Intra-Subject Clustering of ECG Heartbeats from Wearable Devices Using Deep Learning and Feature Engineering

Intra-Subject Clustering of ECG Heartbeats from Wearable Devices using Deep Learning and Feature Engineering

1st Federico Digiacomo
Politecnico di Torino
Turin, Italy
federico.digiacomo@polito.it

2nd Gabriella Olmo
Politecnico di Torino
Turin, Italy
gabriella.olmo@polito.it

3rd Alessandro Gumiero
STMicroelectronics
Agrate Brianza, Italy
alessandro.gumiero@st.com

***Abstract*—Cardiovascular diseases (CVDs) represent a significant global health concern, necessitating effective early detection methods. The advent of wearable electrocardiogram (ECG) devices offers the potential for enhanced portability and continuous monitoring. However, the substantial volume of data generated underscores the need for automated analysis techniques. This paper presents a framework for intra-subject clustering of ECG heartbeats, employing a robust P-QRS-T wave classifier based on a Multiscale Convolutional Neural Network (Multiscale CNN) combined with a Bidirectional Gated Recurrent Unit (biGRU). This is followed by heartbeat segmentation, feature engineering using both manual descriptors and the Time Series Feature Extraction Library (TSFEL), and subsequent clustering via a k-means algorithm. Furthermore, a novel ECG database, the Hi Database (HiDB), is introduced. It was acquired using the Hi 3-Leads ECG (Hi-ECG), a wearable three-lead device commercialised by CGM in collaboration with STMicroelectronics. The proposed wave classifier demonstrated strong performance, tested according to the ANSI/AAMI EC57 standard. It achieved an average sensitivity of 97.79% and precision of 96% for QRS detection across the MIT-BIH Arrhythmia Database (MITDB), the American Heart Association ECG Database (AHADB), and the MIT-BIH Noise Stress Test Database (NSTDB). Intra-subject clustering on the MITDB yielded a mean Adjusted Rand Index (ARI) of 0.78 ± 0.19 and a mean Silhouette score of 0.65 ± 0.14. Application of the clustering approach to the HiDB resulted in an average Silhouette score of 0.74 ± 0.13. The findings suggest that the presented framework can support clinicians in ECG beat annotation tasks. Its modular design enables adaptability to additional objectives, such as rhythm anomaly detection, by leveraging QRS information from the wave classifier. Both the wave classifier and the feature-based clustering model demonstrated the robustness of the approach across different ECG data sources. Moreover, the intra-subject setting highlights its potential for personalised cardiac monitoring.**

***Index Terms*—Wearable ECG device, Cardiovascular diseases (CVDs), Clustering, Machine learning, Unsupervised Learning, Heartbeat Anomalies**

I. INTRODUCTION

Cardiovascular diseases (CVDs) are a group of disorders affecting heart and blood vessels, including coronary heart disease, cerebrovascular disease, peripheral arterial disease, rheumatic heart disease, congenital heart disease, deep vein thrombosis and pulmonary embolism. They represent the leading cause of mortality worldwide, with 32% of all global deaths attributed to CVDs in 2019. Specifically, it is estimated that CVDs account for 17.9 million deaths per year, with 85% due to heart attack and stroke. Furthermore, approximately one third of these deaths occur prematurely in individuals under the age of 70 [1], [2].

Electrocardiograms (ECGs) have been considered a routine diagnostic method since their discovery, providing valuable insight into the bio-electric response of the heart in a non-invasive way [3]. Moreover, the need for early detection techniques, together with the adoption of intelligent algorithms, has led to the widespread use of remote wearable ECG devices. These systems are capable of providing early detection and real-time monitoring using a reduced number of leads, thereby enhancing the portability and convenience for the patient [4]. These devices provide complementary information to traditional diagnostic techniques, allowing for the detection of transient cardiac events that may not occur during a scheduled medical appointment in out-of-hospital contexts [5]. On the other hand, these technologies present several limitations. Power consumption for data acquisition and wireless transmission remains a critical issue. Ensuring high signal quality and effective noise suppression is also challenging, particularly when the device is self-applied by the subject or used during physical activity. Furthermore, the reduced number of leads provides less comprehensive spatial information about the heart's electrical activity, while aspects such as patient privacy and long-term safety also require further attention [3], [6]. The analysis of the electrical signal associated with each heartbeat enables the extraction of a wide range of information to assess cardiac health and identify abnormalities. However, manually analysis of long ECG recordings poses a significant challenge for the medical community due to the time and resources required to review them [7]. This underscores the need for reliable and automatic heart anomaly detection algorithms to assist specialists in managing this large volume of data. Currently, the automation of these approaches primarily relies on machine learning (ML) and deep learning (DL) techniques. The selection between these methods depends on factors such as dataset size, computational constraints, clinical interpretability requirements, detection performance,

and generalisation capability. Current ECG-based Computer-Aided Diagnosis (CAD) systems typically employ two main approaches: classification and clustering. The classification process is generally divided into two categories: single-beat classification, where the algorithm detects abnormal morphology within individual beats, and rhythm classification, which analyses sequences of beats to identify rhythm disorders [8]. Clustering, on the other hand, enables the unsupervised discovery of inter- and intra-individual patterns within ECG data, facilitating the identification of underlying structures, potential rhythm anomalies, and the extraction of clinically meaningful insights from unlabelled signals [9].

