



Article

# Statistical and Diagnostic Properties of pRRx Parameters in Atrial Fibrillation Detection

Szymon Buś<sup>1,\*</sup> , Konrad Jędrzejewski<sup>1</sup> and Przemysław Guzik<sup>2</sup>

<sup>1</sup> Institute of Electronic Systems, Faculty of Electronics and Information Technology, Warsaw University of Technology, Nowowiejska 15/19, 00-665 Warsaw, Poland

<sup>2</sup> Department of Cardiology-Intensive Therapy and Internal Disease, Poznan University of Medical Sciences, 60-355 Poznan, Poland

\* Correspondence: szymon.bus.dokt@pw.edu.pl; Tel.: +48-22-2345883

**Abstract:** Background: We studied the diagnostic properties of the percentage of successive RR intervals differing by at least  $x$  ms (pRR $x$ ) as functions of the threshold value  $x$  in a range of 7 to 195 ms for the differentiation of atrial fibrillation (AF) from sinus rhythm (SR). Methods: RR intervals were measured in 60-s electrocardiogram (ECG) segments with either AF (32,141 segments) or SR (32,769 segments) from the publicly available Physionet Long-Term Atrial Fibrillation Database (LTAfDB). For validation, we have used ECGs from the Massachusetts Institute of Technology–Beth Israel Hospital (MIT–BIH) Atrial Fibrillation Database. The pRR $x$  distributions in AF and SR in relation to  $x$  were studied by histograms, along with the mutual association by the nonparametric Spearman correlations for all pairs of pRR $x$ , and separately for AF or SR. The optimal cutoff values for all pRR $x$  were determined using the receiver operator curve characteristic. A nonparametric bootstrap with 5000 samples was used to calculate a 95% confidence interval for several classification metrics. Results: The distributions of pRR $x$  for  $x$  in the 7–195 ms range are significantly different in AF than in SR. The sensitivity, specificity, accuracy, and diagnostic odds ratios differ for pRR $x$ , with the highest values for  $x = 31$  ms (pRR31) rather than  $x = 50$  (pRR50), which is most commonly applied in studies on heart rate variability. For the optimal cutoff of pRR31 (68.79%), the sensitivity is 90.42%, specificity 95.37%, and the diagnostic odds ratio is 194.11. Validation with the ECGs from the MIT–BIH Atrial Fibrillation Database confirmed our findings. Conclusions: We demonstrate that the diagnostic properties of pRR $x$  depend on  $x$ , and pRR31 outperforms pRR50, at least for ECGs of 60-s duration.

**Keywords:** atrial fibrillation detection; cardiac arrhythmia; electrocardiography; heart rate variability



**Citation:** Buś, S.; Jędrzejewski, K.; Guzik, P. Statistical and Diagnostic Properties of pRR $x$  Parameters in Atrial Fibrillation Detection. *J. Clin. Med.* **2022**, *11*, 5702. <https://doi.org/10.3390/jcm11195702>

Academic Editors: Christian Sohns and Paweł T. Matusik

Received: 1 August 2022

Accepted: 23 September 2022

Published: 27 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Atrial fibrillation (AF) is a tachyarrhythmia with uncoordinated atrial electrical activation and ineffective atrial contraction [1–3]. Diagnosis of AF is based on an electrocardiogram (ECG) with irregular RR intervals (the distances between the peaks of R waves of the QRS complexes reflecting the electrical depolarization of ventricles and measuring the duration of each cardiac cycle), the absence of distinct repeating P waves, and irregular atrial activation. AF must be documented in an entire 12-lead ECG of a duration of at least 8–10-s. If a single ECG strip is used, an arrhythmic episode is considered AF if it lasts at least 30 s [1,2].

AF is the most common sustained cardiac arrhythmia, and it is the only available cardiac rhythm for many people. Heart palpitations, symptoms of irregular pulse, worsened exercise tolerance, dyspnea, and sometimes angina can accompany AF. However, many individuals are asymptomatic or experience mild or unspecific symptoms easily related to advancing age, emotions or effort, or other diseases such as hyperthyroidism [4]. Consequently, AF may be undetected or is found accidentally during a routine check-up or a medical visit for another reason [1,2].

AF increases the risk of excessive morbidity with debilitating clinical consequences and premature death. This arrhythmia always increases the risk of severe arterial thromboembolism, resulting, for example, in an ischaemic stroke of the brain. Therefore, AF must be appropriately diagnosed and actively screened for, particularly in high-risk populations such as older age and patients with hypertension, diabetes, heart failure, or valvular disease [5,6].

Many systems used for AF screening can record ECGs of various lengths, ranging from a couple-of-second 12-lead ECG to recordings of several weeks with the long-term Holter ECGs or ECG bio-patches attached to the skin. Implantable loop recorders are an invasive option for even more prolonged ECG monitoring lasting up to two years.

Visual inspection and ECG analysis are the gold standards for AF diagnosis. With technological advancement, the amount of data with long-term ECGs gradually increases, so a quick analysis of such recordings becomes challenging. New approaches to automatic or semiautomatic analysis are proposed, many of them based on heart rate variability (HRV) analysis [7]. HRV uses mathematical analysis of RR interval time series. RR interval is the cardiac cycle duration measured between two consecutive QRS complexes in a continuous ECG. HRV was developed many years ago. Its primary use was limited only to RR intervals of sinus rhythm (SR) origin for either physiological analysis or prediction of mortality in different diseases.

Due to different distributions of HRV-derived parameters in AF and SR [8], HRV has gained new interest in AF detection in ECG [9–11], as well as from wearable devices [12,13]. Several authors used feature selection methods to find the most relevant HRV parameters for AF detection [14,15]. Others incorporated HRV to predict future occurrence of AF [16,17].

