applied sciences

MDPI

Review

Towards Reliable ECG Analysis: Addressing Validation Gaps in the Electrocardiographic R-Peak Detection

Syed Talha Abid Ali [1,2,†], Sebin Kim [1,†] and Young-Joon Kim [1,2,*]

[1] Department of Electronic Engineering, Gachon University, Seongnam 13120, Republic of Korea; talha15@gachon.ac.kr (S.T.A.A.); sbscsue@gmail.com (S.K.)
[2] Department of Semiconductor Engineering, Gachon University, Seongnam 13120, Republic of Korea
* Correspondence: youngkim@gachon.ac.kr
† These authors contributed equally to this work.

Abstract: Electrocardiographic (ECG) R-peak detection is essential for every sensor-based cardiovascular health monitoring system. To validate R-peak detectors, comparing the predicted results with reference annotations is crucial. This comparison is typically performed using tools provided by the waveform database (WFDB) or custom methods. However, many studies fail to provide detailed information on the validation process. The literature also highlights inconsistencies in reporting window size, a crucial parameter used to compare predictions with expert annotations to distinguish false peaks from the true R-peak. Additionally, there is also a need for uniformity in reporting the total number of beats for individual or collective records of the widely used MIT-BIH arrhythmia database. Thus, we aim to review validation methods of various R-peak detection methodologies before their implementation in real time. This review discusses the impact of non-beat annotations when using a custom validation method, allowable window tolerance, the effects of window size deviations, and implications of varying numbers of beats and skipping segments on ECG testing, providing a comprehensive guide for researchers. Addressing these validation gaps is critical as they can significantly affect validatory outcomes. Finally, the conclusion section proposes a structured concept as a future approach, a guide to integrate WFDB R-peak validation tools for testing any QRS annotated ECG database. Overall, this review underscores the importance of complete transparency in reporting testing procedures, which prevents misleading assessments of R-peak detection algorithms and enables fair methodological comparison.

Keywords: arrhythmia; MIT-BIH dataset; R-peak detection; WFDB; R-peak validation

check for updates

Citation: Ali, S.T.A.; Kim, S.; Kim, Y.-J. Towards Reliable ECG Analysis: Addressing Validation Gaps in the Electrocardiographic R-Peak Detection. *Appl. Sci.* **2024**, *14*, 10078. https://doi.org/10.3390/app142110078

Academic Editors: Žydrūnas Kavaliauskas and Vitas Valincius

Received: 29 August 2024
Revised: 25 October 2024
Accepted: 29 October 2024
Published: 4 November 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cardiac health is one of the most critical issues in modern times. Every year, millions of people die from various cardiovascular diseases (CVDs) [1]. Early diagnosis of CVDs through electrocardiography (ECG) can significantly reduce mortality rates, as timely intervention allows more effective cardiological treatments [2]. Therefore, numerous studies are conducted, and many sensor-based ECG processing devices are proposed. The accuracy of these electronic devices depends on the precision of the R-peak detectors. The detection is usually performed using real-time ECG signal acquisition. The process typically involves monitoring and measuring heart activity for a certain period by placing electrodes in different combinations at specific points on the human body [3] and then applying a peak detection algorithm to it. However, this is only possible if the algorithm is previously validated on publicly available ECG datasets.

The publicly available ECG datasets can be categorized into two groups. The first group consists of general datasets that cover a wide range of clinical real-life conditions, including different ECG noises and various heart arrhythmic anomalies. Examples of these include the MIT-BIH arrhythmia dataset [4], QT database (QTDB) [5], Physikalisch-Technische Bundesanstalt eXtended (PTB-XL) diagnostic database [6], and St. Petersburg

Institute of Cardiological Technics (INCART) database [7]. The second group consists of specific datasets dedicated to addressing certain noisy conditions or a particular arrhythmic heart anomalies, such as the MIT-BIH Noise Stress Test database (NSTDB) [8], European ST-T database [9], and Long-Term Atrial Fibrillation database [10]. The MIT-BIH arrhythmia database is the benchmark for evaluating R-peak detectors before applying methodologies to real-time sensor-based ECG devices [11,12]. This is due to its historical significance as the first publicly available dataset designed for arrhythmia analysis, featuring diverse cardiac abnormalities, clinical noise, and artifacts. It is a widely adopted dataset by researchers for comparison with earlier studies. This dataset is also well-organized and comes with dedicated tools to simplify its processing.

Each recording in these datasets consists of multiple ECG signals, including components like the P-wave, QRS complex, and T-wave, which offer valuable insights into the heart's electrical activity and overall health condition [13]. CVDs, especially arrhythmic pathologies, are analyzed by detecting R-peaks within QRS complexes [14,15]. R-peaks are important for arrhythmia detection and analysis for several reasons. First, they are the gold standard for calculating heart rate, measured by the time interval between consecutive R-peaks (RR intervals) [16,17]. Since arrhythmias often involve irregular heart rhythms, detecting R-peaks enables precise measurement of these fluctuations, whether too fast or too slow [18]. Second, R-peaks mark ventricular depolarization, a key phase of the cardiac cycle. Their timing and shape, in relation to other ECG components (P-wave, T-wave), help diagnose arrhythmic pathologies and assess their effect on heart function [19]. Third, R-peaks are crucial for arrhythmia classification, as specific RR interval patterns are key to distinguishing different arrhythmias [20]. For example, atrial fibrillation shows irregular RR intervals, while ventricular tachycardia presents rapid, repetitive R-peaks [21,22]. Therefore, accurate R-peak detection is vital for properly classifying these conditions. Once these R-peaks are detected while developing a detection algorithm, the next step is to validate them by comparing the results with reference annotations using the window concept. This validation can be performed using tools provided by the waveform database (WFDB) or by developing a custom testing methodology. WFDB tools, version 0.10.0 include beat-by-beat (bxb) [23], run-by-run (rxr), measurement-by-measurement (mxm), event processing interface comparator (epic), and annotation to RR interval (ann2rr) [24,25] comparators, all of which comply with the American National Standards Institute (ANSI)/Association for the Advancement of Medical Instrumentation (AAMI) protocols for testing ECG-derived results [26]. Despite the availability of these standardized tools, our review identifies several gaps and inconsistencies in the literature as shown in Figure 1, regarding the validation of predicted R-peaks.

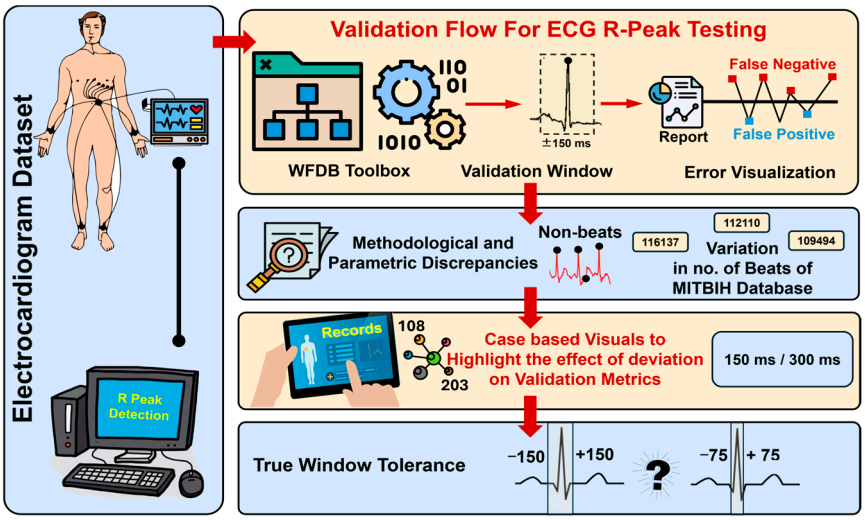


Figure 1. Graphical representation of various discrepancies found in the literature.

Many studies fail to specify tools or custom validation methods used, and they often lack detailed descriptions of their validation processes. While some studies mention the use of a WFDB comparator, they do not provide sufficient details about its integration or the parameters involved [24,25,27–30]. Additionally, several studies listed in Table 1 overlooked the presence of annotated non-beats in the MIT-BIH arrhythmia dataset and failed to explain how these beats were filtered when using custom validation methods. Discrepancies are also found in the selection of validation window size, with some studies deviating from the AAMI-sanctioned standard [28,31–34]. Although several studies have reported low false positive (FP) and false negative (FN) metrics, they failed to specify the employed window size [35–39]. Similarly, inconsistencies in reporting allowable window size tolerance [30] and variations in reported total number of beats for individual or complete MIT-BIH arrhythmia records were also identified [36,40–44]. Furthermore, many studies validate algorithms on selected segments of individual records without specifying timestamps or providing a rationale for it [45–48]. These gaps introduce uncertainty and confusion regarding the reliability of the reported statistical outcomes, affecting methodological comparability due to apparently misleading assessments of the detection algorithm's performance. Additionally, ambiguity in usage and integration of WFDB validation tools is another issue when reviewing the PhysioNet database and available literature. For instance, installation process and backend framework of the WFDB toolbox was extensively discussed in ref. [49]. However, their manuscript lacks a practical guide for validation variants to simultaneously test the entire dataset. On the other hand, implementation details using bxb for individual record testing of the MIT-BIH database is mentioned in ref. [31]. However, they did not explain all the parameters involved in the R-peak validation process. This underscores the need for a narrative review that thoroughly details validation methodologies for R-peak detectors, addresses the aforementioned discrepancies, and explains their effect on R-peak validation statistics.