This work integrates both approaches, employing them for different purposes and is structured around three main objectives. The first is to develop a reliable deep learning-based model for the identification of P waves, QRS complexes, and T waves. Specifically, we adopt a 1D Multiscale Convolutional Neural Network (Multiscale CNN) combined with a Bidirectional Gated Recurrent Network (biGRU) to capture both local patterns and the temporal dependencies between ECG waves by operating in the frequency domain. Training data were obtained from the Lobachevsky University Electrocardiography Database (LUDB) [10], where specialists manually delineated the boundaries of ECG waves. The second objective is to establish a framework that supports specialists in data labelling, in fact, because the number of abnormal ECG samples labelled by cardiologists is limited, there is a growing need for unsupervised learning methods in ECG analysis [9]. In this framework, following the recognition of ECG waves, the ECG traces are segmented into individual heartbeats, and for each segment, features from different domains are extracted by integrating clinical and physiological knowledge. Python packages, such as the Time Series Feature Extraction Library (TSFEL) [11], are employed to obtain a representative set of features for each beat. A k-means clustering algorithm is then used to detect potential individual abnormal heartbeats within each subject's ECG trace. The clustering results were evaluated using the Massachusetts Institute of Technology Beth Israel Hospital Arrhythmia Database (MITDB) [12], [13]. This database contains expert-annotated heartbeats that belong to the five superclasses of arrhythmias—Normal (N), Supraventricular Ectopic (S), Ventricular Ectopic (V), Fusion (F), and Unknown (Q)—and complies with the ANSI/AAMI EC57:1998/(R) 2008 standard proposed by the Association for the Advancement of Medical Instrumentation (AAMI) [14]. Finally, the third objective is to adapt the approach to a newly created ECG database, referred to as the Hi Database (HiDB). The data were collected during an experimental session involving volunteer subjects from two residential care facilities (RCFs) in Piedmont, Italy, managed by the Valdocco Cooperative [15]. The device used for data collection was the CompuGroup Medical (CGM) Hi 3 Leads ECG (Hi-ECG), a CE-certified wearable ECG recording device (class IIa medical device) commercialised by CGM and co-developed with STMicroelectronics [16].

II. MATERIALS AND METHODS

A. LUDB and MITDB Overview

The LUDB is a public ECG database containing 200 twelve-lead, 10-s recordings sampled at 500 Hz, with manual annotations of P, QRS, and T waves. The dataset, collected at the Nizhny Novgorod City Hospital from 200 healthy volunteers and patients with various cardiovascular conditions (age 11–89) , includes 58,429 annotated waves across 8 seconds per recording [10]. The MITDB contains 48 half-hour two-channel recordings from 47 subjects, sampled at 360 Hz, with annotations for both heartbeat types and rhythms. Only the first annotations for the modified lead II (MLII) were used. A total of 46 subjects were considered in this work, resulting in 104,748 heartbeats categorised into the five AAMI heartbeat classes. Figure 1 illustrates the total number of labelled beats and their class distribution. It is evident that the AAMI class distribution is highly imbalanced, with class N being the most represented. Figure 2 shows an example of a heartbeat from each of the four anomaly classes in the dataset.

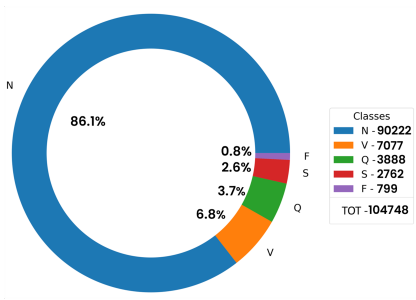

Fig. 1. Distribution of classes in the MITDB (MLII).

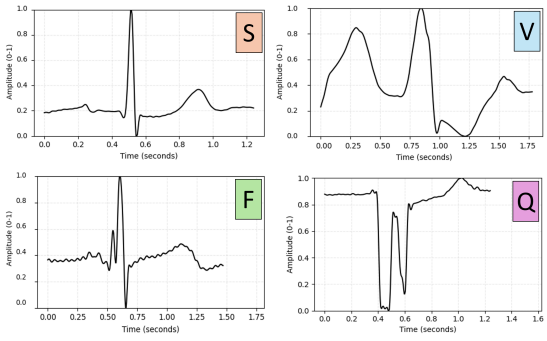

Fig. 2. Examples of four types of AAMI heartbeat anomalies (S, V, F, Q) from the MITDB.

B. Hi-ECG Device and HiDB Construction

The Hi-ECG device is equipped with a three bipolar ECG system, with two electrode connectors located directly on the central unit and a third, referred to as the "mobile electrode", connected via cable. Additionally, it includes a micro-electromechanical systems (MEMS) triaxial accelerometer, allowing simultaneous recording of the subject's body position and physical activity. According to the device manual, signals

can be acquired using one of three possible configurations shown in Figure 3, each providing comprehensive information on cardiac electrical activity in the frontal plane.

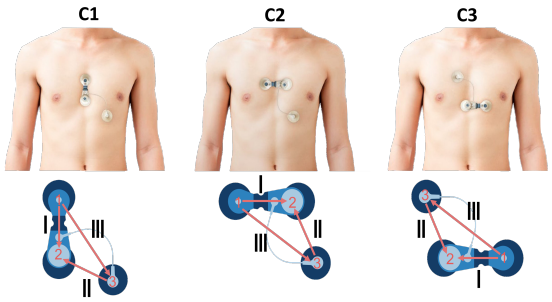

Fig. 3. C1, C2, and C3 configurations of Hi-ECG (from Hi-ECG manual)

The device supports three operational modes: Tele-Holter, Tele-Monitoring, and Tele-ECG. The Tele-ECG mode enables remote at rest ECG acquisition [16]. In this study, this mode was used and data were transmitted via Bluetooth 4.2 to a mini PC emulating the patient application, equipped with a 12th Gen Intel(R) Core(TM) i7-12700H (2.30 GHz quad-core), 16 GB RAM, NVIDIA GeForce RTX 3070ti 8GB GDDR6, HDMI and USB3 ports, and running Windows 11.

