Effortless Beat-Level Classification of PACs and PVCs in PPG Signals

Eniko Vargova[1], Andrea Nemcova[1], Lucie Saclova[1], Radovan Smisek[1,2], Vratislav Saleny[1], Milan Sepsi[3,4], Zuzana Novakova[3]
[1]Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Brno, Czechia
[2]Institute of Scientific Instruments of the Czech Academy of Sciences, Brno, Czechia
[3]Faculty of Medicine, Masaryk University, Brno, Czechia
[4]Department of Internal Medicine and Cardiology, University Hospital Brno, Brno, Czechia

Abstract

Accurate classification of premature beats, particularly premature atrial contractions (PACs) and premature ventricular contractions (PVCs), is clinically important but has not yet been thoroughly explored in PPG signals. While ECG remains the gold standard, PPG offers a non-invasive, user-friendly, and easy-to-use alternative that can be integrated into wearable devices, enabling continuous monitoring, early screening, and assessment across a broader population. Most existing studies focus on detecting only PVCs or classifying PPG segments rather than individual beats, and are often trained on small datasets, which limits their generalizability.

We propose a simple and effortless method based on only two inter-beat interval (IBI)-derived features to distinguish PACs from PVCs at the individual beat level in PPG. To overcome the limitation of small PPG datasets, the method is trained on simulated PPG signals generated from RR intervals derived from real ECG recordings of over 500 individuals, providing a large-scale and diverse dataset that enables a robust and generalizable approach.

The approach is evaluated on both simulated and real-world PPG recordings, achieving F1-scores of 0.77 and 0.76, respectively. These results demonstrate that the method is robust, computationally efficient, and capable of classifying PACs and PVCs at the beat level, making it suitable for practical applications in wearable PPG monitoring and screening across a broad population.

1. Introduction

Premature contractions (PCs), also called ectopic beats, occur independently of the normal heart rhythm and are commonly observed even in healthy individuals. They are classified into premature atrial contractions (PACs) and premature ventricular contractions (PVCs). A higher burden, typically exceeding 1,000 beats per day, is associated with increased cardiovascular risk, including atrial fibrillation (AF), heart failure, stroke, and higher mortality [1].

PACs originate from early depolarization in atrial tissue outside the sinoatrial node, whereas PVCs arise from the Purkinje fibers or ventricular myocardium. On ECG, PACs typically present with premature, often abnormal P waves, while PVCs are characterized by wide QRS complexes without a preceding P wave [1]. PCs may occur as couplets, triplets, or in patterns such as bigeminy and trigeminy. These patterns, together with symptoms such as palpitations, dizziness, and dyspnea, may indicate underlying cardiac disease or increased arrhythmic risk [2]. Importantly, distinguishing between PACs and PVCs is clinically relevant: frequent PACs may precede AF, whereas a high burden of PVCs is associated with ventricular dysfunction, heart failure, and potentially life-threatening arrhythmias.

Reliable identification of PCs is therefore of considerable clinical importance. While ECG remains the gold standard for detection, its use is typically limited to clinical settings or short-term monitoring. Photoplethysmography (PPG) offers a non-invasive and user-friendly alternative that can be seamlessly integrated into consumer wearables such as smartwatches, enabling continuous monitoring in everyday life and across broader populations, including preventive screening. Importantly, accurate detection and classification of PCs from PPG could not only support early diagnosis but also reduce false AF alarms by distinguishing benign ectopic activity from true arrhythmic episodes.

Despite the potential of PPG for detecting PCs, research in this area remains relatively limited. Existing studies focus on PVC detection [3, 4, 5], likely because PVCs are more prevalent in the general population and are often considered more clinically significant, while PACs have received comparatively less attention. There is a study [6] that addresses both PACs and PVCs, however, it performs classification at the level of 30-second segments rather than on individual beats. In practice, a common limitation of PPG signals is that PCs are not always clearly visible and may merge with the preceding waveform, which can hinder their detection. However, even when PCs are

detectable, a more challenging and still underexplored problem is the reliable classification into PACs and PVCs, as most existing approaches focus mainly on detection of PCs rather than this classification task. Furthermore, the majority of proposed detection methods rely on machine learning (ML) or deep learning (DL) techniques trained on relatively small PPG datasets, often comprising recordings from only a limited number of subjects, which raises concerns about their generalizability to real-world scenarios.