The percentage of successive RR intervals differing by at least 50 ms (pRR50) is a particular form of pRR $x$  parameter. Ewing et al. [18] proposed the total number of successive RR intervals that differ by at least  $x$  ms (RR $x$  count) for monitoring cardiac parasympathetic activity. They analyzed two threshold values:  $x = 50$  ms and  $x = 6.25\%$  of the previous RR interval. The computationally simpler 50 ms threshold was later widely adopted, and Bigger et al. [19] proposed a relative statistic pNN50 (percentage of normal-to-normal RR interval differences over 50 ms). The “NN” in the name emphasizes that the analyzed R-waves are of sinus origin [1,19]. However, the same mathematical analysis can be conducted for different cardiac rhythms, including AF. RR intervals are not normal (i.e., of sinus origin) in AF and thus should always be labeled as RR and not as NN.

So far, the threshold values  $x$  other than 50 ms have rarely been analyzed. Mitetus et al. [20] used pRR $x$  with  $x$  ranging from 4 ms to 100 ms to compare various groups such as healthy people versus patients with heart failure, sleeping vs. awake states, or young vs. elderly subjects. In all cases, thresholds  $<50$  ms allowed for better discrimination between the studied groups. Torres et al. [21] applied pRR $x$  with  $x$  from 10 ms to 50 ms to distinguish healthy subjects from survivors of acute myocardial infarction, and thresholds  $<50$  ms performed better. Saiz-Vivo et al. [22,23] used HRV indices, including pRR20 and pRR50, to analyze 500 beats preceding onsets of AF and distinguish between healthy and AF subjects. In several studies [24–27], various pRR $x$  indices with  $x$  from 5 ms to 500 ms were used to distinguish between four cardiac rhythms (AF/SR/noisy/other) [28]. Jovic et al. [29] combined various HRV parameters with pRR5, pRR10, pRR20, and pRR50 to differentiate between nine cardiac rhythms, including AF [30].

We have recently analyzed the diagnostic properties of several HRV parameters, including the mean of RR intervals (mean RR); the standard deviation of RR intervals (SDRR); the standard deviation of points perpendicular to the line of identity (SD1) and along the line of identity (SD2) from the Poincare plot analysis; the power of low-frequency (LF: 0.04–0.15 Hz) and high-frequency (HF: 0.15–0.4 Hz); and pRR10, pRR30, pRR50, pRR70, and pRR90 [31]. Among these parameters, pRR30 had the highest area under the curve (AUC) [32]. Different combinations of the HRV parameters, used as input features for ML classifiers, showed that the sets with pRR30 outperformed other HRV feature sets, including

those with pRR50. Conroy et al. [33] used parameters analogous to pRR $x$  to detect AF in the photoplethysmographic signal (PPG). Instead of RR intervals, interbeat intervals (IBI) from PPG were used, and the highest AUC was yielded for  $x = 35$  ms. Ramesh et al. [34] used HRV features, including pRR20, RR20, pRR50, and RR50 (and analogous parameters from PPG), for AF detection in ECG and PPG.

Altogether, these data strongly suggest that various  $x$  threshold values for the difference between two successive RR intervals might be more helpful in detecting AF. However, the 50 ms threshold (pRR50) has been used for many years as a commonly accepted practice in HRV analysis.

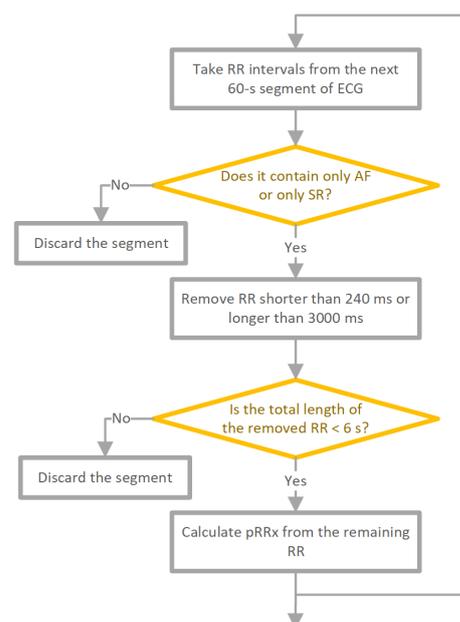
The primary aim of this study was to systematically explore the diagnostic properties of the pRR $x$  as a function of the different  $x$  thresholds for differentiating AF from SR in 60-s ECG segments. The secondary aim was to compare the diagnostic properties of the pRR50 with the pRR $x$  found to have the optimal diagnostic properties for AF detection in the same 60-s ECG segments. This part of the study is presented separately in Appendix A.

## 2. Materials and Methods

### 2.1. Data

For this study, we used anonymized data from the Long-Term Atrial Fibrillation Database (LTAfDB) [35,36]. It contains 84 long 24-h Holter electrocardiographic (ECG) recordings sampled at a rate of 128 Hz, the information about locations of R-waves, and the type of corresponding cardiac beats (normal, supraventricular, ventricular, atrial fibrillation, and technical artifact). The LTAfDB database contains ECG recordings from patients with paroxysmal AF and other arrhythmias. We selected only uninterrupted ECG fragments with either AF or SR of at least 60-s duration for further analysis. We discarded the segments labelled as different rhythms.

The data-preprocessing scheme is presented in Figure 1. The RR interval time series were cut into 60-s separate, neighboring segments. For a segment to be labeled SR, each RR interval had to be of SR origin. If it was not, for example, it was a atrial or ventricular beat, the segment was removed from further analysis. For AF segments, each cardiac beat needed to be AF, and if a ventricular beat was found, such a segment was also excluded.



**Figure 1.** Data preprocessing scheme. RR—interval between peaks of consecutive R-waves; ECG—electrocardiogram; AF—atrial fibrillation; SR—sinus rhythm; and pRR $x$ —percentage of successive RR intervals differing by at least  $x$  ms.