Data collection was carried out following a formal agreement with the Valdocco Association. A total of 25 volunteers (13 women, mean age 84.6 ± 12.2 years) participated in the experiment. A formal agreement was signed with the Valdocco Association to enable data collection during the study. Inclusion criteria required stable general health, with no major comorbidities or cognitive impairments, to ensure participant safety and the reliability of data collection. All subjects remained awake and seated during the recordings. The distribution of confirmed clinical conditions was highly variable and included healthy individuals, stroke survivors, individuals with multiple sclerosis, and those with Parkinson's disease. Additionally, most subjects suffered from atrial hypertension, with some cases of atrial fibrillation and mitral insufficiency. Only one subject also had a pacemaker.

A consistent recording protocol was used for each subject to ensure reproducibility. After skin preparation, FIAB disposable ECG electrodes (36×45 mm, model F9089) with solid gel, Ag/AgCl sensor, and stainless steel snap [17] were connected to the device and placed on the subject's chest. For each subject, recordings were performed in all the three device configurations, as described in the user manual. To minimise the risk of misplacement, device positioning was carefully monitored during each acquisition. However, due to individual physiology or the presence of braces and pain relief patches, the recommended placement could not always be achieved; such cases were documented. ECG signals were acquired at 512 Hz, while acceleration data were sampled at 8 Hz. It was noted that the electrode positioned under the battery of the central unit often had poor skin adhesion, leading to inconsistent signal quality in leads 1 and 2. As a result, only lead 3 was used for further analysis, resulting in a total of 88 individual ECG traces. The average duration of ECG trace acquisition was 150 ± 41 s (median 135 s, interquartile range 123-162 s).

C. Common Processing Pipeline

The overall system pipeline is illustrated in Figure 4. All steps were implemented in Python 3.10, primarily using the `PyTorch` and `scikit-learn` libraries [18], [19]. The pipeline consists of three main stages: (i) Heartbeat Segmentation Module, (ii) Feature Engineering, and (iii) Clustering, described in detail in the following subsections. The workflow has been applied to both MITDB and HiDB signals; however, only MITDB contains labelled beats, allowing supervised evaluation and feature selection.

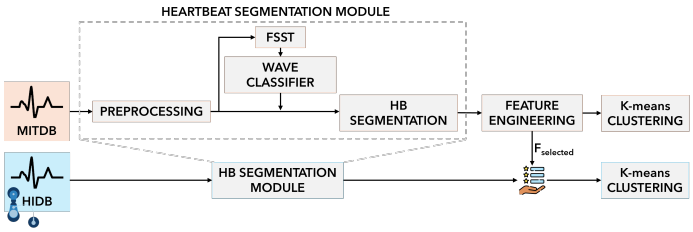

Fig. 4. Architecture of the proposed system.

D. Heartbeat Segmentation Module

A first preprocessing stage includes signal filtering and downsampling. A 4th-order Butterworth bandpass filter (0.5-45 Hz) was applied using zero-phase forward-backward filtering. Subsequently, signals were downsampled to a standard rate of 360 Hz to match the native frequency of the MITDB. As in [20], the Fourier Synchrosqueezed Transform (FSST) was then computed on signal windows of 3000 samples. FSST improves upon the traditional Short-Time Fourier Transform (STFT) by redistributing the energy in the time–frequency (TF) plane based on the instantaneous frequency of the STFT phase . A 72 sample Kaiser-Bessel window (approximately 0.2 s) with shape parameter $\beta = 10$ was used for analysis. The temporal and spectral resolutions were set to 1 sample and 1 Hz respectively, and analysis was limited to 46 frequencies in the 0–45 Hz range. The real and imaginary components of the FSST were standardised using the Z-score method and concatenated to form the input matrix for the subsequent model. The architecture of the wave classifier, shown in Figure 5 along with the full training pipeline, was optimised for computational efficiency. It takes the FSST as input and first applies a 1D Multiscale CNN with 16 filters of sizes 3, 5, and 7 to capture features at different temporal resolutions. These feature maps are concatenated and passed to the BiGRU with 32 hidden units. The GRU was chosen for its simple structure, relying only on update gates and reset gates. The bidirectional design facilitates the capture of temporal dependencies in both forward and backward directions, thereby enabling a more comprehensive interpretation of ECG waveform dynamics. Finally, two linear layers transform the BiGRU output to

generate, for each FSST sample, a prediction in one of the four classes: P, QRS, T, or n/a (no wave). To mitigate class imbalance, particularly the overrepresentation of the n/a class, a Weighted Focal Loss was adopted. Class weights were defined inversely to class frequency, increasing the contribution of underrepresented classes (e.g., P waves).

To train the model, the LUDB was split by subjects: 70% (144 subjects) for training, 20% (36 subjects) for validation, and 10% (20 subjects) for testing.

The final step involves segmenting the ECG signal into individual heartbeats, starting from QRS detection. The R peak is identified as the absolute maximum within a window equal to 20% of the identified QRS duration, centred at the midpoint between its onset and offset. The heart rate (HR) is estimated using a sliding window of five consecutive R–R intervals centred on the segment of interest and constrained between 30 and 240 beats per minute (BPM) to ensure physiological plausibility. Each heartbeat segment is defined around the R peak, including 40% of the estimated cycle length before and 60% after it. Furthermore, if the segment overlaps by at least 20% with a wave labelled differently from QRS that wave is included in the segment. This strategy avoids imposing a fixed P-QRS-T structure, which may not always be present in abnormal heartbeats.