Table 1. Provided ECG testing details in literature for evaluating MIT-BIH arrhythmia dataset.

Literature	R-Peak Validation	MIT-BIH Beats	Testing Format	Window Size (ms)	Statistical Metrics			
					FP	FN	Se (%)	+P (%)
Mahmoodabadi et al. [24]	Bxb	104,988	Selective	NG	NG	NG	99.18	98.00
Mahmoodabadi et al. [25]	Bxb	104,988	Selective	NG	NG	NG	99.18	98.00
Doyen et al. [28]	Bxb	109,494	Entire Record	50	3195	NG	87.48	89.39
Ledezma et al. [29]	Bxb	91,285	Entire Record	150	295	267	NG	NG
Bachi et al. [30]	Bxb	109,494	Abnormal Beats	300	NG	NG	94.81	NG
Zhu et al. [31]	Bxb	109,401	Entire Record	100	91	93	99.92	99.92
Qin et al. [32]	NG	109,966	NG	50	561	668	99.39	99.49
Pandit et al. [33]	NG	109,809	Entire Record	80	369	389	99.65	99.66
Kaur et al. [34]	NG	109,498	Selective Records	100	55	54	99.95	99.94
Gupta et al. [35]	NG	109,494	NG	NG	39	49	99.96	99.96
Park et al. [36]	NG	109,494	NG	NG	99	79	99.93	99.91
Manikandan et al. [37]	NG	109,496	NG	NG	53	76	99.93	99.86
Kaur et al. [38]	NG	109,494	Segment	NG	76	53	99.93	99.95
Rakshit et al. [39]	NG	109,474	NG	NG	116	58	99.95	99.88
Afonso et al. [40]	NG	91,314	Selective	NG	406	374	99.59	99.56
Tompkins et al. [41]	NG	116,137	NG	NG	507	277	99.80	99.76
Varghese et al. [42]	NG	109,021	NG	NG	195	381	99.65	99.82
Yeh et al. [43]	NG	116,137	NG	NG	58	166	99.95	99.85
Chen et al. [44]	NG	109,494	Entire Record	150	63	124	99.89	99.97
Tang at al. [45]	NG	109,966	Entire Record	NG	494	911	99.17	99.55
Ravanshad et al. [46]	NG	109,428	Entire Record	NG	1216	651	98.89	99.44
Elgendi et al. [47]	NG	109,985	NG	NG	82	247	99.78	99.92
Mondelo et al. [50]	wxw	NG	Selective	NG	NG	NG	NG	NG
Abdullah et al. [51]	NG	109,494	Entire Record	NG	NG	59	NG	NG
Modak et al. [52]	NG	109,494	Entire Record	NG	136	200	99.82	99.88
Rahul et al. [53]	NG	109,494	NG	NG	155	193	99.82	99.85
Saadi et al. [54]	NG	91,285	NG	NG	NG	NG	99.90	99.87
Hammad et al. [55]	NG	NG	Segment	NG	NG	NG	99.98	100
Li et al. [56]	NG	109,497	Entire Record	NG	138	67	99.94	99.87

Table 1. *Cont.*

Literature	R-Peak Validation	MIT-BIH Beats	Testing Format	Window Size (ms)	Statistical Metrics			
					FP	FN	Se (%)	+P (%)
Burguera et al. [57]	NG	109,985	Entire Record	NG	NG	NG	97.93	98.84
Dohare et al. [58]	NG	109,966	Entire Record	NG	728	870	99.21	99.34
Banerjee et al. [59]	NG	19,098	Selective Records	NG	40	76	99.66	99.55
Pal et al. [60]	NG	45,936	PVC, BBB	NG	17	54	99.98	99.96
Ning et al. [61]	NG	109,452	Segment	NG	127	138	99.87	99.88
Gutiérrez-Rivas et al. [62]	NG	109,949	Entire Record	NG	289	502	99.54	99.73
Hossain et al. [63]	NG	109,441	Segment	NG	122	46	99.97	99.93
Yazdani et al. [64]	NG	109,494	Entire Record	NG	108	137	99.87	99.99
Sabor et al. [65]	NG	109,494	Entire Record	NG	104	126	99.89	99.91
Nayak et al. [66]	NG	109,494	Entire Record	NG	70	52	99.95	99.94
Hamilton et al. [67]	NG	109,267	Selective	NG	248	340	99.68	99.77
Zhang et al. [68]	NG	109,510	Entire Record	NG	204	213	99.81	99.80
Zidelmal et al. [69]	NG	108,494	Selective	NG	97	171	99.84	99.91
Yochum et al. [70]	NG	109,491	Entire Record	NG	574	160	99.85	99.85

This table examines the details of ECG testing in the literature, considering which R-peak validation tool was utilized for ECG testing. It also explores testing formats, whether all records were tested collectively or individually with sample exclusion, adherence to the AAMI standard for window size if applied, and the statistical metrics. It also shows the variation in the reported number of total beats for the MIT-BIH arrhythmia dataset. BBB = Bundle block branch beat, NG = Not given in the literature, PVC = Premature ventricular contraction.

This review is structured as follows: Section 2 highlights literature search methods to provide a comprehensive overview of the search and evaluation strategy for the studies included in this narrative review. Section 3 covers available R-peak validation tools and their testing mechanism. Section 4 discusses the MIT-BIH dataset, window validation concept, identified discrepancies in the literature, their impact on statistical outcomes, and tolerance variations within a 0.15 s window. Section 5 addresses non-beat annotations and studies that fail to specify validation methods. Section 6 highlights numeric discrepancies in the literature regarding the individual or total number of beats for the MIT-BIH records and issues related to selective segment testing. Section 7 introduces a structured concept as a future prospect to provide a guide for integrating WFDB tools to simultaneously test, generate validation statistics and errors reports, and visualize FP and FN. Finally, Section 8 presents the conclusions of the review.

2. Literature Search Methods

We perform a narrative review to highlight validation gaps in electrocardiographic R-peak detection. Information on validation studies of R-peak detection algorithms and recognized databases was analyzed to identify discrepancies by examining commonalities and differences among the included studies. The review highlights discrepancies, key conclusions, and future prospects.

The scientific literature search was performed using several databases (Scopus, Web of Science). Specific terms such as ECG R-peak detection, ECG validation techniques, MIT-BIH arrhythmia, WFDB tools, ECG window size, ECG window tolerance, false positive ECG, false negative ECG, ECG statistics, and ECG validation tools were used in the search query. The search was not restricted by a particular time period but was limited to English-language articles. To determine if a study met the inclusion criteria for this review, two reviewers independently screened each record, and a third reviewer double-checked the results for accuracy and fairness. No automation tools were used in the screening process.

The search provides a substantial number of peer-reviewed research articles and conference papers relevant to R-peak detection validation. After removing duplicates and excluding papers that did not meet the eligibility criteria, the remaining articles were further processed. Each study was analyzed for R-peak validation method, beat numbers, testing format, window size, and statistical metrics, and the findings were tabulated to ensure key aspects of ECG validation methods were addressed. The gathered information was then classified from different perspectives: adopted validation methodologies (custom methods),

use of WFDB tools to validate R-peaks, mention of window size parameters, numeric discrepancies in beat reporting, and methodologies that reported only validation metrics.

3. Tools for R-Peak Validation

Accurate validation of detected R-peaks with reference annotations is a crucial step for developing a reliable detection algorithm. However, the literature usually does not provide detailed descriptions on the contextual variables for the employed validation methodologies, making it difficult to replicate the exact conditions for testing and conducting a fair comparison of peak detection results. To address this issue, standard systems and benchmarks were established by ANSI/AAMI to ensure transparency and consistent testing conditions. Various tools have been established to facilitate the R-peak validation process, many of which are part of the waveform database (WFDB) software package, version 0.10.0 available on PhysioNet [23]. Below, we discuss these tools and their testing mechanisms.

Beat-by-beat (bxb): The bxb tool is a utility comparator of the WFDB software package available on PhysioNet [23]. This tool complies with the EC57-2012 standards for ECG R-peak validation [71,72]. The testing mechanism of bxb starts with taking a predicted sample file generated by a custom peak detection algorithm and creates corresponding annotations using the writing annotation (WRANN) function. After running a peak detection algorithm on an ECG signal, a list of predicted R-peak locations, typically represented as time instances or sample numbers, is obtained. These predictions are formatted into a text file, with each line containing the sample location for the detected R-peak. WRANN reads this input file and generates a corresponding predicted annotation file in the WFDB format, similar to the provided reference atr files with clinical annotation or beat types, aligning and labeling the predicted sample or time location with a user-defined symbol [73]. This symbol can be any letter and does not need to follow the AAMI beat notations, as the focus is solely on detection validation not classification, meaning only the predicted time locations are required. This predicted annotation file can then be used with WFDB validation tools to quantitatively assess the performance of the detection algorithm in comparison with reference annotations using a comparison window size of 0.15 s or 150 ms [74]. To handle non-beats or segments of the ECG without real annotated R-peaks, bxb uses a shutdown and resume function. It pauses the comparison process during noisy segments or annotated non-beats and resumes when those segments end, ensuring that noise and non-beat annotations do not affect validation statistics [75]. The bxb program generates various performance reports, including summaries of annotation discrepancies and statistics on RR interval errors, false positives, and missed beats. It also offers options to generate condensed reports, verbose output for mismatches, and labeled error files, allowing developers to thoroughly evaluate their detection algorithms.