For SR and AF, to limit the number of potentially unidentified technical artifacts, RR intervals shorter than 240 ms or longer than 3000 ms were removed. Additionally, RR intervals corresponding to ventricular premature beats were also removed from both SR and AF ECGs. For SR only, premature supraventricular beats were removed too. Segments with a total length of excluded RR intervals exceeding 6 s were also discarded from the analysis. The total number of 60-s RR series after preprocessing was 64,910 (32,141 AF, 32,769 SR).

As the sampling rate of 128 Hz corresponds to the precision of 7.8125 ms, we quantized pRRx thresholds  $x$  into 7.8125 ms bins. The absolute values of the differences between consecutive RR intervals were measured and used to calculate pRRx for the  $x$  ranging from 7.8125 to 195.3125 ms in 7.8125 ms steps for both SR and AF 60-s segments. To improve the paper's readability, we use only the integer part of  $x$  in ms in the names of parameters, e.g., pRR7 instead of pRR7.8125.

## 2.2. Software Tools

We used Python programming language (version 3.9, Python Software Foundation, Wilmington, DE, USA) for all the analyses.

## 2.3. Statistical Analysis

First, we analyzed the distributions of pRRx obtained for different  $x$  values separately for SR and AF using histograms. Based on the histograms and Shapiro–Wilk test results, we concluded that pRRx do not have normal distributions either in AF or in SR. Consequently, we used a percentile scale to describe the distributions and mostly applied nonparametric statistical techniques for data analysis.

Next, using the Wilcoxon test [37], we made paired comparisons for pRRx with different  $x$  (separately for SR and AF). We conducted the unpaired analysis comparing pRRx for the same  $x$  between SR and AF using the Mann–Whitney test [38]. The associations between pRRx values for different  $x$  were analyzed with the Spearman correlation [39] and presented as heatmaps with rho correlation coefficients. To analyze the differences between two pRRx with different  $x$ , we calculated the mean difference for quantifying bias, and the standard deviation (SD) of the differences for each pair of pRRx. As the data distribution was not normal, we defined the limits of agreement (LoA) as the range between the 2.5th and 97.5th percentile of the distribution of the differences. To analyze the diagnostic properties of pRRx with different  $x$ , the area under the curve (AUC) from the receiver operator curve (ROC) characteristics was calculated [32]. We identified the optimal cutoff values for each  $x$  using Youden's Index [40], which maximizes the sum of sensitivity and specificity of AF/SR differentiation.

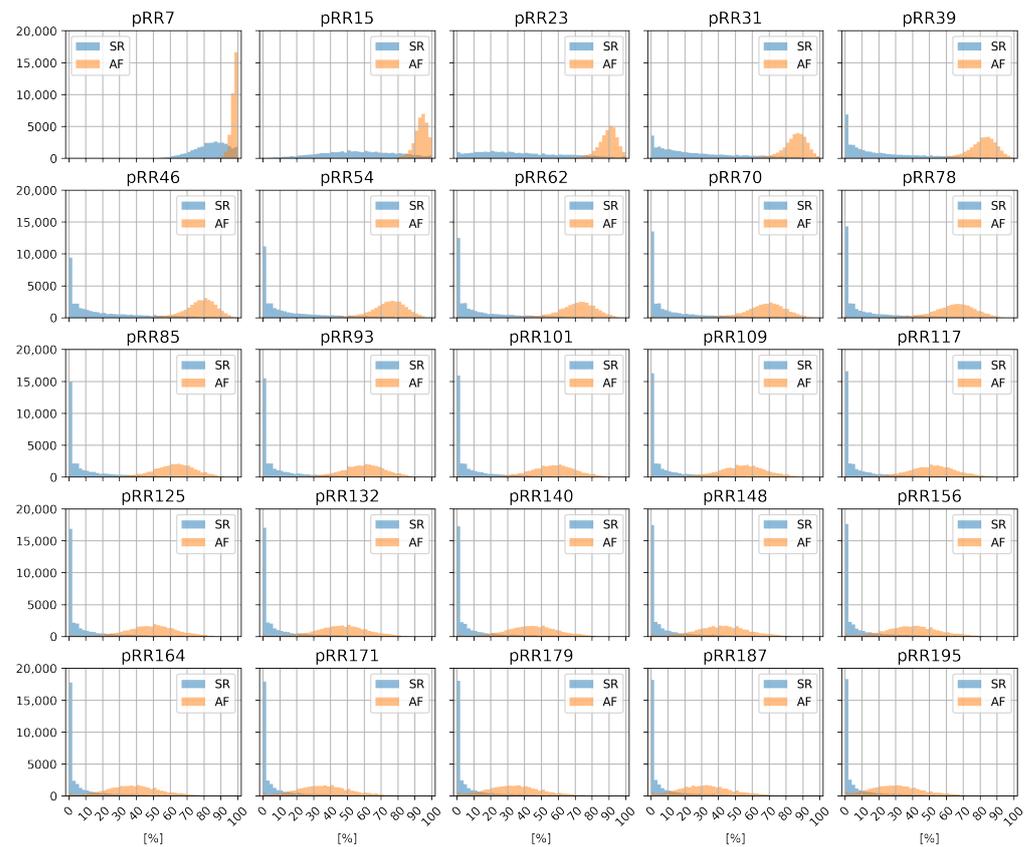
For each threshold  $x$  with the optimal cutoff value for differentiating between AF and SR, we calculated several classification metrics [41], namely, accuracy, specificity, sensitivity, F1-score [42], positive predictive value (PPV), negative predictive value (NPV), and diagnostic odds ratio (DOR) [43]. For the estimation of classification metrics' 95% confidence interval (CI), we used a nonparametric bootstrap with 5000 samples [44]. We analyzed all measures as functions of the threshold value  $x$  of pRRx in the range of 7 to 195 ms.