In contrast, this work focuses specifically on the classification of already detected PCs into PACs and PVCs. To address the limited availability of large-scale wearable PPG datasets, we utilize simulated PPG signals generated from RR intervals derived from real ECG recordings with PACs/PVCs from 714 individuals. Although simulated signals cannot fully capture all aspects of real PPG data, they provide a sufficiently representative and scalable basis for the development of the proposed method. The method itself relies on only two simple and computationally efficient features, the intervals to the preceding and subsequent peaks (RR/IBIs). While morphological features were explored, they did not lead to performance improvements. The goal is therefore to distinguish between PACs and PVCs using a minimal and robust feature set that remains practical for real-world deployment, and the proposed approach is subsequently evaluated on real-world PPG recordings to assess its applicability in realistic conditions.

2. Material and Methods

2.1. Data

Simulated PPG dataset

To generate the simulated PPG dataset predominantly for training, recordings from the China Physiological Signal Challenge 2018 (CPSC2018) [7] are utilized. Of all the recordings, we focused on 463 recordings labeled with PACs and 543 labeled with PVCs. The data were resampled from 500 Hz to 250 Hz and segmented into 10-second-long signals, with only high-quality segments, determined based on the expert annotations, retained for further analysis. In total, 822 10-second-long recordings from 714 individuals were included in the final dataset.

For these signals, annotations marking the exact positions of all normal QRS complexes, PVCs, and PACs were created by our team, which has many years of experience in labelling data with arrhythmias. Based on these annotations, RR intervals were extracted and used as input for a PPG simulator designed to simulate arrhythmias [8]. In the simulation, data are generated with varied pulse peak types, enabling the creation of diverse signal patterns.

From the simulated signals, a total of 8,589 PPG peaks were obtained, of which 7,352 corresponded to normal beats and 1,237 to abnormal premature beats (714 PVCs and 523 PACs). Due to the physiological delay between the ECG and PPG, which was also modeled by the simulator, the PPG signals had to be temporally aligned with the ECG annotations. Due to this shift each segment was shortened by one second at both the beginning and the end, resulting in final simulated PPG recordings of 8-seconds in length. This process resulted in a robust dataset, providing PPG signals derived from a large number of individuals (714), allowing the model to learn from a diverse set of normal and abnormal cardiac events.

Real PPG dataset

For evaluation, real-world PPG recordings (1–2 hours each) from 16 subjects were obtained from the MIMIC database [9, 10], all sampled at 250 Hz. Annotations of PACs and PVCs in the PPG signals were performed using the simultaneously acquired ECG recordings as a reference. The corresponding PPG pulses were annotated by our team as PACs, PVCs, or normal PPG pulses corresponding to normal QRS complexes. PPG segments of poor signal quality were identified and annotated. The data were subsequently divided into 30-second windows, with any segment containing low quality excluded from further analysis. From the remaining windows, only those containing at least one PC were retained for the testing dataset. This process yielded a dataset comprising over 10,000 normal peaks and nearly 2,500 PCs (ca. 60% PVCs).

2.2. Preprocessing

All PPG signals underwent a preprocessing pipeline. The procedure included detrending, followed by band-pass filtering with a 3rd-order Butterworth filter in the range of 0.5–2 Hz. The filtered signals were then smoothed using a Savitzky–Golay filter (window length 21, polynomial order 3) and normalized using Min-Max normalization to the [0, 1] range.

2.3. Feature extraction

The proposed algorithm first detects systolic peaks in the PPG signal. For each detected peak, the corresponding onset and offset points are determined, thereby delineating individual PPG waveforms. A label is then assigned to each waveform based on the available annotations, and only those labeled as PAC or PVC are retained for further analysis.

A common challenge in PPG signals is that PCs may not be distinctly visible and can merge with the preceding waveform (Figure 4). In such cases, two annotations (a normal QRS complex and a PAC/PVC event) may correspond to a single PPG waveform. In our approach, the entire waveform is therefore labeled as PAC/PVC.

For each selected waveform, the main peak is located, and two temporal features are extracted: the interval to the preceding peak and the interval to the subsequent peak. These features are normalized by dividing them by the

local heart rate, defined as the median IBI computed over a window of ±5 neighboring beats. The resulting normalized intervals constitute the feature set used in the subsequent analysis. This procedure is applied consistently across the entire dataset.

2.4. Classification method

The modeling stage uses simulated data, divided into training and testing sets (training: 622 signals from 533 subjects including 534 PVC and 422 PAC beats, testing: 200 signals from 181 subjects including 180 PVC and 101 PAC beats). To address the slight imbalance between classes in the training set, the Synthetic Minority Over-sampling Technique (SMOTE) is used to generate additional samples for the minority class. The two features are then standardized using Z-score, calculated from the training data and applied to both the simulated test set and the real-world PPG dataset.