E. Feature Engineering

For this step, labelled segments from MITDB were used. Feature extraction was performed by combining manually designed morphological and spectral descriptors with automated extraction. Morphological features included wave durations, amplitudes and areas (P, QRS, T), inter-wave intervals and heart rate variability metrics. Spectral features were calculated from the FSST, including absolute and normalised power in the wave bands, dominant frequency, spectral entropy and related metrics. In addition, two features denote the presence and number of P and T waves in each segment. This resulted in 73 morphological and 16 FSST-based features. 60 additional features were extracted using the TSFEL Python library, covering statistical, temporal, fractal and spectral domains. Due to some features returning multiple coefficients rather than a single value, the number of TSFEL-derived features increased to 156. In total, each ECG segment was described by 245 features.

Following feature extraction, approximately 30% of the beats in MITDB exhibited NaN values in P wave-related features, due to both the difficulty in detecting the P wave and its absence in certain abnormal classes. To assess their relevance to AAMI heartbeat classification, the dataset was temporarily restricted to segments with predicted P waves. Mutual Information (MI), which captures both linear and non-linear relationships between each feature and the class label, was computed on this subset, and features with MI scores above the 90th percentile were selected. Among them, only one P wave-related feature (*PR slope*) was found to be informative. As a result, all P wave-related features were excluded from the full dataset. MI-based selection was then repeated on the remaining features, using the same thresholding approach.

F. Clustering

The standardised features were then used as input to a k-means clustering algorithm. For each MITDB subject, clustering performance was evaluated using both the Adjusted Rand Index (ARI) and the Silhouette score. The ARI measures the similarity between predicted clusters and ground truth labels, corrected for the chance of random agreement, with values ranging from -1 to 1. The Silhouette score, on the other hand, assesses cluster quality by measuring how well each sample fits into its assigned cluster compared to others, with values ranging from -1 to 1. To determine the optimal number of clusters k, values from 2 to 5 were tested for each MITDB subject. Clustering performance was prioritised based on the highest ARI value, provided that the Silhouette score exceeded a predefined threshold, thus ensuring a meaningful cluster structure. The same procedure was used for the HiDB subjects, except that only the Silhouette score was used, as no ground truth labels were available. These metrics are computed on a per-subject basis and should be interpreted as relative indicators of intra-subject heartbeat morphology variability.

III. RESULTS

A. Wave Delineation and Segmentation

Figure 6 shows the sample-wise confusion matrix for the wave classifier on the LUDB test set, where each value represents the percentage of correctly classified samples over the total number of samples for that class. The model demonstrates robust QRS detection (97.6%) and a high recall for T waves (95.5%), although a notable number of *n/a* samples are misclassified as T. This is likely due to the longer temporal extent of the T wave compared to other waves. Although the P wave also achieves high recall, a tendency towards over-segmentation is observed, reflecting the inherent difficulty in detecting P waves, especially within a clinically diverse population such as that represented in the LUDB data. Table I summarises precision, recall, and F1-score for each class, along with an overall accuracy of 88%. The low precision (67%) associated with the P wave supports the previous observation, while QRS and T waves achieve a good balance between precision and recall, resulting in F1-scores of 91% and 87%, respectively. Despite *n/a* being the class with the highest support, the impact of this imbalance is mitigated by the use of the Weighted Focal Loss.

The most critical aspect of a wave classifier is the accurate detection of QRS complexes. Consequently, the model was tested for QRS detection in accordance with the ANSI/AAMI EC57:1998/(R)2008 recommendations [14]. This evaluation was conducted using the American Heart Association ECG Database (AHADB) [21], the MITDB, and the MIT-BIH Noise Stress Test Database (NSTDB) [22]. The results, in terms of QRS sensitivity and precision, are summarised in Table II which confirms the model's generalisability across different ECG datasets. The table also includes two examples

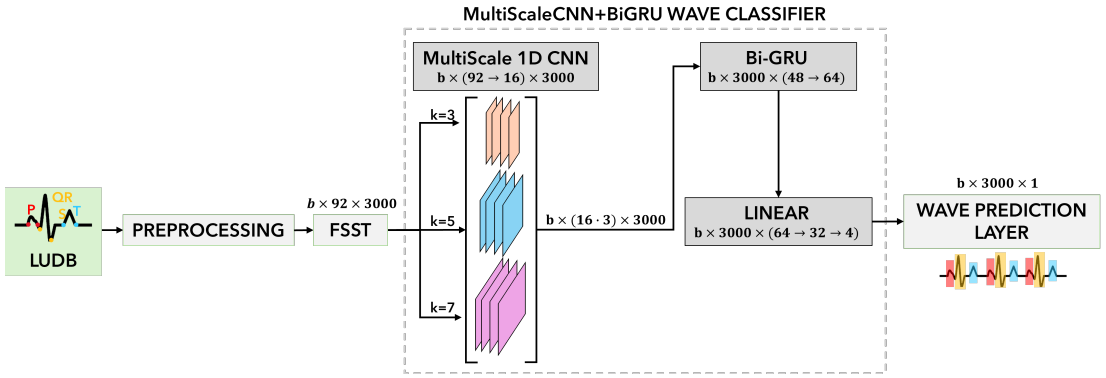

Fig. 5. Overview of the training pipeline and architecture of the wave classifier.