As the use of the pRR50 parameter is a de facto standard, we decided to compare its diagnostic properties against the optimal pRRx for AF detection. The optimal cutoff values for both parameters were estimated with the Youden criterion [40], and we used a nonparametric bootstrap to compare the distributions of the classification metrics with optimal cutoffs to verify whether the diagnostic properties of different thresholds are not random. For the sampling frequency of 128 Hz in the LTAfDB, the threshold  $x = 54$  ms is equivalent to  $x = 50$  ms, so pRR54 was used as an equivalent of pRR50.

### 3. Results

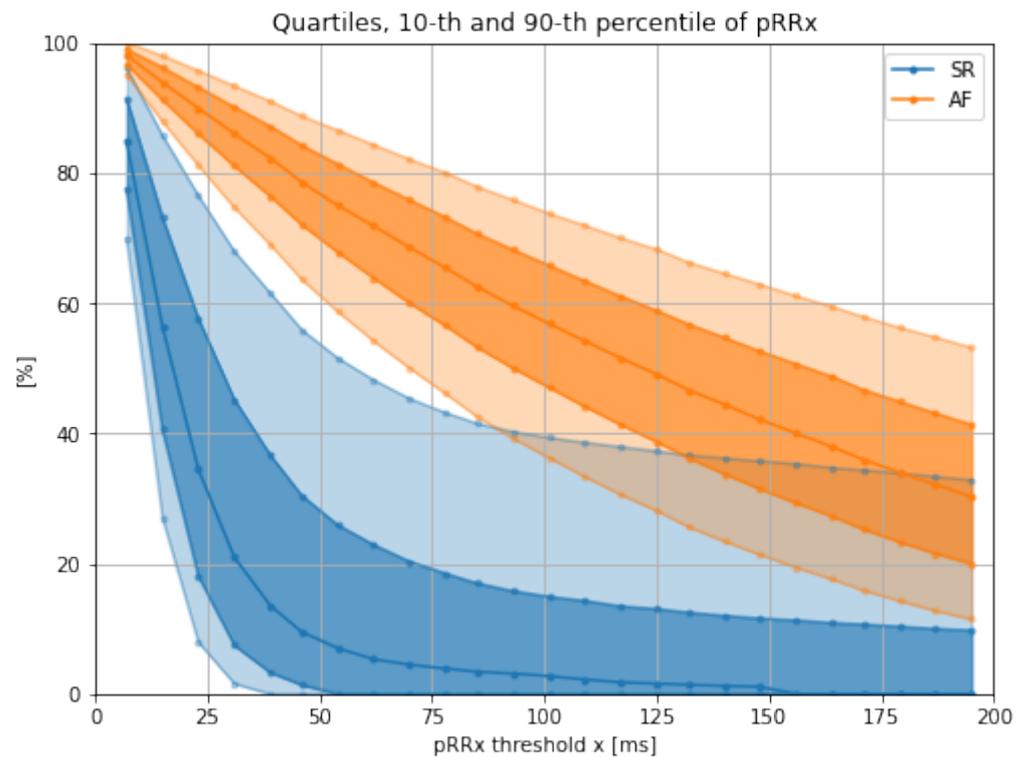
#### 3.1. Data Distribution Analysis

Figure 2 shows pRRx histograms for SR and AF for different x threshold values between 7 and 195 ms. The histograms for SR and AF partially overlap—for the same x, the left part of the SR distribution covers the right part of the AF distribution. The overlap is notable for very low (pRR7, pRR15) and very high values of the threshold x (pRR101–pRR195), but the distributions are better separated between these extremes.



**Figure 2.** Histograms of percentages of successive RR intervals differing by at least x ms (pRRx) for sinus rhythm (SR) and atrial fibrillation (AF) for different x thresholds.

Figure 3 presents medians (bolder line), the 25th to 75th percentile range (darker band), and the 10th to 90th percentile range (lighter band) for AF (orange) and SR (blue). Median values of SR and AF never cross or overlap in the whole range of studied x. Interestingly, there is also no overlap between the 75th percentiles of pRRx for SR and the 25th percentiles of pRRx for AF. Additionally, the 90th percentile of pRRx for SR does not cross with the 10th percentile for AF in the range of x from 15 to 85 ms. It suggests that the optimal value of threshold x for differentiating between AF and SR lies within this range.



**Figure 3.** Medians (bolder line), interquartile ranges (darker band), and 10th to 90th percentile ranges (lighter band) of the percentages of successive RR intervals differing by at least  $x$  ms (pRR $x$ ) for sinus rhythm (SR) and atrial fibrillation (AF) for different  $x$  thresholds.