Run-by-run (rxr): The rxr tool is another WFDB comparator that aligns and compares annotation files from reference and predicted samples within a 0.15 s window [76]. This tool was primarily designed to compare RR intervals but can be used to evaluate the accuracy of automated algorithms for detecting R-waves. It mainly reports statistics such as mean error and standard deviation. For segments other than actual beats, it uses a similar concept of shutdown and resume like bxb.

Measurement-by-measurement (mxm): The mxm tool is another validation tool of WFDB, used to validate heart rate measurements from multichannel ECG. It starts comparison after the first 5 min of a given recording to allow the signal to stabilize and minimize the impact of initial artifacts and noise [77]. Marsili et al. applied a similar concept with bxb to validate atrial fibrillation [27]. For each test measurement, mxm calculates the error relative to the closest reference measurement, reporting normalized or unnormalized root mean square (RMS) errors.

Event processing interface comparator (epic): The epic tool compares episodes of arrhythmias and ischemic events, such as ventricular or atrial fibrillation, flutters, and ischemic ST episodes, assigning weight on the basis of episode length or duration. The tool then uses

weights in comparison to reference annotated markings to assess the overlap, which must be at least 50% [78].

Annotation to RR interval (aa2rr): Another tool, ann2rr, can also validate peak predictions [79]. It primarily extracts RR intervals by converting annotations into interval lists and comparing them with reference markings. However, other options are available for customization, such as specifying time intervals to analyze, filtering event type, or handling specific formats. The tool is useful in analyzing heart rate variability or detecting irregularities in heartbeat timing.

These are some of the validation tools provided by WFDB based on the ECG testing guidelines by ANSI/AAMI, which can be modified and incorporated into detection methodologies to verify predicted R-peaks. Some authors have also proposed their methods to make the validation process more effective. An example of this is wave-by-wave (wxw), a QRS delineation percentage comparator proposed by [50]. It distributes the ECG into its components like P-wave, QRS-wave, and T-wave and then compares them with expert annotations using specific scoring rules. It requires precise determination of wave positions, including onset and offset, to improve diagnostic accuracy.

Compared to the other tools, bxb is more often used for ECG R-peak validation because it was specifically designed to compare predicted and reference annotation markings beat-by-beat [80], as shown in Figure 2. However, due to scattered and unclear information regarding these tools within the PhysioNet database, and lack of comprehensive guidance in the available literature, a detailed guide is needed to integrate WFDB tools and implement additional validation pipelines (series of steps or processes). These pipelines include tool integration, simultaneous testing of all records, generation of validation statistics, error reports, extraction of FP and FN, and their accurate visualization in the reference signal [23].

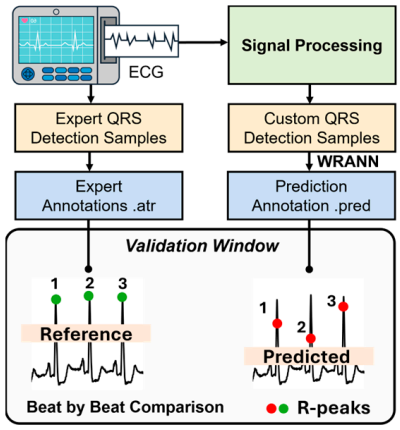


Figure 2. Process Flow of the bxb Annotation Comparator.

4. WFDB Window-Based Evaluation

The MIT-BIH arrythmia database is a publicly available database from the Massachusetts Institute of Technology–Beth Israel Hospital (MIT-BIH) comprising ECG recordings with supervised expert annotations. This dataset consists of 48 dual-channeled recordings, each 30 min long, sampled at 360 Hz with variables like artifacts, wave shapes, morphologically altered beat types, and extreme noises. Each recording is accompanied by a reference annotation (atr) file created by cardiovascular experts and used as the gold standard to evaluate R-peak detectors. The annotation consists of two key aspects for each ECG signal: the precise marking of the R-peak location and the labeling of any related arrhythmic anomalies, such as left or right bundle branch block beats, atrial premature beats, premature ventricular contractions, ventricular escape beats, and others. These anomalies are labeled according to the categories established by the ANSI/AAMI [4]. Each label consists of a single letter, such as “L” for left bundle branch block beat, “R” for right bundle branch block beat, “A” for atrial premature beat, “V” for premature ventricular

contraction, and "e" for ventricular escape beat. These labels or beat types can then be used for classification purposes after detection, as they help in classifying key events and particular shapes and features associated with different cardiovascular conditions in the physiological signals.

The primary purpose of R-peak detectors is to predict R-peak time location. The window concept is essential for evaluating time location and is considered the foundation of all WFDB tools. It plays a critical role in validating the robustness of every R-peak detection methodology. It allows for the comparison of WRANN-generated annotations (based on predicted peaks) with manually edited reference annotations marked by a cardiologist, using the benchmark window size of 0.15 s [81]. This helps distinguish false peaks from true R-peaks. Using this standard, two R-waves are considered matching if their labels are, at most, 0.15 s apart [82]. This precise comparison ensures not only the accurate diagnosis of arrhythmia but also can be used for applications like heart rate variability (HRV) [16], fetal heart rate monitoring (fHRM) [83], and biometric recognition [84].

4.1. SUMSTAT Measured Statistical Metrics

The summary statistics (SUMSTAT) metrics are the derived aggregate statistics from bxb, rxr, etc. using the comparable window concept that can be categorized mainly into three groups: true positives (TPs), false positives (FPs), and false negatives (FNs). The SUMSTAT function counts each of these statistical parameter for every record to generate a report file [32,51,52,85]. These parameters are defined as follows:

TP: If the predicted sample falls within the 150 ms window.

FP: If the predicted sample falls outside the 150 ms window.

FN: If the predicted sample does not fall within the 150 ms window.

Figure 3 illustrates the basic working of the window and how metrics like TP, FP, and FN are measured. We can also use other metrics to evaluate the detector based on TP, FP, and FN. These metrics are positive predictability (+P), accuracy (Acc), and the detection error rate (DER) [36,86,87]. Each of these evaluation metrics is defined as follows:

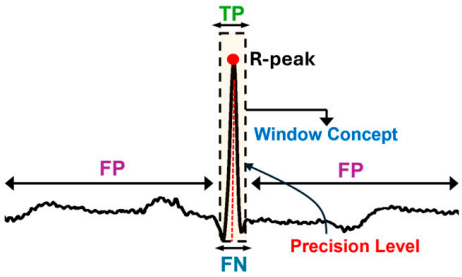


Figure 3. Calculation of Window-based Statistical Metrics.

Positive predictability: It quantifies the algorithm's ability to differentiate true beats and false beats, and it is defined as

$$+\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \times 100\% \tag{1}$$

Sensitivity: It measures the algorithm's ability to detect true beats accurately, and it is defined as

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\% \tag{2}$$

Accuracy: It represents the overall correctness of the algorithm, and it is defined as

$$\mathrm{Acc} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \times 100\% \tag{3}$$

Detection error rate (DER): It helps evaluate the detection algorithm's performance. It is important to note that a lower DER indicates better performance. It is defined as

$$\text{DER} = \frac{\text{FP} + \text{FN}}{\text{TP}} \times 100\% \quad (4)$$

4.2. Discrepancies in the Selection of Window Size

The benchmark window size used for R-peak validation is 150 ms, as matching with the exact location of the R-peak defined by the cardiologist is not essential as long as the predicted sample falls within the QRS or, more precisely, the start and end of the R-wave [88]. The QRS complex represents ventricular depolarization, and the R-peak is the highest standard point in ECG. Given the variability in the signal and adopted methodologies, slight differences in timing are expected. So, a window of 150 ms is benchmarked by ANSI/AAMI to cover the entire R-wave to ensure the accuracy of predicted R-peak validation without compromising the overall clinical interpretation. This 150 ms window defines the maximum absolute difference in annotations that can be considered a match. Suppose a predicted and a reference annotation coincide within this span. In that case, the predicted beat is classified as a legitimate peak, similar to those defined by medical experts [30,81,86,89]. AAMI defines this standard to maintain consistent testing protocols across the numerous methodologies proposed for R-peak detection. Reviewing the literature, we found manuscripts that employ different window sizes or the same window size of 150 ms without providing a reason for deviation from the standard. These manuscripts can be categorized as follows:

- Doyen et al. [28] and Pandit et al. [33] mentioned window sizes of 50 ms and 80 ms, respectively. Zidelma et al. [90] and Kaur et al. [34] both mentioned a window size of 100 ms, smaller than the defined AAMI benchmark.
- Ledezma et al. [29] and Chen et al. [44] mentioned a window size of 150 ms.
- Bachi et al. [30] and Chen et al. [91] reported a window size of 300 ms, but they did not specify whether this refers to a fixed window size of 0.30 s, a user-defined parametric value for comparing predicted results to reference annotations, or simply the total of $\pm$150 ms tolerance (+150 ms and $-$150 ms) which is a benchmark window size of 0.15 s. If they set the value of 0.30 s for comparison, this would represent a significant deviation from the AAMI benchmark for comparing R-peak detection results as the total becomes 600 ms. Similarly, Rincón et al. also reported a time window of 320 ms [92] with similar ambiguity in their reporting.