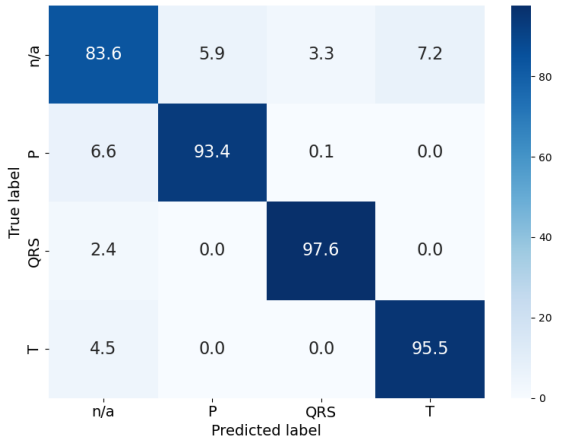

Fig. 6. Confusion matrix on the LUDB test set.

TABLE I
CLASSIFICATION REPORT FOR THE WAVE CLASSIFIER ON THE LUDB.

Class	Precision	Recall	F1-score	Support
n/a	0.97	0.84	0.90	447408
P	0.67	0.93	0.78	56388
QRS	0.85	0.98	0.91	83446
T	0.80	0.96	0.87	132758
Overall Accuracy		**0.88**		720000

of ML-based QRS detection approaches from the literature, which report results consistent with those obtained in this work for MITDB and NSTDB. Each detected QRS complex was assigned to the corresponding AAMI class, followed by segmentation, resulting in 104,747 segments for MITDB and 16,952 for HiDB.

B. Feature Selection and Analysis

A total of 23 features were selected, having MI values above the applied threshold of 0.205. Two additional categorical features, related to the number of P and T waves in each segment (*p count* and *t count*), were included, resulting in a final set of 25 features per heartbeat. Of these, 17 of

TABLE II
QRS DETECTION PERFORMANCE ACROSS DATASETS AND LITERATURE

Author	Database	QRS S (%)	QRS +P (%)
This Work	MITDB	99.75	99.80
This Work	AHADB	99.62	99.75
This Work	NSTDB	94.00	88.47
Literature			
Zahid et al. [23]	MITDB	99.85	99.82
Zhao et al. [24]	NSTDB	95.59	91.03

the top selected features originate from manually designed morphological and spectral descriptors (e.g. RR intervals, QT interval, wave amplitudes), while the remaining 8 were extracted using the TSFEL library and primarily capture spectral and statistical characteristics of the ECG segments (e.g. mean absolute deviation, autocorrelation, absolute energy, average power). This distribution highlights the importance of domain-specific feature engineering tailored to the characteristics of the specific signal under analysis. The inclusion of P- and T-wave counting features was motivated by the observation that the presence or absence of these waves suggests important relationships with the AAMI classes, which could support the clustering model. In particular, Figure 7 shows the distributions of the P waves (blue) and the T waves (green), indicating whether they are absent, present once, or present more than once in each segment. The absence of P waves appears to be especially relevant for classes of ventricular ectopic beats (V) and unknown beats (Q), whose morphologies typically overrides the P wave. A similar pattern is observed in class S, where the supraventricular region is affected, and in the fusion beat class (F), which combines normal and ectopic beats. Furthermore, the number of T waves becomes critical for classes V and S, which often contain two T waves per segment - frequently associated with suppressed or absent P waves. Notably, approximately 20% of segments in the normal (N) class show no predicted P wave. This may be due to classification errors, technical limitations of the ECG, physiological characteristics of the subject, or overlap of the P wave with other waves.

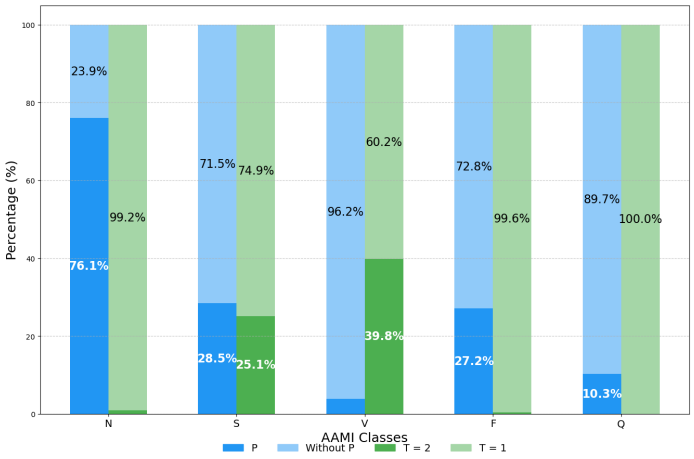

Fig. 7. Distribution of segments in the MITDB based on the number of P and T waves and AAMI classes.

C. Clustering Performance on MITDB

Figure 8 presents the ARI scores for each MITDB subject. All 29 subjects with ARI $>$ 0 exhibit a Silhouette score above the empirical threshold of 0.2, with the exception of subjects 13, 19, and 31, whose traces are entirely classified as normal (N). For these 29 subjects, the mean ARI is 0.78 ± 0.19, and the mean Silhouette score is 0.65 ± 0.14, indicating strong agreement between clustering results and ground truth labels, as well as good intra-cluster cohesion. Conversely, data from subjects with ARI $=$ 0 form a single cluster despite the presence of multiple AAMI classes. These cases are characterised by a strong class imbalance in favour of normal beats (N), and the few abnormal heartbeats typically belong to the supraventricular ectopic class (S) or the unknown class (Q). The former often present morphological patterns similar to normal beats, while the latter lack a well-defined structure and may still overlap with normal beats. An analysis of the AAMI class distribution across all ECG segments of these subjects showed that 89.4% belong to class N, 4.8% to class S, and 5.3% to class Q.