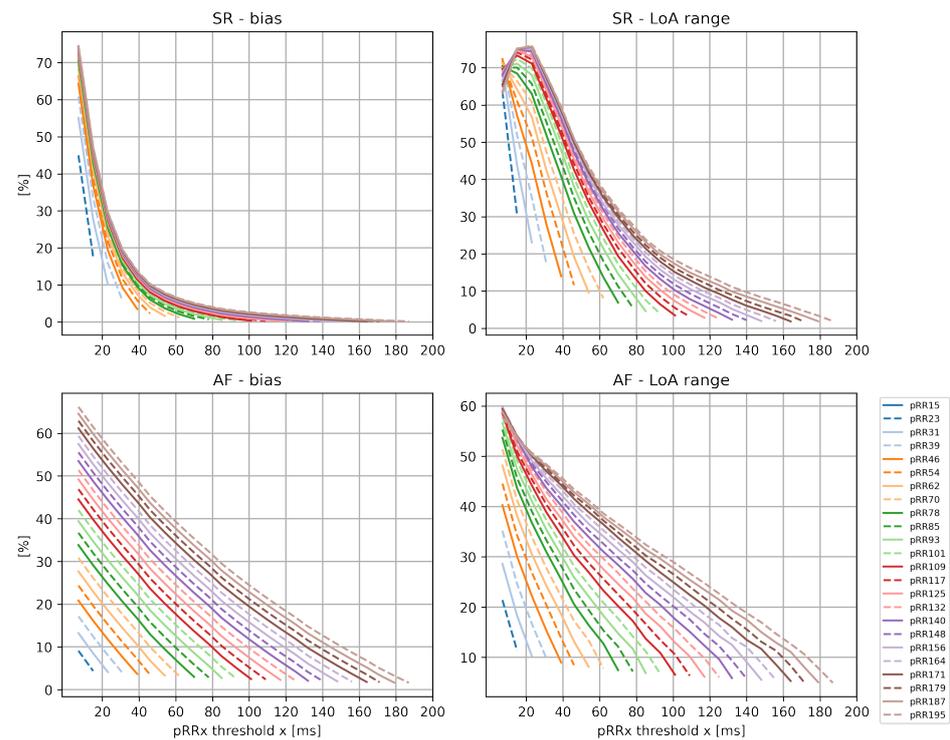
### 3.2. Correlation

Figure 4 shows two heatmaps with rho coefficients of Spearman correlations between pRR $x$  for the different  $x$  thresholds in SR (the upper panel) and AF (the lower panel). For most  $x$  thresholds, particularly for lower values, the correlation coefficients are higher for SR than for AF. For the highest values of  $x$ , these correlations are above 0.95, both for SR and AF.

The rho coefficients decrease for SR and AF as the distance between two pRR $x$  increases. Interestingly, this effect is more pronounced for lower  $x$ . For instance, in AF recordings, the rho coefficient drops from 0.7 for the pair (pRR7, pRR15) to 0.63 for (pRR7, pRR23). In contrast, rho is 1 for two pairs (pRR195, pRR187) and (pRR195, pRR179).

In AF, the pRR $x$  with the lowest  $x$  (pRR7) has the strongest correlation with pRR15 (0.7), which is lower than in SR (0.92). However, the weakest correlation (between pRR7 and pRR195) is higher in AF (0.42) than in SR (0.38). It shows that for the lower values of  $x$ , the range of correlations is much wider for SR than for AF. Additionally, it suggests that the strength of correlations between various pRR $x$  changes non-linearly for SR and AF.

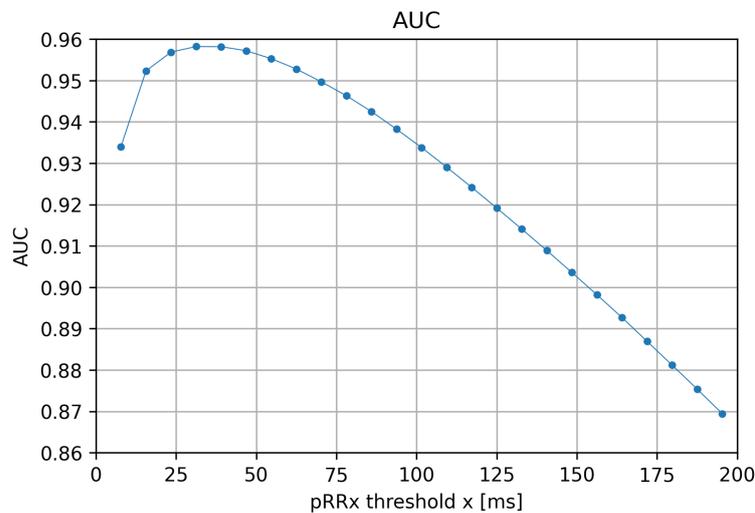




**Figure 5.** Summary of the analysis of the differences between the pairs of the percentages of successive RR intervals differing by at least  $x$  ms (pRRx parameters). Mean difference (bias) and 95% limits of agreement (LoA) range in sinus rhythm (SR, top) and atrial fibrillation (AF, bottom).

### 3.4. Area under ROC Curve (AUC)

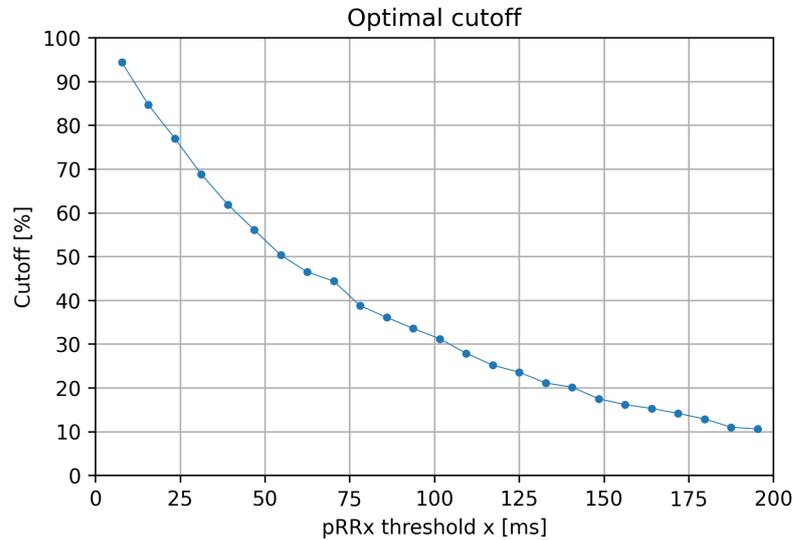
The AUCs for the differentiation of AF from SR, shown in Figure 6, have their peak values exceeding 0.94 for the range of  $x$  from 15 ms to 85 ms (identical with the no overlapping zone between the 90th percentile of the pRRx distribution of SR and 10th percentile for AF as visible in Figure 3). AUCs gradually decrease to 0.87 for the highest  $x$  thresholds. In other words, pRRx for lower  $x$  values has a much better AUC for distinguishing the SR from AF recordings. The maximal AUC value can be one of the criteria for selecting the optimal  $x$  threshold for differentiating SR from AF with pRRx. In our study, the maximal AUC value (0.958) was for  $x = 31$  ms, i.e., pRR31.