These variations in comparable window size can lead to inconsistent evaluations and misinterpretations of the detector's effectiveness. To understand this, consider the case in Figure 4a, from record 108 of the MIT-BIH database; suppose the R-peak detector picks a non-peak and classifies it as an FP using the standard AAMI benchmark of a 150 ms window. However, increasing the window size can transform this non-peak into a TP, as shown in Figure 4a. Similarly, consider the case of record 203, as shown in Figure 4b. The peaks in this record have obscured amplitudes, leading to more missed beats or FNs. Suppose a non-peak is incorrectly picked by the detector initially, increasing the FP count during validation. However, if the window size is widened to 300 ms for retesting, the same FP would be reclassified as a TP, reducing the final FP and FN counts. These cases from the MIT-BIH database illustrate how variations in window size can significantly impact the statistical outcomes. Chauhan et al. [93] criticized the literature for using large window sizes up to 320 ms, which exceed the maximum allowable tolerance of 150 ms. They emphasized the need for greater transparency and adherence to standardized evaluation protocols in ECG R-peak validation.

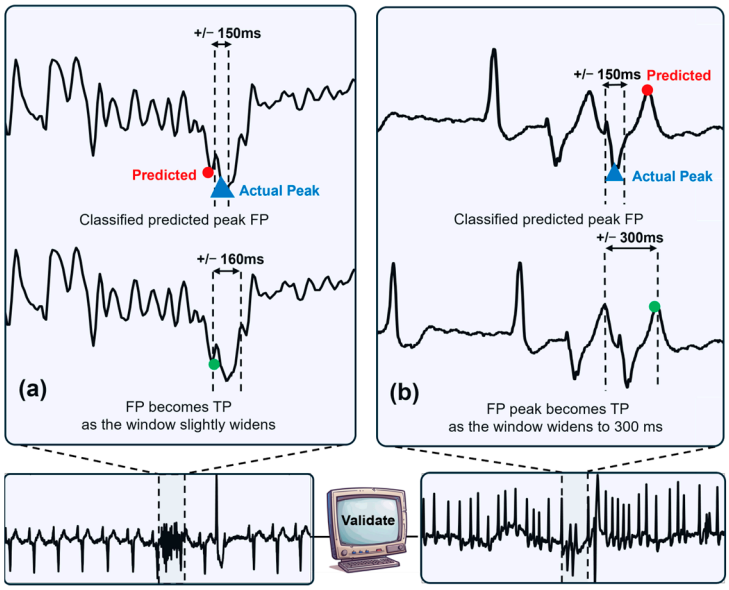


Figure 4. Case-based analysis of window size and its effect on FPs/FNs. (**a**) presents case study 1 featuring record 108 from the MIT-BIH dataset. (**b**) displays case study 2 with record 203 from the MIT-BIH arrhythmia dataset. These cases are included to demonstrate how altering window size can affect FP and FN counts.

4.3. Allowable Window Tolerance

Another issue observed in the literature is the ambiguity of the acceptable window tolerance. Online WFDB documentations mention a window of 150 ms or 0.15 s without specifying the acceptable ± tolerance range, whether it is ±75 ms [44] or ±150 ms [30], as shown in Figure 5a. There is further confusion about whether the WFDB R-peak validation tools use exact ±75, ±150 sample values or any other sample limits. Online resources typically indicate a window size of 0.15 s, but R-peak validation is usually conducted on samples. Therefore, the time unit must be converted first into samples as

$$\text{Sample Size} = 0.15 \times F_s \tag{5}$$

where Fs is the sampling frequency of the validated dataset. For the MIT-BIH arrhythmia database with sampling frequency of 360 Hz, a ±150 ms window corresponds to ±54 samples, and a ±75 ms window corresponds to ±27 samples, as shown in Figure 5b.

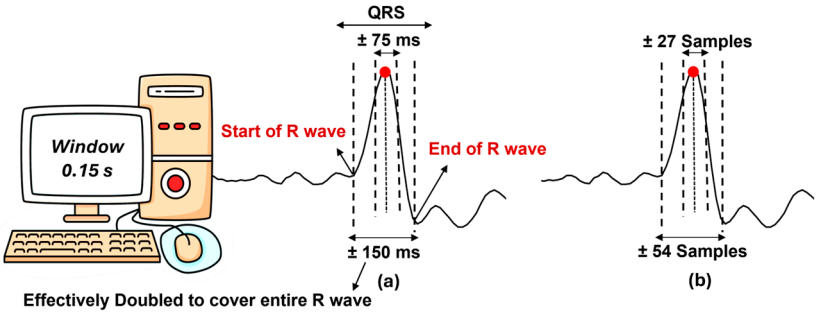


Figure 5. Acceptable Window Sample Tolerance. (**a**) Window tolerance reported in literature and (**b**) window tolerance in samples.

The actual allowable sample limit for the R-peak validation window is ±54 samples or ±150 ms. To confirm this acceptable tolerance range, one can take the following steps:

1: Choose a single reference sample value, for example, 550.

2: Calculate the predicted sample value by adding or subtracting 28 samples from the reference sample value. In this case, subtract 28 samples from 550 to obtain the predicted sample of 522.

3: When validating using WFDB tools with the AAMI benchmark window size of 0.15 s, it doubles the window size to cover the entire R-wave by uniformly distributing the total size of 0.30 s on either side of the reference annotation.

4: During validation, check for FNs and FPs. If they are reported, the allowable tolerance for the benchmark of 150 ms is ±75 ms or ±27 samples, as shown in Figure 6, because our predicted sample exceeds the window size by ±1 sample. If FNs and FPs are not reported, then the allowable tolerance would be ±54 samples for a 150 ms window, which is equivalent to a ±150 millisecond tolerance.

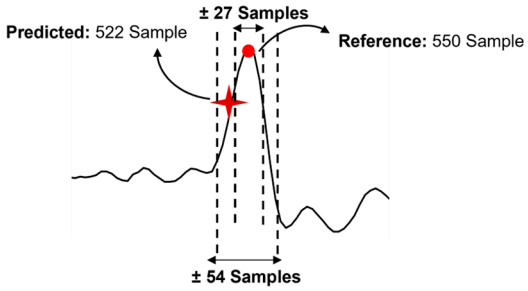


Figure 6. Validation of the actual tolerance for the AAMI benchmark window of 0.15 s.

This proves that whenever a WFDB R-peak validation tool is used with a window size of 0.15 s (or 150 ms), it employs a tolerance of ±54 samples or ±150 ms before and after the reference annotation, effectively doubling the window size to 0.3 ms to check if the predicted sample is within or outside this evenly distributed window size. The reason for ±tolerance is because the prediction can be either side of the reference R-peak time location. While WFDB validation tools handle this automatically, custom validation methodologies require implementation of this mechanism, rather than simply defining the window size of 0.15 s. Therefore, it is important to specify the tolerance used, whether ±75 ms or ±150 ms.

5. Non-Beats and Custom R-Peak Validation Method

The MIT-BIH arrhythmia database is classified into two main annotation categories: beats with actual R-peaks, which have 15 subtypes and consist of 109,494 beat annotations, and non-beats without QRS or R-peaks, which have 24 additional subtypes consisting of 3153 non-beat annotations including 42 beats of ventricular flutter from record 207 [94]. We refer to them as non-beats because they do not represent QRS-waves, making them irrelevant for R-peak detection as shown in Figure 7.

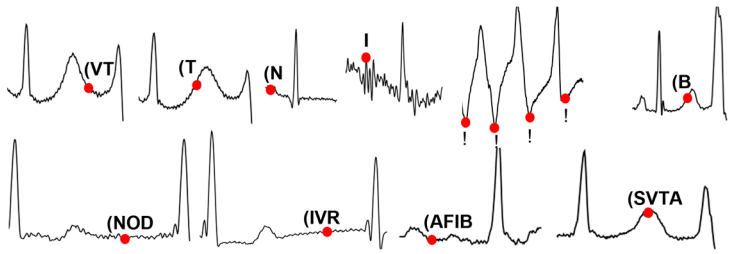


Figure 7. Some of the Non-beats in MIT-BIH Arrhythmia Dataset. VT: Ventricular tachycardia, T: Ventricular trigeminy, N: Normal sinus rhythm, |: Isolated QRS-like artifact, !: Ventricular flutter, B: Ventricular bigeminy, NOD: Nodal (A-V junctional) rhythm, IVR: Idioventricular rhythm, AFIB: Atrial fibrillation, and SVTA: Supraventricular tachyarrhythmia.

Validating a custom R-peak methodology involves predicting a combination of beats and non-beats. However, if the predicted samples are tested against reference annotations without using WFDB comparators and the method instead relies solely on window size adjustments through a self-developed R-peak validation methodology, problems may arise. In this case, the methodology would be unable to distinguish between beat and non-beat annotations, resulting in an increase in TPs, FPs, and FNs. WFDB-provided comparators address this issue by filtering out all non-beats and only validate the actual beat annotations, as shown in Figure 8.