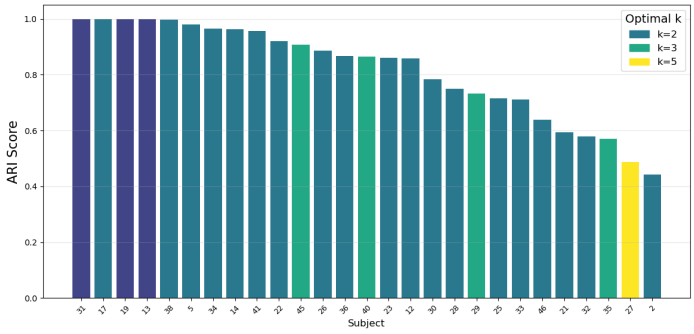

Fig. 8. ARI scores per subject in MITDB (ARI $>$ 0)

These metrics are computed for each subject, thereby capturing individual morphological variability and class distribution. As a result, the model highlights beats that deviate from each subject's typical pattern, aligning with the aim of developing a subject-specific ECG labelling tool. Figure 9 shows the t-SNE projection of the feature distribution for Subject 27 from the MITDB dataset, with AAMI-labeled classes on the left and clustering results on the right. It can be observed that the clustering model tends to group normal (N) and supraventricular (S) beats into a single cluster, while ventricular (V) ectopic beats are more clearly isolated. Additionally, some normal beats with distinct feature profiles are assigned to a separate cluster, suggesting that the model captures subtle morphological differences within the normal class.

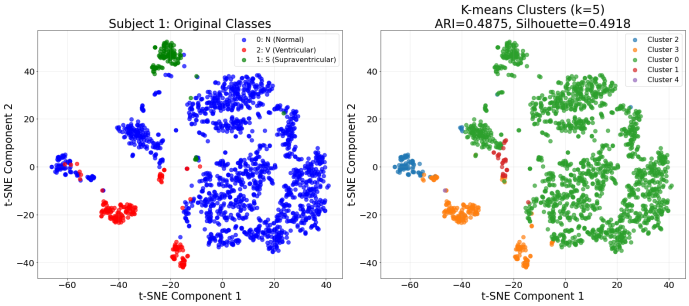

Fig. 9. t-SNE projection of ECG segments for subject 27 from the MITDB dataset.

D. Application to HiDB

In contrast to MITDB, multiple acquisitions in HiDB may refer to the same subject recorded at different times or with different device configurations. As a result, each acquisition exhibits unique characteristics related to electrode placement on the chest, as well as potential noise, artifacts, or ECG anomalies specific to the moment in which the recording was performed. Given the smaller number of segments per subject compared with MITDB, a stricter Silhouette threshold of 0.30 was adopted to ensure a more robust evaluation of clustering quality. Out of 88 acquisitions, 54 achieved a Silhouette score above the 0.30 threshold (mean 0.74 ± 0.13), while the remaining 34 consistently scored below it for all tested k, resulting in the formation of a single cluster. Figure 10 displays the Silhouette scores for the 54 HiDB acquisitions. This substantial number of single-cluster acquisitions could be due to the controlled experimental environment and the relatively short recording time, which limited the intrinsic process variability and the likelihood of capturing non-physiological beats. Additionally, excessively noisy leads were excluded from the analysis, thereby further reducing variability within subjects. As further confirmation, only three acquisitions exhibited more than two clusters.

Given the supervised nature of the study, it is important to apply indirect validation to support the interpretation of the clustering results. Rather than relying solely on the absolute value of the Silhouette score, it is crucial to consider it within the context of the HiDB dataset, using specific examples to illustrate relevant scenarios. In several cases, high Silhouette values were observed due to the presence of two clusters with

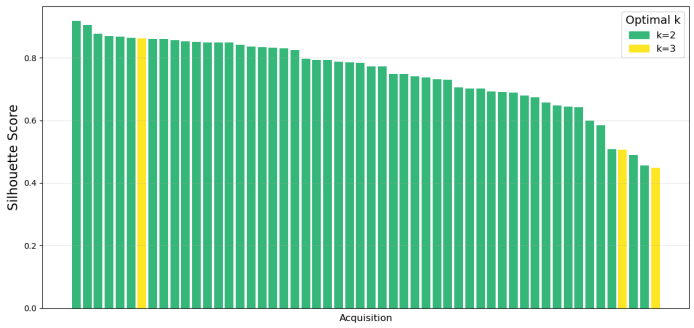

Fig. 10. Silhouette scores for the 54 HiDB acquisitions with values above the 0.30 threshold.

highly imbalanced sizes. Such imbalance may arise not only from the higher prevalence of normal beats in ECG signals, but also from movement artefacts or electrode detachment, which are typical of wearable settings. These issues can lead to abnormal morphologies that differ from the subject's dominant beat pattern. This is exemplified by Acquisition 63 (Silhouette $= 0.81$) from the HiDB dataset, as shown in Figure 11, which reports the median and interquartile range of the heartbeats associated with each cluster. Cluster 1 (in green) appears to contain noise, with beats lacking a recognisable morphology. Conversely, when multiple morphologies differ from the standard pattern, Silhouette values decrease. This is due to the more dispersed distribution of anomalous beats, which makes cluster boundaries less well defined. In the HiDB dataset, the morphologies that are most effectively identified by the clustering algorithm typically involve morphological and spectral features such as QRS complex and T wave duration, area, and power. This is illustrated by Acquisition 1 (Silhouette $= 0.74$) in Figure 11, who exhibits a widened QRS complex and an inverted T wave in the beats of Cluster 1 (in green). Rhythm-related features, such as the RR intervals preceding and following each beat, also contribute to beat differentiation. For example, Acquisition 27 (Silhouette $= 0.51$) from HiDB, as shown in Figure 11, shows in Cluster 0 (in orange) a set of beats lacking a visible P wave, highlighting a different rhythm pattern.