For classification, we use a Support Vector Classifier (SVC) with a radial basis function (RBF) kernel. Hyperparameters were optimized via 3-fold cross-validation on the training set, resulting in C=1 and γ=0.1. The relationship between the two features is not linearly separable, so simple linear models or basic threshold rules do not perform well.

3. Results

The results are summarized in Table 1, which reports precision, recall, and F1-score for the simulated training set, the simulated test set, and the real-world PPG test set. The F1-score on the training set is 0.75, slightly lower than 0.77 and 0.76 observed on the simulated and real test sets, respectively, indicating consistent performance across both simulated and real data.

Table 1. Results

Dataset	Precision	Recall	F1
Train sim. PPG	0.75	0.75	0.75
Test sim. PPG	0.77	0.79	0.77
Test real PPG	0.77	0.77	0.76

This difference reflects the composition of the datasets: the training data include signals from over 500 individuals, resulting in high variability and many challenging cases, whereas the test sets contain fewer subjects, making the classification task slightly less complex. Figures 1 and 2 show examples of correct classifications on real PPG data: Figure 1 illustrates a well-classified PAC with a typical waveform, while Figure 2 shows a well-classified PVC.

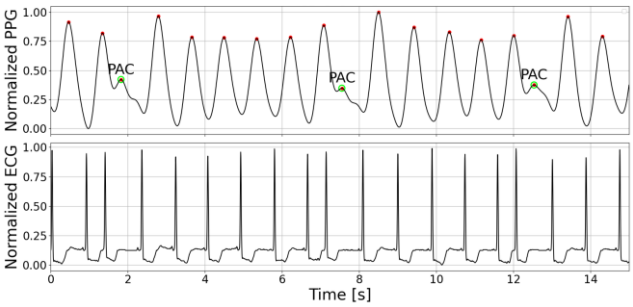

Figure 1. Correctly Classified PACs in real PPG Signal

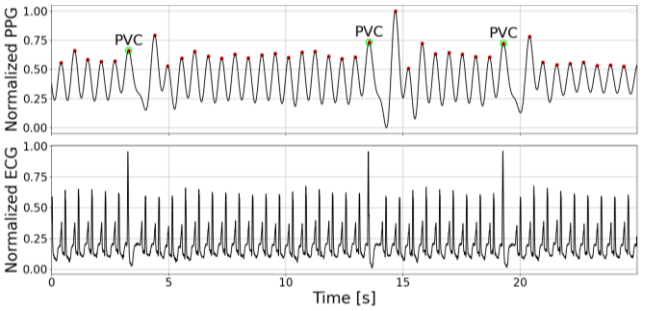

Figure 2. Correctly Classified PVCs in real PPG Signal

4. Discussion

Waveform shape alone does not always allow a reliable distinction between PAC and PVC. In Figure 3, for example, PAC and PVC exhibit very similar morphologies, highlighting that classification based solely on shape can be ambiguous. In these situations, IBI features play a decisive role in correct identification.

We also trained models by adding waveform-based features, such as peak amplitude, onset height, peak width, and area under the curve, to the two IBI features. The results showed that these additional features did not improve performance and, in some cases, slightly reduced it. This confirms that the method performs reliably using only the two simple IBI-based features, keeping the approach lightweight, efficient, and still effective at distinguishing PACs and PVCs in real-world recordings.

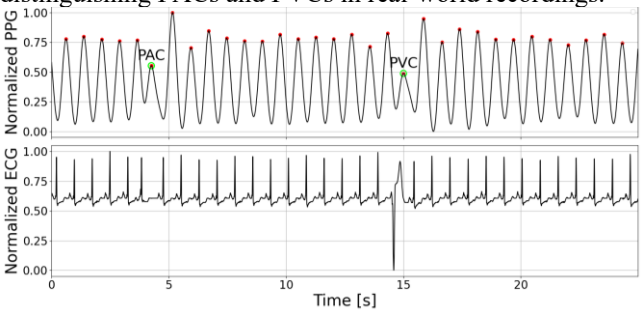

Figure 3. Correctly Classified PAC and PVC in real PPG signal with Similar Morphology

A limitation of the proposed classification is, among others, linked to the detection of PACs and PVCs in PPG signals. PCs do not always produce a distinct, separate

peak, as previously mentioned. Instead, they can merge with the preceding waveform corresponding to a normal QRS complex. This occurs for both PACs and PVCs, although our findings indicate that it is more frequent in PVCs. As a result, the detected segment may include both the normal beat and the extrasystole.