**Figure 6.** Area under curve (AUC) of the percentages of successive RR intervals differing by at least  $x$  ms (pRRx parameters) for the differentiation of AF from SR as a function of the  $x$  threshold.

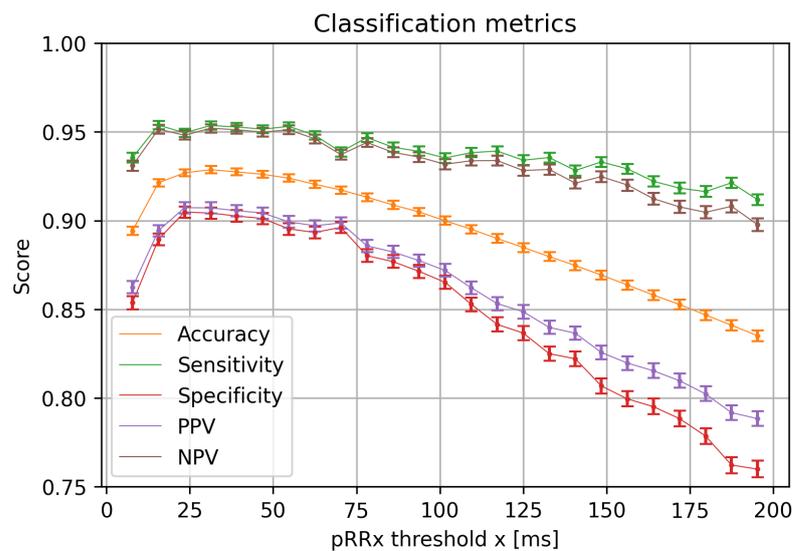
### 3.5. Determining Optimal Cutoff Values for Different pRRx

Using Youden’s index, we selected optimal cutoff values for all pRRx (Figure 7). Notably, the cutoff values for pRRx strongly and negatively depended on their x threshold.



**Figure 7.** Optimal cutoff values based on Youden Index the percentages of successive RR intervals differing by at least x ms (pRRx parameters) as a function of the x threshold.

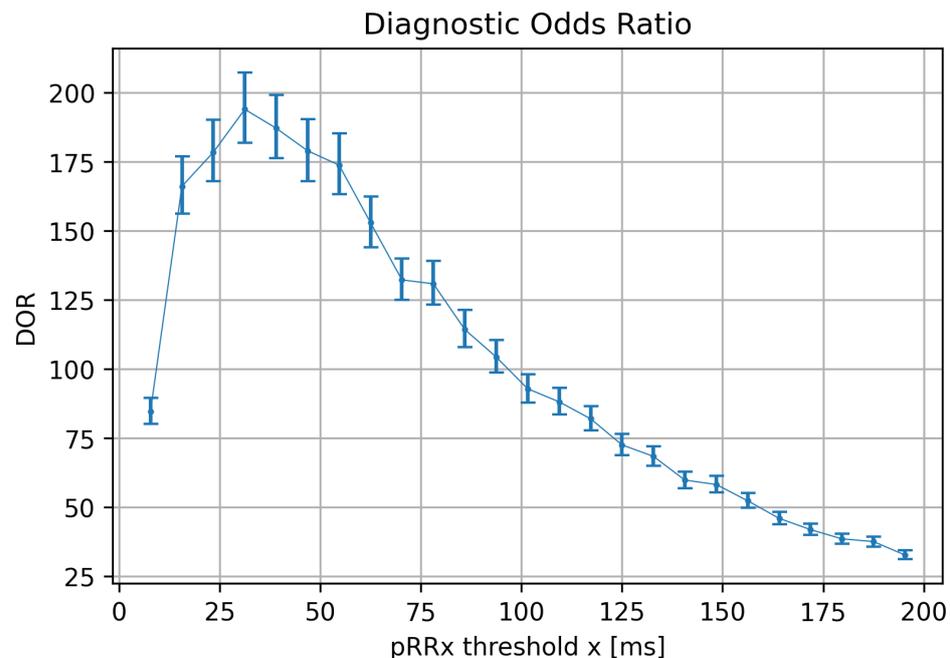
Next, the nonparametric bootstrap estimated the statistical metrics of AF detection using pRRx parameters with the optimal cutoffs. Figure 8 shows the median values and 95% CI of the classification metrics for optimal cutoffs of pRRx in relation to the x threshold. Although there is a gradual slow decline in the sensitivity of pRRx, its median exceeds 0.9 in the whole studied range of x. For the NPV, this relation is similar, with the lowest values above 0.89 for the largest x = 195 ms. Gradual decline with the increasing x values is also visible for both PPV and specificity. However, the decline is more rapid and deeper, from 0.89 to 0.77 (PPV), and from 0.88 to 0.72 (specificity) for x thresholds above 50 ms.



**Figure 8.** Medians and 95% confidence intervals of diagnostic classification metrics for optimum cut-off values for the percentages of successive RR intervals differing by at least x ms (pRRx parameters) as a function of the x threshold. PPV—positive predictive value, NPV—negative predictive value.