IV. CONCLUSIONS

Cardiovascular diseases remain a major health concern worldwide, highlighting the need for robust, early detection systems based on ECG signals. Given the significant burden required to label large ECG datasets, there is a need for automated algorithms that can identify anomalous patterns and clustering of heartbeats. In this article, we presented the creation of HiDB, a new ECG database acquired with the CGM Hi 3-Leads device.

The primary aim of HiDB is to support studies involving wearable devices designed to acquire ECG signals using a reduced number of leads, and to assess the feasibility of introducing automatic anomaly detection models to assist clinicians in the data labeling process. In line with this objective, we developed a framework that clusters ECG heartbeats in an

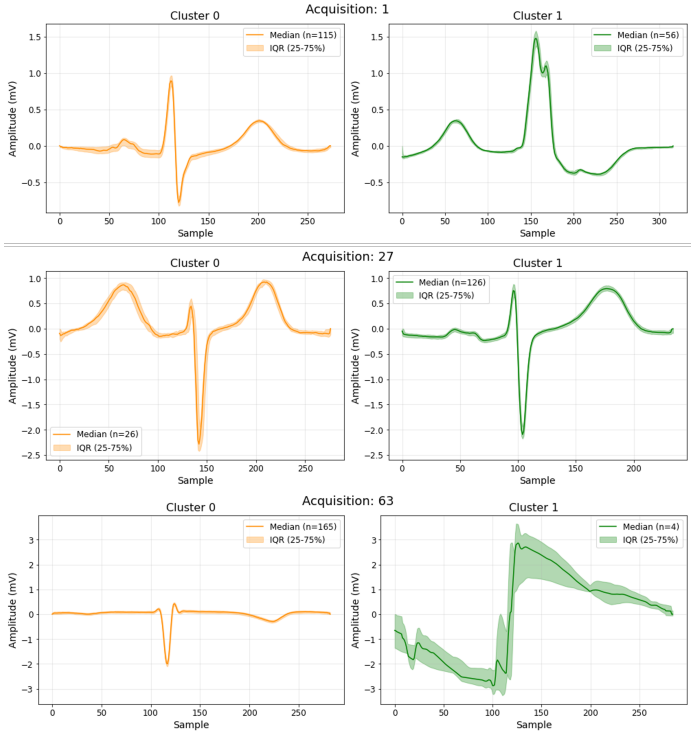

Fig. 11. Median and interquartile range of the two clusters identified in Acquisitions 1, 27, and 63 from the HiDB dataset.

intra-subject setting using a robust P-QRS-T wave classifier. This classifier achieved an overall accuracy of 88% on the LUDB dataset and demonstrated strong generalisation on the MITDB, AHADB, and NSTDB, correctly identifying 99.75%, 99.62%, and 94% of QRS complexes, respectively.

We then applied a supervised feature extraction pipeline driven by MI, highlighting the importance of domain-specific feature engineering: 17 of the 23 features selected were carefully designed morphological or spectral descriptors. An additional insight emerged from analysing wave presence. In fact, the absence of P waves and the number of T waves significantly discriminated among AAMI classes. In particular, classes V and Q contained 94.1% and 97.3% of segments classified as P-wave absent, respectively. In addition, a double T-wave pattern was observed in 47.8% of class V segments and 23.2% of class S segments.

Finally, a k-means algorithm with acquisition-specific k provided a first-level discrimination of beat morphologies, achieving a mean ARI of 0.78 ± 0.19 and a mean Silhouette score of 0.65 ± 0.14 for the MITDB dataset. This suggests that for the majority of subjects, the identified feature subset effectively discriminates among classes, particularly for the V class. This result can be explained by the subset of features identified during feature selection. Examples include the RR interval, QRS and T wave duration and area, and the corrected QT interval, which are all directly influenced by ventricular ectopic beats. However, the clustering model experiments challenges due to the morphological similarities between S and

N beats and the inconsistent morphology of Q beats, which tend to be uniformly distributed across the feature space. For the HiDB dataset, the approach yielded an average Silhouette score of 0.74 ± 0.13, indicating good cluster separability. In most recordings, it was possible to identify beats with distinct morphologies. This suggests that the proposed method may serve as a useful starting point to support clinicians during the labelling process, ultimately facilitating the integration of supervised models on the data acquired by the Hi-ECG device. For acquisitions that resulted in a single cluster, this is not necessarily indicative of clustering failure, but rather reflects the limited morphological variability captured within short ECG recordings in HiDB.

For this reason, we intend to extend the HiDB dataset by including additional subjects, longer recording durations, and a wider variety of subject activities and postures. Such expansions would provide more comprehensive data, supporting the development of more generalisable models. Moreover, the introduction of a quality assessment metric for each acquired lead in Hi-ECG could allow the integration of multichannel information, making analyses of individual heartbeats more robust. In line with this, incorporating accelerometry data could help to associate heartbeat morphologies with the subject's state of motion or posture. In addition, extending the pipeline to detect rhythm abnormalities by leveraging information from QRS complexes identified by the wave classifier would enable the detection of arrhythmic patterns occurring across multiple heartbeats. Finally, from a feature selection perspective, the development of subject-specific feature extraction pipelines would ensure optimal beat discrimination tailored to each individual. This would be coupled with a fusion of sample-level temporal information obtained from the wave classifier, combining them in a feature fusion approach to improve the discrimination of beats that currently present classification difficulties, such as supraventricular ectopic beats (S) and unknown beats (Q).