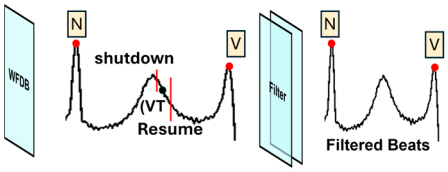


Figure 8. WFDB comparator bxb excluding a non-beat.

Here, the WFDB comparator's general working is shown to illustrate how it processes the actual beats of normal (N) and premature ventricular contraction (V) beats but excludes a non-beat ventricular tachycardia (VT) using the shutdown and resume time mechanism. Shutdown is the time when excluded parts consisting of non-beats start and resume is the time when these parts end. The literature, such as [45–47] and many others as given in Table 1, does not provide the details of the employed R-peak validation procedures. Although these studies provide statistical metrics, it is reasonable to assume that they used custom validation methods for testing R-peaks that raise a question about how they addressed the non-beat issue, as simply adjusting the window size is insufficient to ensure accurate statistical outcomes. In contrast, WFDB R-peak validation tools filter out non-beats, highlighting the importance of a standard R-peak validation method to tackle these issues effectively.

6. Beat Variations for MIT-BIH Dataset

Another discrepancy found while reviewing the existing literature using the MIT-BIH arrhythmia database is that not all studies considered the same number of beats, which impacts the R-peak validation results and fair methodological comparison. The total number of actual beats is 109,494 for all 48 records as available in the PhysioNet database. This number, 109,494, serves as the standard from which we can deduce the following [94]:

- If the literature shows a higher number, it means non-beats have been included in the results.
- If the numbers are lower than 109,494, it indicates that beats have been removed, which should be thoroughly discussed in the manuscript.

In the literature, Zhu et al. reported the total number of beats as 109,401, 93 fewer than the standard 109,494, indicating they missed valid beats in their final statistical results [31]. Additionally, they claimed that the total beats in the MIT-BIH dataset are 116,137 but did not provide a rationale for this figure. Even after reviewing the PhysioNet database and archives, the total number of actual beats (109,494) and non-beats (3153) amounts to 112,647, not 116,137. Rakshit et al. reported the total number of beats as 109,474 [39]. They used fewer beats for records 100, 106, 107, 108, 116, 118, 119, 201, 205, 207, 208, 209, 210, 213, 215, 222, and 233, indicating that they missed many actual beats. They also used more beats in records 101, 105, 115, 121, 123, 124, 200, 203, and 214, showing that they have added some non-beats to the results of these records. Nayak et al. reported the total number of beats as 109,497 [66]. Upon reviewing the number of beats for each record in their manuscript, we found that the reported beats for record 109 are 3 more than the original 2532 beats, which is only possible if all 3 non-beats are added to the final count. Hamilton et al. reported a total number of beats of 109,267, fewer than 109,494 [67]. They also skipped beats from every record of the MIT-BIH dataset but did not mention the timestamps to show which

samples were skipped in their manuscript. Zhang et al. reported the total number of beats as 109,510 [68]. They used more beats in records 103, 108, 207, 214, and 217, indicating that they have added non-beats into the final count without specifying. Afonso et al. reported 91,314 beats using fewer samples from each record [40]. They skipped the first 5 min of each record to comply with the WFDB validation tool mxm settings. The rationale is to allow the signal to stabilize and reduce the impact of any initial artifacts and noise that may be present. Yeh et al. claimed to use a total of 116,137 beats [43]. However, contrary to their claim, the actual number of beats used per record in their manuscript yields a total of not more than 109,809, which is still higher than the actual beat count of 109,494. Similarly, Tompkins et al. [41] also reported the total number of beats as 116,137, but recalculation shows the total number to be 109,809 beats used in their manuscript. Zidelmal et al. [69] also claimed to use the complete number of beats of the MIT-BIH arrhythmia dataset but reported 108,494, 1000 fewer than the actual count of 109,494. These discrepancies emphasize the necessity of uniformity and transparency in reporting the total beats used and timestamps for the skipped segments while validating the R-peak detector. Table 1 also provides a comprehensive overview of the information about beats skipped, where NG indicates unavailable information from referenced papers. This overall discussion leads to three conclusions:

1. Inconsistency in reporting no. of beats makes it difficult to compare results directly since performance metrics (accuracy, precision, recall, etc.) vary depending on the size and composition of the dataset.
2. The selected beats may not uniformly represent various arrhythmia types as some researchers might choose more challenging beats and some just skip them, impacting their proposed detectors' perceived performance.
3. Results from non-standardized datasets cannot be reliably compared; one researcher's high accuracy might be achieved on a simpler subset, while another works with more complex data.

To ensure transparency and comparability, standardized protocols for selecting beats from the MIT-BIH arrhythmia dataset should be followed to properly document and report the exact beats used, including exclusion or selection criteria, along with timestamps. This will provide a complete understanding of the proposed methodology and ensure its reproducibility for meaningful comparisons.

7. Guide for Integrating WFDB Tools, a Future Prospect

WFDB tools are crucial for validating predicted R-peaks, as they handle non-beats and compare actual beats based on defined ANSI/AAMI testing protocols and defined guidelines. However, upon reviewing the available literature and PhysioNet databases related to WFDB tools, it becomes evident that there is lack of guidance and the related information is scattered. This makes it challenging to determine which parameters are critical for R-peak validation, as various parameters are required to obtain SUMSTAT evaluation metrics, and additional parameters are necessary to identify annotation discrepancies and their sample values according to AAMI standard mnemonics. Similarly, there is a lack of clarity on how to effectively integrate WFDB tools to conduct the validation process simultaneously across the entire dataset. To bridge this gap, a structured hierarchy is proposed, which can be coded to develop a graphical user interface (GUI). The proposed structured hierarchy is shown in Figure 9. The main directory is named "Validate", and it contains various subfolders:

1. **Peak Samples**: It stores the predicted R-peak samples as individual CSV files for each record. Each predicted file is placed in a separate folder, with the folder name corresponding to the processed individual record. Each CSV file must contain the predicted samples in a single-column matrix.
2. **Comparator**: This folder stores the predicted annotation files (pred) for the R-peak samples produced by the WRANN function. These annotations are generated by processing subfolders for each record within the "Peak Samples" folder. The Compara-

tor folder also stores the reference annotations (atr) and their corresponding records downloaded from the PhysioNet online directory. These predicted and reference R-peak annotations will be used for window-based comparison.

3. **Results**: It is the target folder to store all resulting metrics, including TPs, FPs, FNs, positive predictivity, and sensitivity in the form of a CSV report file.
4. **CM:** It is a subfolder of "Results" that stores confusion matrices for each record in txt or text format.
5. **Error**: It is also a subfolder of "Results" that stores details of the actual annotation discrepancies for each record in txt format.

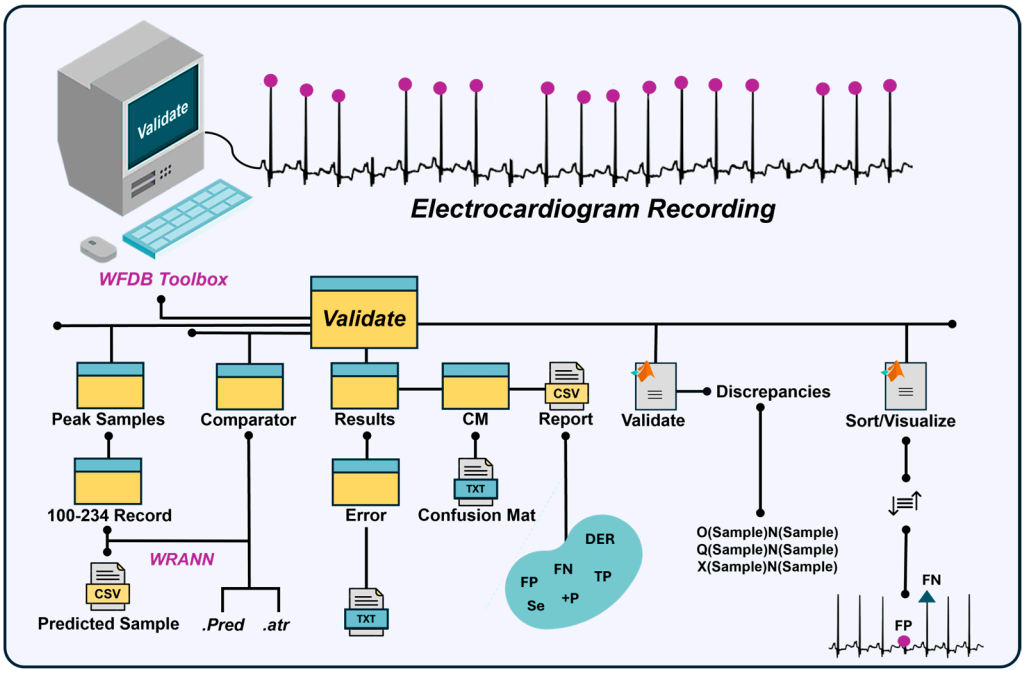


Figure 9. Proposed structured hierarchy that could be utilized for ECG R-peak Validation.