Figure 4 illustrates this effect: the entire waveform is classified as a PVC, which is correct in terms of distinguishing PACs from PVCs, but the PPG signal blends the PVC with the preceding normal beat. Consequently, the apparent rhythm pattern may be distorted. For example, a true sequence such as normal–normal–normal–PVC (quadrigeminy) may appear as normal–normal–PVC due to the merging of the PVC with the preceding beat, thus resembling trigeminy in the PPG signal. This can make the rhythm look different from the real one. Additionally, the detection of PCs may be further limited in the presence of AF due to the highly irregular rhythm. Future work could focus on a more detailed investigation of overlapping beat separation, which represents a highly challenging problem and may not be feasible in all cases. Another important research direction using ML models has already been initiated, aiming to improve the reliability of PAC and PVC classification in personalized signals (i.e., for individual subjects), with encouraging initial results.

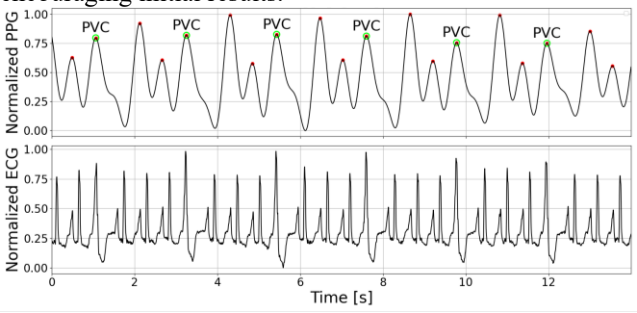

Figure 4. Example of a merged PVC in real PPG signal and its effect on apparent rhythm pattern.

Despite these limitations, the proposed approach still provides a simple and effective method for distinguishing between PACs and PVCs in PPG signals.

Conclusion

This work presents a simple and efficient method for distinguishing between PACs and PVCs in PPG signals using only two IBI-based features. Despite its simplicity, the approach achieves robust performance, reaching an F1-score of 0.77 on the simulated test set and 0.76 on the real-world PPG test set, demonstrating its potential for practical applications.

Acknowledgments

Brno Ph.D. Talent Scholarship Holder – Funded by the Brno City Municipality. This work was partially co-financed by the Technology Agency of the Czech Republic under the TREND 11 Programme (project No. FW11020227).

References

[1] Orini M, Stefan van Duijvenboden, William J. Young, Ramírez J, Jones A, Tinker A, et al., “Premature atrial and ventricular contractions detected on wearable-format electrocardiograms and prediction of cardiovascular events,” *European Heart Journal - Digital Health*, vol. 4, no. 2, pp. 112-118, Feb. 2023.

[2] Zeppenfeld K, Tfelt-Hansen J, de Riva M, Winkel BG, Behr ER, Blom NA, et al., “2022 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death,” *European Heart Journal*, vol. 43, no. 40, pp. 3997-4126, Aug. 2022.

[3] Solosenko A, Petrenas A, Marozas V., “Photoplethysmography-Based Method for Automatic Detection of Premature Ventricular Contractions,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 5, pp. 662-669, 2015.

[4] Yousefi MR, Khezri M, Bagheri R, Jafari R., “Automatic Detection of Premature Ventricular Contraction Based on Photoplethysmography Using Chaotic Features and High Order Statistics,” *2018 IEEE International Symposium on Medical Measurements and Applications*, pp. 1-5, 2018.

[5] Raj R, Kumar US, Maik V., “Enhanced premature ventricular contraction pulse detection and classification using deep convolutional neural networ,” *Physical and Engineering Sciences in Medicine*, vol. 46, no. 7, pp. 1677-1691, Sep. 2023.

[6] Liu Z, Zhou B, Jiang Z, Chen X, Li Y, Tang M, et al., “Multiclass Arrhythmia Detection and Classification From Photoplethysmography Signals Using a Deep Convolutional Neural Network,” *Journal of the American Heart Association*, vol. 11, no. 7, Apr. 2022.

[7] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al., “An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection,” *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368-1373, Sep. 2018.

[8] Sološenko A, Petrėnas A, Paliakaitė B, Marozas V, Sörnmo L. Model for Simulating ECG and PPG Signals with Arrhythmia Episodes. 2022. Available from: 10.13026/s32e-sv15

[9] Moody GB, Mark RG. A Database to Support Development and Evaluation of Intelligent Intensive Care Monitoring. *Computers in Cardiology*. 1996.

[10] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al., “Physiobank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. 215–220, 2000.

Address for correspondence:
Enikö Vargová
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication, Brno University of Technology, Technická 3058/12, 61600 Brno, Czech Republic. xvargo02@vutbr.cz