ACKNOWLEDGMENT

The Authors would like to thank the Cooperativa di Animazione Valdocco (CAV), Via Sondrio 13, Turin, Italy, for their valuable cooperation. The study was conducted in accordance with the Declaration of Helsinki; informed consent was obtained from all participants. The study protocol does not involve patients but only healthy volunteers. It was approved by the CAV and a convention among all the involved Researchers was properly signed (no. 1BD261). The data presented in this study are available for research purposes from the corresponding author upon request.

REFERENCES

[1] World Health Organization, "Cardiovascular diseases - who." Online. url: https://www.who.int/health-topics/cardiovascular-diseasestab=tab$_1$.

[2] World Health Organization, "Invisible numbers: The true extent of noncommunicable diseases and what to do about them," tech. rep., World Health Organization, Geneva, Switzerland, 2022.

[3] E. S. Dahiya, A. M. Kalra, A. Lowe, and G. Anand, "Wearable technology for monitoring electrocardiograms (ecgs) in adults: a scoping review," *Sensors*, vol. 24, no. 4, p. 1318, 2024.

[4] H. Ozkan, O. Ozhan, Y. Karadana, M. Gulcu, S. Macit, and F. Husain, "A portable wearable tele-ecg monitoring system," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 1, pp. 173–182, 2019.

[5] Z. Bouzid, S. S. Al-Zaiti, R. Bond, and E. Sejdić, "Remote and wearable ecg devices with diagnostic abilities in adults: a state-of-the-science scoping review," *Heart Rhythm*, vol. 19, no. 7, pp. 1192–1201, 2022.

[6] N. Wang, J. Zhou, G. Dai, J. Huang, and Y. Xie, "Energy-efficient intelligent ecg monitoring for wearable devices," *IEEE transactions on biomedical circuits and systems*, vol. 13, no. 5, pp. 1112–1121, 2019.

[7] H. Li and P. Boulanger, "A survey of heart anomaly detection using ambulatory electrocardiogram (ecg)," *Sensors*, vol. 20, no. 5, p. 1461, 2020.

[8] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer methods and programs in biomedicine*, vol. 127, pp. 144–164, 2016.

[9] K. Nezamabadi, N. Sardaripour, B. Haghi, and M. Forouzanfar, "Unsupervised ecg analysis: A review," *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 208–224, 2022.

[10] A. Kalyakulina, I. Yusipov, V. Moskalenko, A. Nikolskiy, K. Kosonogov, N. Zolotykh, and M. Ivanchenko, "Lobachevsky university electrocardiography database," *Type: Dataset. Available online: https://physionet.org/content/ludb/1.0. 0(accessed on 10 July 2021)*, 2020.

[11] M. Barandas, D. Folgado, L. Fernandes, S. Santos, M. Abreu, P. Bota, H. Liu, T. Schultz, and H. Gamboa, "Tsfel: Time series feature extraction library," *SoftwareX*, vol. 11, p. 100456, 2020.

[12] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[13] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[14] A. AAMI and A. EC57, "Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms," *American National Standards Institute, Arlington, VA, USA*, p. 43, 2008.

[15] Centro Aiuto alla Vita La Valdocco. Identità - Cooperativa Animazione Valdocco Onlus, "Identità - cooperativa animazione valdocco onlus." Online, 2024. Accessed: 2024-06-12, url: https://cav.lavaldocco.it/identita.html (cit. on p. 48).

[16] CompuGroup Medical, "Cgm hi 3 leads ecg." Online. url: https://www.cgm.com/ita$_i t/telemedicina/cgm-hi-3-leads-ecg-en.html$.

[17] FIAB S.p.A., "F9089 disposable ecg electrodes – technical data sheet." Online, 2024. Accessed: 2024-06-12, url: https://www.elettromedicali.it/wp-content/plugins/ant-webcommerce/global//files/ECGXF9089.pdf?x89244=.

[18] A. Paszke, "Pytorch: An imperative style, high-performance deep learning library," *arXiv preprint arXiv:1912.01703*, 2019.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[20] N. Zhang, R. A. R. Melendez, J. Butler, W. H. Mahmoud, and L. A. Thompson, "Ecg waveform segmentation for p-qrs-t detection using deep learning based on fourier synchrosqueezed transform," in *ASEE Mid-Atlantic Section Spring Conference*, 2024.

[21] E. Foster, "Aha database.," *Respiratory Care*, vol. 33, no. 10, pp. 984–984, 1988.

[22] G. B. Moody, W. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," *Computers in cardiology*, vol. 11, no. 3, pp. 381–384, 1984.

[23] M. U. Zahid, S. Kiranyaz, T. Ince, O. C. Devecioglu, M. E. Chowdhury, A. Khandakar, A. Tahir, and M. Gabbouj, "Robust r-peak detection in low-quality holter ecgs using 1d convolutional neural network," *IEEE Transactions on Biomedical Engineering*, vol. 69, no. 1, pp. 119–128, 2021.

[24] W. Zhao, Z. Li, J. Hu, and Y. Ma, "A simple and effective deep neural network based qrs complex detection method on ecg signal," *Frontiers in Physiology*, vol. 15, p. 1384356, 2024.