The "Validate" algorithm can serve as the main component of this hierarchy, performing validation for all predicted R-peaks and providing statistical results and annotation discrepancies. The "Sort/Visualize" algorithm can then categorize these discrepancies into FPs and FNs according to standard label mismatches [74]. To sort the discrepancies, a loop can be set up to check if "O" or "X" appears on the left side in any annotation difference file, which will be considered an FP, and if it appears on the right side, it will be considered an FN, as illustrated in Figure 9. Any WFDB comparator can be integrated into this pipeline; here, we use the bxb function as an example due to its frequent use in the literature. The bxb function requires several inputs, including the path to the folder containing expert and predicted annotations with extensions atr and pred and a path where the generated CSV will be saved. It further requires a start time, end time, and window size [75,95] as given below.

bxb (Input path, 'atr', 'pred', path out, 'start time', 'end time', 'window size')

Moreover, to generate annotation differences, a command line can be encoded in the same file as given below.

bxb (-r file, -a atr pred, -f 0, -o, -v 2, target folder)

Here, -r file specifies the record to analyze, -a atr pred indicates the respective annotation files to compare, -f 0 ensures that the analysis starts from the beginning of the record, and -o enables detailed output for each annotation difference in the AAMI-defined mnemonics

format. The -v 2 sets the verbosity level and the target folder is the destination folder, which, according to the proposed structured hierarchy, is the "Error" subfolder.

The step-by-step procedure to use the proposed structured hierarchy is defined as follows:

1. Establish structural pathways for streamlined continuous inputs and outputs as shown in Figure 9.
2. Iterate through each record individually from the "Peak Sample" subfolder, ensuring proper column orientation by transposing each sample point.
3. Use the WFDB WRANN function to generate annotation files (the dot extension can be configured as pred or any other desired name) for each predicted sample for every single record and output the results into the main "Comparator" subfolder. Once all predicted annotations are generated and stored as pred files in the "Comparator" folder, activate the bxb function.
4. The bxb function takes atr and pred files from the "Comparator" folder as inputs to analyze whether each predicted beat is an actual beat, using a 150 ms ($\pm$54 samples) window span. By default, it sets the window to a benchmark size of 0.15 s. Additionally, one has to define extensions for reference and predicted annotation files to configure bxb parameters.
5. Calculate necessary metrics such as TPs, FNs, FPs, sensitivity, and predictability using SUMSTAT.
6. Calculate a confusion matrix for each record, saved in the "Results" subfolder as CM.
7. Append SUMSTAT metrics to a CSV report saved in the "Results" folder.
8. FPs and FNs can then be extracted using the annotation difference command line by defining comparable annotation dot extensions. After extraction, it will be a mix of FPs and FNs which then can be sorted based on the left and right occurrence of "O" or "X".
9. For visualization, one can define a one-minute period or other span of time within the full-length ECG recording of the validating dataset for the error observation feasibility. Visualizing FPs and FNs helps identify all errors and their causes, offering insights into algorithm performance and aiding in quality assessments to ensure reliable ECG interpretations and refinements.

The same process can be repeated using other WFDB validation tools replacing bxb with the respective name of the comparator.

8. Conclusions

This review focuses on transparent, complete, and reliable reporting for ECG R-peak validation. By examining existing literature, it provided an overview of key discrepancies while reporting testing methodologies, window size and tolerance, actual beats, and skipped segments and also provided their potential effects on statistical outcomes. Additionally, it also addresses the challenges of filtering non-beats in custom methods, highlighting the importance of standard WFDB tools and AAMI benchmark parameters. Transparency in reporting validation procedures and parameters is crucial. Transparency is crucial when reporting R-peak validation procedures. This includes clear explanations of how annotated non-beats are filtered when using custom validation methods, the chosen window size, its exact tolerance, and proper documentation of any deviations in the total number of beats for datasets like MIT-BIH or others available on PhysioNet. Similarly, skipped beat segments, along with their exact timestamps and reasons, as well as any custom validation method, should also be fully explained. This is essential to ensure complete and reliable reporting of the validation process, preventing misleading assessments of R-peak detection algorithms, and enabling better understanding and fair comparisons between different methodologies. It allows for a comprehensive understanding of the proposed methodology, replication of exact conditions for R-peak validation, and fair comparison with other ECG R-peak studies. Moreover, a structured hierarchy is conceptualized to integrate WFDB validation tools that can help in understanding provided tools

for validating the ECG R-peak. This hierarchy could pave the way for developing GUIs that simplify R-peak validation procedures and contribute to advancements in the medical research community.

Author Contributions: Conceptualization, S.T.A.A.; methodology, S.T.A.A. and S.K.; Conceptualized software, S.T.A.A. and S.K.; Formal analysis, S.T.A.A. and S.K.; investigation, S.T.A.A.; resources, S.T.A.A. and S.K.; data curation, S.T.A.A. and S.K.; writing—original draft preparation, S.T.A.A.; writing—review and editing, S.T.A.A.; visualization, S.T.A.A.; supervision, Y.-J.K.; project administration, Y.-J.K.; funding acquisition, Y.-J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Gachon University (202307870001, 202404160001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were produced.

Acknowledgments: We would like to thank K Theyagarajan for his assistance with the writing and editing.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Balakumar, P.; Maung-U, K.; Jagadeesh, G. Prevalence and prevention of cardiovascular disease and diabetes mellitus. *Pharmacol. Res.* **2016**, *113*, 600–609. [CrossRef] [PubMed]
2. Almansouri, N.E.; Awe, M.; Rajavelu, S.; Jahnavi, K.; Shastry, R.; Hasan, A.; Hasan, H.; Lakkimsetti, M.; AlAbbasi, R.K.; Gutiérrez, B.C. Early Diagnosis of Cardiovascular Diseases in the Era of Artificial Intelligence: An In-Depth Review. *Cureus J. Med. Sci.* **2024**, *16*, e55869. [CrossRef] [PubMed]
3. Pereira, T.M.; Conceição, R.C.; Sencadas, V.; Sebastião, R. Biometric recognition: A systematic review on electrocardiogram data acquisition methods. *Sensors* **2023**, *23*, 1507. [CrossRef] [PubMed]
4. Moody, G.B.; Mark, R.G. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [CrossRef]
5. Laguna, P.; Mark, R.G.; Goldberg, A.; Moody, G.B. A database for evaluation of algorithms for measurement of QT and other waveform intervals in the ECG. In Proceedings of the Computers in Cardiology 1997, Lund, Sweden, 7–10 September 1997; IEEE: Piscataway, NJ, USA, 1997; pp. 673–676.
6. Wagner, P.; Strodthoff, N.; Bousseljot, R.-D.; Kreiseler, D.; Lunze, F.I.; Samek, W.; Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset. *Sci. Data* **2020**, *7*, 1–15. [CrossRef]
7. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.-K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [CrossRef]
8. Moody, G.B.; Muldrow, W.; Mark, R.G. A noise stress test for arrhythmia detectors. *Comput. Cardiol.* **1984**, *11*, 381–384.
9. Taddei, A.; Distante, G.; Emdin, M.; Pisani, P.; Moody, G.; Zeelenberg, C.; Marchesi, C. The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur. Heart J.* **1992**, *13*, 1164–1172. [CrossRef]
10. Petrutiu, S.; Sahakian, A.V.; Swiryn, S. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *Europace* **2007**, *9*, 466–470. [CrossRef]
11. Alfaras, M.; Soriano, M.C.; Ortín, S. A fast machine learning model for ECG-based heartbeat classification and arrhythmia detection. *Front. Phys.* **2019**, *7*, 103. [CrossRef]
12. Wang, L.-H.; Yu, Y.-T.; Liu, W.; Xu, L.; Xie, C.-X.; Yang, T.; Kuo, I.-C.; Wang, X.-K.; Gao, J.; Huang, P.-C. Three-heartbeat multilead ECG recognition method for arrhythmia classification. *IEEE Access* **2022**, *10*, 44046–44061. [CrossRef]
13. Gacek, A.; Pedrycz, W. *ECG Signal Processing, Classification and Interpretation: A Comprehensive Framework of Computational Intelligence*, 2011th ed.; Springer: Berlin/Heidelberg, Germany, 2011.
14. Zhao, K.; Li, Y.; Wang, G.; Pu, Y.; Lian, Y. A robust QRS detection and accurate R-peak identification algorithm for wearable ECG sensors. *Sci. China Inf. Sci.* **2021**, *64*, 182401. [CrossRef]
15. Rahul, J.; Sora, M.; Sharma, L.D. Exploratory data analysis based efficient QRS-complex detection technique with minimal computational load. *Phys. Eng. Sci. Med.* **2020**, *43*, 1049–1067. [CrossRef] [PubMed]
16. Gliner, V.; Behar, J.; Yaniv, Y. Novel method to efficiently create an mHealth app: Implementation of a real-time electrocardiogram R peak detector. *JMIR mHealth uHealth* **2018**, *6*, e8429. [CrossRef]
17. Sartor, F.; Papini, G.; Cox, L.G.E.; Cleland, J. Methodological shortcomings of wrist-worn heart rate monitors validations. *J. Med. Internet Res.* **2018**, *20*, e10108. [CrossRef]