### 3.6. The Diagnostic Odds Ratios for Optimal Cutoffs for Different pRRx

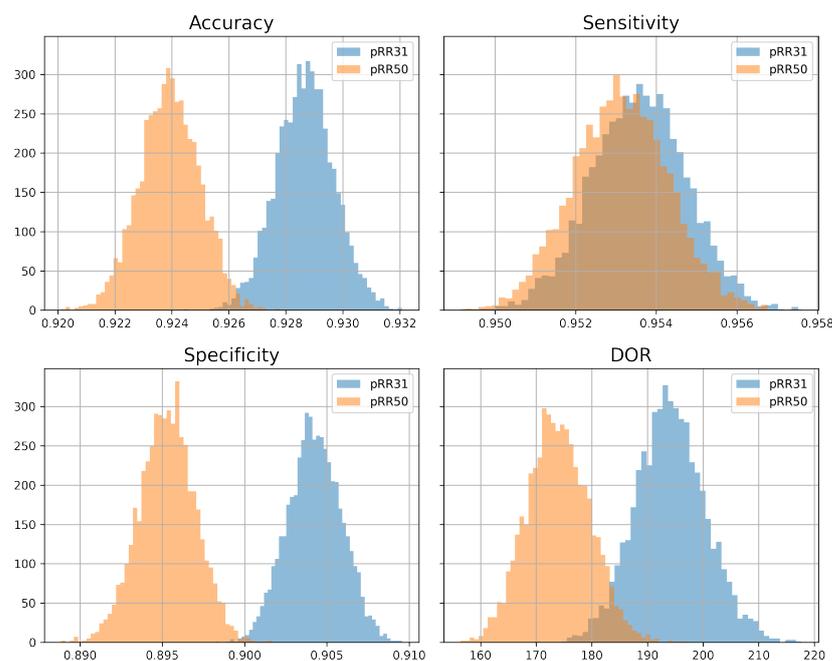
Figure 9 presents median values of DOR with 95% CI for optimal cutoffs for different pRRx in relation to the x value. The highest DOR of 194.01 is observed for pRR31. In other words, the odds of pRR31 being greater or equal to the cutoff = 68.79% (Figure 7) is nearly 200 higher for the AF presence than in its absence. The next highest DORs are for pRR39, pRR46, and pRR25, all exceeding 175. DORs for pRRx for x up to 81 ms are at least 125, then gradually decline with increasing x values, reaching a minimum of 32.68 for x = 195 ms.



**Figure 9.** Medians and 95% confidence intervals of diagnostic odds ratio (DOR) for optimum cutoff values for the percentages of successive RR intervals differing by at least x ms (pRRx parameters) as a function of the x threshold.

### 3.7. Comparison of pRR50 and pRR31

The threshold x = 50 ms of pRRx is most broadly used and analyzed in HRV literature, including HRV-based AF detection methods. On the other hand, in our study, pRR31 achieved the highest accuracy (Figure 8) and DOR (Figure 9). We used a nonparametric bootstrap to compare the distributions of classification metrics obtained by pRR31 and pRR50 with optimal cutoffs to verify whether better diagnostic properties of different thresholds are not random. For the sampling frequency of 128 Hz in the LTAfDB, the threshold x = 54 ms is equivalent to x = 50 ms, so pRR54 was used as an equivalent of pRR50. Figure 10 shows the histograms of accuracy, sensitivity, specificity, and DOR for pRR50 and pRR31. All the metrics are visibly higher for pRR31 than for pRR50, except for sensitivity, which is only slightly higher for pRR31. The Shapiro–Wilk test verified that accuracy, sensitivity, and specificity have normal distributions and DOR does not. The histograms and the results of paired t-test (accuracy, sensitivity, and specificity) and Wilcoxon test (DOR) demonstrate that better performance of pRR31 is not random.



**Figure 10.** Histograms of classification metrics for optimum cutoff values for pRR31 and pRR50 (the percentages of successive RR intervals differing by at least 31 ms and 50 ms, respectively). DOR—diagnostic odds ratio.

#### 4. Discussion

In this study, we summarize various statistical properties of pRR $x$  as functions of the  $x$  threshold both for AF and SR. All possible pairs of pRR $x$ , regardless of the  $x$  distance separating them, are well correlated (Figure 4). Good correlations with  $\rho$  at least 0.7 were present for all pRR $x$  starting from  $x = 46$  ms and most for  $x = 39$  ms in SR recordings. In AF, all pairs of pRR $x$  are well correlated, starting from  $x = 31$  ms and then for most  $x = 23$  ms. For higher  $x$ , pairs of pRR $x$  are perfectly correlated. Lower pRR $x$  values provide more information than higher pRR $x$ . For instance, compared to pRR100, the pRR31 informs about the percentage of neighboring RR intervals, which differ not only from 100 ms but also in the lower range between 31 and 100 ms. In other words, using pRR $x$  with higher  $x$  thresholds is more aggressive as it filters out much important information in the lower  $x$  ranges. Another issue is that good correlation does not mean that all pRR $x$  parameters are the same and thus replaceable. Besides, caution should be exercised for the lowest  $x = 7$  ms, for which the  $\rho$  values drop to 0.38 and 0.42 for SR and AF, respectively. This analysis shows that various pRR $x$  provide different information, particularly those with lower  $x$ .

The analysis of differences between different pRR $x$  (Figure 5) reveals that their mean difference (bias) and the 95% limits of agreement range have distinct properties in AF and in SR. It shows that the decline of pRR $x$  differences is more dynamic for SR than AF, even for the neighboring  $x$ . This analysis also demonstrates that pRR $x$  with different  $x$ , even for the closest pairs of pRR $x$ , do not provide identical information.

The optimal cutoffs for pRR $x$  depended differently on the threshold  $x$  in a wide range from 7 ms to 195 ms. The best diagnostic features for AF detection are for the  $x$  range between 15 and 85 ms. It was repeatedly indicated by comparing pRR $x$  90th percentile for SR and 10th percentile for AF (Figure 3), then AUC over 0.94 (Figure 6), and finally with odds ratios exceeding 110 for specific cutoffs (Figure 9). Out of several possible  $x$  values within this range,  $x = 31$  ms appears to be the best.

pRR50 is the most broadly studied pRR $x$  parameter in physiological and clinical studies with ECG with sinus rhythm and for the differentiation of AF from SR [45–48]. Only a few studies explored pRR $x$  with threshold values  $x$  different than 50 ms in AF detection [22,23,25–27,29,31,33]. Among them, only one study contained an analysis of

the distribution of pRRx in AF and SR [26], and two had AUCs of the parameters [31,33]. However, neither analysis was in-depth.

