for data labeling and handling missing data in long-term monitoring using wearable technology [3], [22]. By personalizing each model to subject data and extracting the most informative features, we successfully mitigated the challenges associated with local anomaly detection, improving the identification of MAs up to 41% for the F1-score on the IF.

V. CONCLUSIONS

This study explores motion artifact detection in PPG data using unsupervised algorithms trained with catch22 features. We compare anomaly detection algorithms, evaluating their efficacy when tailored for specific subjects versus conventional training on the entire dataset. Our findings show significant improvement in motion artifact detection through subject-specific hyperparameter fine-tuning. The study highlights the effectiveness of catch22 features even with personalized model training. Overall, it underscores the importance of developing personalized models in health monitoring to better cater to individual needs and enhance personal healthcare.

ACKNOWLEDGMENT

We would like to acknowledge and thank the participants, researchers, and clinicians for their contributions to this work.

References

[1] H. Herrmann and H. Ewald, “False Positives Avoidance in Pulse Detection from ECG and PPG Sensor Signals,” in *Proceedings of IEEE Sensors*, Institute of Electrical and Electronics Engineers Inc., 2021. doi: 10.1109/SENSORS47087.2021.9639525.

[2] J. Park, H. S. Seok, S. S. Kim, and H. Shin, “Photoplethysmogram Analysis and Applications: An Integrative Review,” *Frontiers in Physiology*, vol. 12. Frontiers Media S.A., Mar. 01, 2022. doi: 10.3389/fphys.2021.808451.

[3] Z. Guo, C. Ding, X. Hu, and C. Rudin, “A supervised machine learning semantic segmentation approach for detecting artifacts in plethysmography signals from wearables,” *Physiol Meas*, vol. 42, no. 12, Dec. 2021, doi: 10.1088/1361-6579/ac3b3d.

[4] C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones, “catch22: CAnonical Time-series CHaracteristics: Selected through highly comparative time-series analysis,” *Data Min Knowl Discov*, vol. 33, no. 6, pp. 1821–1852, Nov. 2019, doi: 10.1007/s10618-019-00647-x.

[5] Z. Guo, C. Ding, X. Hu, and C. Rudin, “A supervised machine learning semantic segmentation approach for detecting artifacts in plethysmography signals from wearables,” *Physiol Meas*, vol. 42, no. 12, Dec. 2021, doi: 10.1088/1361-6579/ac3b3d.

[6] P. H. Charlton, P. Celka, B. Farukh, P. Chowienczyk, and J. Alastruey, “Assessing mental stress from the photoplethysmogram: A numerical study,” *Physiol Meas*, vol. 39, no. 5, May 2018, doi: 10.1088/1361-6579/aabe6a.

[7] L. Zhang and L. Liu, “Data Anomaly Detection Based on Isolation Forest Algorithm,” in *Proceedings of the International Conference on Computation, Big-Data and Engineering 2022, ICCBE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 87–89. doi: 10.1109/ICCBE56101.2022.9888169.

[8] A. Chatterjee and B. S. Ahmed, “IoT anomaly detection methods and applications: A survey,” *Internet of Things*, vol. 19, p. 100568, Aug. 2022, doi: 10.1016/j.iot.2022.100568.

[9] A. Valerio *et al.*, “Development of a PPG-based hardware and software system deployable on elbow and thumb for real-time estimation of pulse transit time,” in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE, Jul. 2023, pp. 1–5. doi: 10.1109/EMBC40787.2023.10340784.

[10] Q. Wan *et al.*, “Toward Real-Time Blood Pressure Monitoring via High-Fidelity Iontronic Tonometric Sensors with High Sensitivity and Large Dynamic Ranges,” *Adv Healthc Mater*, vol. 12, no. 17, Jul. 2023, doi: 10.1002/adhm.202202461.

[11] A. Mol, A. B. Maier, R. J. A. van Wezel, and C. G. M. Meskers, “Multimodal Monitoring of Cardiovascular Responses to Postural Changes,” *Front Physiol*, vol. 11, Mar. 2020, doi: 10.3389/fphys.2020.00168.

[12] A. Valerio, D. Demarchi, B. O’Flynn, P. Motto Ros and S. Tedesco, “Development of a personalized multiclass classification model to detect blood pressure variations associated with physical or cognitive workload,” *Sensors* MDPI (*Submitted*).

[13] O. Ye. Rodionova, P. Oliveri, and A. L. Pomerantsev, “Rigorous and compliant approaches to one-class classification,” *Chemometrics and Intelligent Laboratory Systems*, vol. 159, pp. 89–96, Dec. 2016, doi: 10.1016/j.chemolab.2016.10.002.

[14] G. O. Campos *et al.*, “On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study,” *Data Min Knowl Discov*, vol. 30, no. 4, pp. 891–927, Jul. 2016, doi: 10.1007/s10618-015-0444-8.

[15] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *2008 Eighth IEEE International Conference on Data Mining*, IEEE, Dec. 2008, pp. 413–422. doi: 10.1109/ICDM.2008.17.

[16] L. Fan, J. Ma, J. Tian, T. Li, and H. Wang, “Comparative Study of Isolation Forest and LOF algorithm in anomaly detection of data mining,” in *Proceedings - 2021 International Conference on Big Data, Artificial Intelligence and Risk Management, ICBAR 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 1–5. doi: 10.1109/ICBAR55169.2021.00008.

[17] L. Zhang and L. Liu, “Data Anomaly Detection Based on Isolation Forest Algorithm,” in *Proceedings of the International Conference on Computation, Big-Data and Engineering 2022, ICCBE 2022*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 87–89. doi: 10.1109/ICCBE56101.2022.9888169.

[18] 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)*, IEEE, Jul. 2019, pp. 1–5. doi: 10.1109/ICST47872.2019.9166366.

[19] J. Leitner, P. H. Chiang, and S. Dey, “Personalized Blood Pressure Estimation Using Photoplethysmography: A Transfer Learning Approach,” *IEEE J Biomed Health Inform*, vol. 26, no. 1, pp. 218–228, Jan. 2022, doi: 10.1109/JBHI.2021.3085526.

[20] A. Rezgui, “Scaling Personalized Machine Learning through DTW Clustering: Predicting Glycemia Levels as an Example,” in *2023 IEEE International Conference on Big Data (BigData)*, IEEE, Dec. 2023, pp. 5356–5364. doi: 10.1109/BigData59044.2023.10386408.

[21] R. G. Bangani, V. Menon, and E. Jovanov, “Personalized Stress Monitoring AI System for Healthcare Workers,” in *Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 2992–2997. doi: 10.1109/BIBM52615.2021.9669321.

[22] J. S. Sunny *et al.*, “Anomaly Detection Framework for Wearables Data: A Perspective Review on Data Concepts, Data Analysis Algorithms and Prospects,” *Sensors*, vol. 22, no. 3. MDPI, Feb. 01, 2022. doi: 10.3390/s22030756.