18. Kumar, S.S.; Rinku, D.R.; Kumar, A.P.; Maddula, R.; Palagan, C.A. An IOT framework for detecting cardiac arrhythmias in real-time using deep learning resnet model. *Meas. Sens.* **2023**, *29*, 100866. [CrossRef]
19. He, R.; Wang, K.; Li, Q.; Yuan, Y.; Zhao, N.; Liu, Y.; Zhang, H. A novel method for the detection of R-peaks in ECG based on K-Nearest Neighbors and Particle Swarm Optimization. *EURASIP J. Adv. Signal Process.* **2017**, *2017*, 82. [CrossRef]
20. Ansari, Y.; Mourad, O.; Qaraqe, K.; Serpedin, E. Deep learning for ECG Arrhythmia detection and classification: An overview of progress for period 2017–2023. *Front. Physiol.* **2023**, *14*, 1246746. [CrossRef]
21. Duan, J.; Wang, Q.; Zhang, B.; Liu, C.; Li, C.; Wang, L. Accurate detection of atrial fibrillation events with RR intervals from ECG signals. *PLoS ONE* **2022**, *17*, e0271596. [CrossRef]
22. Leandro, H.I.C.; Lebedev, D.S.; Mikhaylov, E.N. Discrimination of ventricular tachycardia and localization of its exit site using surface electrocardiography. *J. Geriatr. Cardiol. JGC* **2019**, *16*, 362.
23. Physionet. BXB—ANSI/AAMI-Standard Beat-by-Beat Annotation Comparator. Available online: www.physionet.org/physiotools/wag/bxb-1.htm (accessed on 25 June 2023).
24. Mahmoodabadi, S.; Ahmadian, A.; Abolhasani, M. ECG feature extraction using Daubechies wavelets. In Proceedings of the Fifth IASTED International Conference on Visualization, Imaging and Image Processing, Benidorm, Spain, 7–9 September 2005; pp. 343–348.
25. Mahmoodabadi, S.; Ahmadian, A.; Abolhasani, M.; Eslami, M.; Bidgoli, J. ECG feature extraction based on multiresolution wavelet transform. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 17–18 January 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 3902–3905.
26. *ANSI/AAMI/IEC 60601–2-47*; Particular Requirements for the Basic Safety and Essential Performance of Ambulatory Electrocardiographic Systems. International Standard; International Electrotechnical Commission: Geneva, Switzerland, 2012.
27. Marsili, I.A.; Biasiolli, L.; Masè, M.; Adami, A.; Andrighetti, A.O.; Ravelli, F.; Nollo, G. Implementation and validation of real-time algorithms for atrial fibrillation detection on a wearable ECG device. *Comput. Biol. Med.* **2020**, *116*, 103540. [CrossRef] [PubMed]
28. Doyen, M.; Ge, D.; Beuchée, A.; Carrault, G.; Hernández, A.I. Robust, real-time generic detector based on a multi-feature probabilistic method. *PLoS ONE* **2019**, *14*, e0223785. [CrossRef]
29. Ledezma, C.A.; Altuve, M. Optimal data fusion for the improvement of QRS complex detection in multi-channel ECG recordings. *Med. Biol. Eng. Comput.* **2019**, *57*, 1673–1681. [CrossRef] [PubMed]
30. Bachi, L.; Billeci, L.; Varanini, M. QRS Detection Based on Medical Knowledge and Cascades of Moving Average Filters. *Appl. Sci.* **2021**, *11*, 6995. [CrossRef]
31. Zhu, H.; Dong, J. An R-peak detection method based on peaks of Shannon energy envelope. *Biomed. Signal Process. Control* **2013**, *8*, 466–474. [CrossRef]
32. Qin, Q.; Li, J.; Yue, Y.; Liu, C. An Adaptive And Time-Efficient ECG R-peak Detection Algorithm. *J. Healthc. Eng.* **2017**, *2017*, 5980541. [CrossRef]
33. Pandit, D.; Zhang, L.; Liu, C.; Chattopadhyay, S.; Aslam, N.; Lim, C.P. A lightweight QRS detector for single lead ECG signals using a max-min difference algorithm. *Comput. Methods Programs Biomed.* **2017**, *144*, 61–75. [CrossRef]
34. Kaur, A.; Kumar, S.; Agarwal, A.; Agarwal, R. An Efficient R-peak Detection Using Riesz Fractional-Order Digital Differentiator. *Circuits Syst. Signal Process.* **2020**, *39*, 1965–1987. [CrossRef]
35. Gupta, V.; Mittal, M.; Mittal, V. R-peak detection based chaos analysis of ECG signal. *Analog Integr. Circuits Signal Process.* **2020**, *102*, 479–490. [CrossRef]
36. Park, J.-S.; Lee, S.-W.; Park, U. R Peak Detection Method Using Wavelet Transform and Modified Shannon Energy Envelope. *J. Healthc. Eng.* **2017**, *2017*, 4901017. [CrossRef]
37. Manikandan, M.S.; Soman, K. A novel method for detecting R-peaks in electrocardiogram (ECG) signal. *Biomed. Signal Process. Control* **2012**, *7*, 118–128. [CrossRef]
38. Kaur, A.; Agarwal, A.; Agarwal, R.; Kumar, S. A Novel Approach to ECG R-peak Detection in Electrocardiogram (ECG) Signal. *Arab. J. Sci. Eng.* **2019**, *44*, 6679–6691. [CrossRef]
39. Rakshit, M.; Panigrahy, D.; Sahu, P. An improved method for R-peak detection by using Shannon energy envelope. *Sādhanā* **2016**, *41*, 469–477. [CrossRef]
40. Afonso, V.X.; Tompkins, W.J.; Nguyen, T.Q.; Luo, S. ECG beat detection using filter banks. *IEEE Trans. Biomed. Eng.* **1999**, *46*, 192–202. [CrossRef]
41. Pan, J.; Tompkins, W.J. A real-time QRS detection algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *BME-32*, 230–236. [CrossRef]
42. Varghese, V.J.; Manikandan, M.S. Fast R-peak detection from compressed ECG sensing measurements without reconstruction for energy-constrained cardiac health monitoring. In Proceedings of the 2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART), Paris, France, 7–9 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–4.
43. Yeh, Y.-C.; Wang, W.-J. QRS complexes detection for ECG signal: The Difference Operation Method. *Comput. Methods Programs Biomed.* **2008**, *91*, 245–254. [CrossRef]
44. Chen, H.; Maharatna, K. An Automatic R and T Peak Detection Method Based on the Combination of Hierarchical Clustering and Discrete Wavelet transform. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 2825–2832. [CrossRef]
45. Tang, X.; Hu, Q.; Tang, W. A Real-Time QRS Detection System with PR/RT Interval and ST Segment Measurements for Wearable ECG Sensors Using Parallel Delta Modulators. *IEEE Trans. Biomed. Circuits Syst.* **2018**, *12*, 751–761. [CrossRef]

46. Ravanshad, N.; Rezaee-Dehsorkh, H.; Lotfi, R.; Lian, Y. A Level-Crossing Based QRS-Detection Algorithm for Wearable ECG Sensors. *IEEE J. Biomed. Health Inform.* **2013**, *18*, 183–192. [CrossRef]
47. Elgendi, M.; Mohamed, A.; Ward, R. Efficient ECG Compression and QRS Detection for E-Health Applications. *Sci. Rep.* **2017**, *7*, 459. [CrossRef]
48. Elgendi, M.; Eskofier, B.; Dokos, S.; Abbott, D. Revisiting QRS detection methodologies for portable, wearable, battery-operated, and wireless ECG systems. *PLoS ONE* **2014**, *9*, e84018. [CrossRef]
49. Silva, I.; Moody, G.B. An Open-Source Toolbox for Analysing and Processing Physionet Databases in MATLAB and Octave. *J. Open Res. Softw.* **2014**, *2*, e27. [CrossRef] [PubMed]
50. Mondelo, V.; Lado, M.J.; Mendez, A.J.; Vila, X.A.; Rodriguez-Linares, L. An evaluation tool for wave delineation in ECG processing: Wxw. In Proceedings of the 2018 13th Iberian Conference on Information Systems and Technologies (CISTI), Caceres, Spain, 13–16 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
51. Abdullah Al, Z.M.; Thapa, K.; Yang, S.-H. Improving R Peak Detection in ECG Signal using Dynamic Mode Selected Energy and Adaptive Window Sizing Algorithm with Decision Tree Algorithm. *Sensors* **2021**, *21*, 6682. [CrossRef] [PubMed]
52. Modak, S.; Taha, L.Y.; Abdel-Raheem, E. A Novel Method of QRS Detection Using Time and Amplitude Thresholds with Statistical False Peak Elimination. *IEEE Access* **2021**, *9*, 46079–46092. [CrossRef]
53. Rahul, J.; Sora, M.; Sharma, L.D. Dynamic thresholding based efficient QRS complex detection with low computational overhead. *Biomed. Signal Process. Control* **2021**, *67*, 102519. [CrossRef]
54. Saadi, D.B.; Tanev, G.; Flintrup, M.; Osmanagic, A.; Egstrup, K.; Hoppe, K.; Jennum, P.; Jeppesen, J.L.; Iversen, H.K.; Sorensen, H.B. Automatic Real-Time Embedded QRS Complex Detection for a Novel Patch-Type Electrocardiogram Recorder. *IEEE J. Transl. Eng. Health Med.* **2015**, *3*, 1–12. [CrossRef]
55. Hammad, M.; Maher, A.; Wang, K.; Jiang, F.; Amrani, M. Detection of abnormal heart conditions based on characteristics of ECG signals. *Measurement* **2018**, *125*, 634–644. [CrossRef]
56. Li, H.; Wang, X.; Chen, L.; Li, E. Denoising and R-Peak Detection of Electrocardiogram Signal Based on EMD and Improved Approximate Envelope. *Circuits Syst. Signal Process.* **2014**, *33*, 1261–1276. [CrossRef]
57. Burguera, A. Fast QRS Detection and ECG Compression Based on Signal Structural Analysis. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 123–131. [CrossRef]
58. Dohare, A.K.; Kumar, V.; Kumar, R. An efficient new method for the detection of QRS in electrocardiogram. *Comput. Electr. Eng.* **2014**, *40*, 1717–1730. [CrossRef]
59. Banerjee, S.; Gupta, R.; Mitra, M. Delineation of ECG characteristic features using multiresolution wavelet analysis method. *Measurement* **2012**, *45*, 474–487. [CrossRef]
60. Pal, S.; Mitra, M. Empirical mode decomposition based ECG enhancement and QRS detection. *Comput. Biol. Med.* **2012**, *42*, 83–92. [CrossRef] [PubMed]
61. Ning, X.; Selesnick, I.W. ECG Enhancement and QRS Detection Based on Sparse Derivatives. *Biomed. Signal Process. Control* **2013**, *8*, 713–723. [CrossRef]
62. Gutiérrez-Rivas, R.; Garcia, J.J.; Marnane, W.P.; Hernández, A. Novel Real-Time Low-Complexity QRS Complex Detector Based on Adaptive Thresholding. *IEEE Sens. J.* **2015**, *15*, 6036–6043. [CrossRef]
63. Hossain, M.B.; Bashar, S.K.; Walkey, A.J.; McManus, D.D.; Chon, K.H. An Accurate QRS Complex and P Wave Detection in ECG Signals Using Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Approach. *IEEE Access* **2019**, *7*, 128869–128880. [CrossRef] [PubMed]
64. Yazdani, S.; Vesin, J.-M. Extraction of QRS fiducial points from the ECG using adaptive mathematical morphology. *Digit. Signal Process.* **2016**, *56*, 100–109. [CrossRef]
65. Sabor, N.; Gendy, G.; Mohammed, H.; Wang, G.; Lian, Y. Robust arrhythmia classification based on QRS detection and a compact 1D-CNN for wearable ECG devices. *IEEE J. Biomed. Health Inform.* **2022**, *26*, 5918–5929. [CrossRef]
66. Nayak, C.; Saha, S.K.; Kar, R.; Mandal, D. An efficient and robust digital fractional order differentiator based ECG pre-processor design for QRS detection. *IEEE Trans. Biomed. Circuits Syst.* **2019**, *13*, 682–696. [CrossRef]
67. Hamilton, P.S.; Tompkins, W.J. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Trans. Biomed. Eng.* **1986**, *BME-33*, 1157–1165. [CrossRef]
68. Zhang, F.; Lian, Y. QRS detection based on multiscale mathematical morphology for wearable ECG devices in body area networks. *IEEE Trans. Biomed. Circuits Syst.* **2009**, *3*, 220–228. [CrossRef]
69. Zidelmal, Z.; Amirou, A.; Ould-Abdeslam, D.; Moukadem, A.; Dieterlen, A. QRS detection using S-Transform and Shannon energy. *Comput. Methods Programs Biomed.* **2014**, *116*, 1–9. [CrossRef] [PubMed]
70. Yochum, M.; Renaud, C.; Jacquir, S. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. *Biomed. Signal Process. Control* **2016**, *25*, 46–52. [CrossRef]
71. Young, B. New standards for ECG equipment. *J. Electrocardiol.* **2019**, *57*, S1–S4. [CrossRef]
72. *EC57-2012*; Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms. ANSI/AAMI: Arlington, VA, USA, 2012; p. 46.
73. Physionet. WRANN—Write a WFDB Annotation File. Available online: https://physionet.org/physiotools/wag/wrann-1.htm (accessed on 23 June 2023).
74. Moody, G.B. *WFDB Applications Guide*, 10th ed.; Massachusetts Instittue of Technology: Cambridge, MA, USA, 2022; p. 173.

75. Physionet. Comparing Annotation Files. Available online: https://www.physionet.org/physiotools/wag/evnode10.htm (accessed on 22 June 2023).
76. Physionet. RXR—ANSI/AAMI-Standard Run-by-Run Annotation Comparator. Available online: https://physionet.org/physiotools/wag/rxr-1.htm (accessed on 22 June 2023).
77. Physionet. MXM—ANSI/AAMI-Standard Measurement-by-Measurement Annotation Comparator. Available online: https://physionet.org/physiotools/wag/mxm-1.htm (accessed on 24 June 2023).
78. Physionet. EPIC—ANSI/AAMI-Standard Episode-by-Episode Annotation Comparator. Available online: https://archive.physionet.org/physiotools/old/dbag/epic-1.htm (accessed on 25 June 2023).
79. Bernat, M.; Piotrowski, Z. Software tool for the analysis of components characteristic for ECG signal. In Proceedings of the 2015 22nd International Conference Mixed Design of Integrated Circuits & Systems (MIXDES), Torun, Poland, 25–27 June 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 104–109.
80. Moody, G.B. *WFDB Programmer's Guide*, 10th ed.; Massachusetts Instittue of Technology: Cambridge, MA, USA, 2022; p. 176.
81. Zong, W.; Heldt, T.; Moody, G.; Mark, R. An open-source algorithm to detect onset of arterial blood pressure pulses. In Proceedings of the Computers in Cardiology, Thessaloniki, Greece, 21–24 September 2003; IEEE: Piscataway, NJ, USA, 2003; pp. 259–262.
82. Zanoli, S.; Ansaloni, G.; Teijeiro, T.; Atienza, D. Event-based sampled ECG morphology reconstruction through self-similarity. *Comput. Methods Programs Biomed.* **2023**, *240*, 107712. [CrossRef]
83. Zhang, L.; Huang, M.-J.; Wang, H.-J. A Novel Technique for Fetal Heart Rate Estimation Based on Ensemble Learning. *Mod. Appl. Sci.* **2019**, *13*, 137. [CrossRef]
84. AlDuwaile, D.A.; Islam, M.S. Using convolutional neural network and a single heartbeat for ECG biometric recognition. *Entropy* **2021**, *23*, 733. [CrossRef]
85. Physionet. SUMSTATS—Derive Aggregate Statistics from bxb, rxr, etc., Line-Format Output. Available online: https://archive.physionet.org/physiotools/wag/sumsta-1.htm (accessed on 23 June 2023).
86. McConnella, M.; Schwerina, B.; Soa, S.; Richardsb, B. RR-APET-Heart rate variability analysis software. *Comput. Methods Programs Biomed.* **2020**, *185*, 105127. [CrossRef]
87. Moody, G.; Moody, B.; Silva, I. Robust detection of heart beats in multimodal data: The physionet/computing in cardiology challenge 2014. In Proceedings of the Computing in Cardiology 2014, Cambridge, MA, USA, 7–10 September 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 549–552.
88. Gibbs, A.; Fitzpatrick, M.; Lilburn, M.; Easlea, H.; Francey, J.; Funston, R.; Diven, J.; Murray, S.; Mitchell, O.G.; Condon, A. A universal, high-performance ECG signal processing engine to reduce clinical burden. *Ann. Noninvasive Electrocardiol.* **2022**, *27*, e12993. [CrossRef]
89. Pino, E.; Ohno–Machado, L.; Wiechmann, E.; Curtis, D. Real–Time ECG Algorithms for Ambulatory Patient Monitoring. *Proc. AMIA Annu. Symp. Proc.* **2005**, *2005*, 604.
90. Zidelmal, Z.; Amirou, A.; Adnane, M.; Belouchrani, A. QRS detection based on wavelet coefficients. *Comput. Methods Programs Biomed.* **2012**, *107*, 490–496. [CrossRef]
91. Chen, A.; Zhang, Y.; Zhang, M.; Liu, W.; Chang, S.; Wang, H.; He, J.; Huang, Q. A real time QRS detection algorithm based on ET and PD controlled threshold strategy. *Sensors* **2020**, *20*, 4003. [CrossRef] [PubMed]
92. Rincón, F.; Recas, J.; Khaled, N.; Atienza, D. Development and evaluation of multilead wavelet-based ECG delineation algorithms for embedded wireless sensor nodes. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 854–863. [CrossRef] [PubMed]
93. Chauhan, C.; Agrawal, M.; Sabherwal, P. Accurate QRS complex detection in 12-lead ECG signals using multi-lead fusion. *Measurement* **2023**, *223*, 113776. [CrossRef]
94. Khalaf, A.J.; Mohammed, S.J. Verification and comparison of MIT-BIH arrhythmia database based on number of beats. *Int. J. Electr. Comput. Eng.* **2021**, *11*, 4950. [CrossRef]
95. Physionet. BXB Varargout Function. Available online: https://archive.physionet.org/physiotools/matlab/wfdb-app-matlab/html/bxb.html (accessed on 20 June 2023).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.