Our study is more detailed, systematic, and focused on the whole pRRx family in a broad range of x thresholds. It is probably the first analysis investigating various statistical properties of pRRx exclusively as a function of x for the differentiation of AF from SR.

This study provides practical consequences for distinguishing AF from SR in 60-s ECG segments. First, although pRR50 is within the range of x values with very good diagnostic properties, it is not the best. Second, for the 60-s ECG segments, the  $x = 31$  ms outperforms other pRRx (including pRR50), particularly when comparing DORs. The odds of pRR31 exceeding 68.79% (Figure 9) is nearly 200 times higher in the presence of AF than in the absence of AF. Third, we demonstrate how distinct the statistical properties of pRRx are in SR and AF. Fourth, if HRV parameters are used to differentiate AF from SR, for example, in the machine learning algorithms, pRRx with x shorter than 50 should be applied.

In SR, the differences between the duration of successive RR intervals are much more limited than in AF. The RR interval for SR usually falls within 80 to 120% of the previous RR interval. For AF, no such limit exists. It results from a couple of physiological regulatory mechanisms controlling SR, which have weaker or no effect on the heart rate during AF.

SR originates in the sinus node, where pacemaker depolarization activity is regulated by several controlling mechanisms. The electrical depolarization from the sinus nodes travels across the right atrium and reaches the atrioventricular node. The atrioventricular conduction undergoes additional controlling regulations. During AF, thousands of cells from the border of pulmonary veins and the left atrium or both atria depolarize spontaneously in a relatively uncontrolled way. Atrial depolarizations reach the atrioventricular node in a more or less random order, which may or may not conduct them through the His system to both ventricles. The refractory period is probably the most important physiological mechanism that is still functioning; it may control the ventricle rate in AF. All cells of the atrioventricular node, His system, and ventricles have various refractory periods. During this period, cardiac cells below the atria cannot respond to electrical depolarization. The refractory period, particularly in the atrioventricular node, is under the strong control of the autonomic nervous system during SR. This control is less effective but is still present during AF. In other words, less controlling mechanisms over the ventricular rate in AF introduce some randomness to RR intervals and thus higher values of pRRx than in SR.

The limitations of the study must be recognized. It is an observational and retrospective study using ECG recordings from a single (Long-Term AF Database) database. To verify whether the results are not database-specific, we used optimal cutoffs of pRRx from LTAfDB and employed them to classify 60-s segments from MIT-BIH Atrial Fibrillation Database (250 Hz sampling frequency) [35,49]. The highest DOR (median 276) and accuracy (median 0.944) were obtained for pRR31 (in LTAfDB 194 and 0.928, respectively—the detailed results are shown in Appendix A). Moreover, these public databases have been used in several studies, and others can easily replicate our results. Other limitations are the arbitrary use of 60-s ECG segments and the application of additional filters removing too short (<240 ms) or too long (>3000 ms) RR intervals to select ECG segments that might result in “too perfect” results. Thus, the interpretation of our findings must be limited only to the specific settings of the filters and ECG recording of 60-s duration.

Comparison of different methods and parameters is always complex and should never be based on a single approach or descriptor. As demonstrated, visual inspection is always essential—such a simple approach as the distribution analysis clearly shows how pRRx changes for various x, both for SR and AF. Unfortunately, presenting the actual data distribution is not a common approach. Next, correlation analysis is the most popular in many studies to present how well some parameters are correlated. Presented correlation heatmaps also show strong correlations for pRRx in a wide range of x values. Many studies stop at this point, without further and more detailed exploration. However, additional analyses with classic statistical methods, starting with differences analysis, followed by more advanced analyses such as ROC, the identification of cutoff points,

and various classification measures, reveal huge differences between parameters that appear so well related. Summarizing this part, a set of several classical statistical methods for comparing various methods or parameters should be constantly employed in clinical medicine. Referring to one method, usually correlation analysis, can produce misleading conclusions and false diagnoses.

## 5. Conclusions

In conclusion, pRRx values for various  $x$  thresholds are not the same or interchangeable. Although most pRRx help detect AF, the parameters around pRR31 outperform others. If in a 60-s ECG the pRR31 is at least 68.79%, it is far more likely to be AF than SR. Using optimal pRRx instead of pRR50 should improve machine learning models for AF detection. Our results potentially apply to different biomedical systems used for screening for AF, e.g., long-term ECG Holter or bio-patch systems. However, we are aware that this potential should be validated using real-life ECG recordings acquired in clinical conditions, and it requires further prospective studies. It is worth noticing that the proposed approach for searching for the most valuable, from the point of view of AF detection, pRRx parameters can be repeated in other conditions and applied, for example, to specific ECG devices. Finally, pRR50 has been used in several studies as one of the HRV features incorporated in various machine learning models detecting AF. Replacing pRR50 with pRRx with a smaller  $x$  might improve the diagnostic properties of such models.

**Author Contributions:** Conceptualization, S.B., K.J. and P.G.; methodology, S.B., K.J. and P.G.; software, S.B.; validation, S.B., K.J. and P.G.; formal analysis, S.B., K.J. and P.G.; investigation, S.B., K.J. and P.G.; data curation, S.B.; writing—original draft preparation, S.B., K.J. and P.G.; writing—review and editing, S.B., K.J. and P.G.; visualization, S.B.; and supervision, K.J. and P.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the Scientific Council for the discipline of Automatic Control, Electronics, and Electrical Engineering of Warsaw University of Technology, Poland, grant “Studies on the effectiveness of atrial fibrillation detection based on the analysis of PPG signal and ECG signal using machine learning techniques, including proposals for new solutions”.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** MIT–BIH Atrial Fibrillation Database (AFDB) [35,49], and Long-Term AF Database (LTAfDB) [35,36] were used in the study. They are available at <https://physionet.org/content/afdb/1.0.0/> (accessed on 1 June 2022) and <https://physionet.org/content/ltafdb/1.0.0/> (accessed on 1 June 2022), respectively.

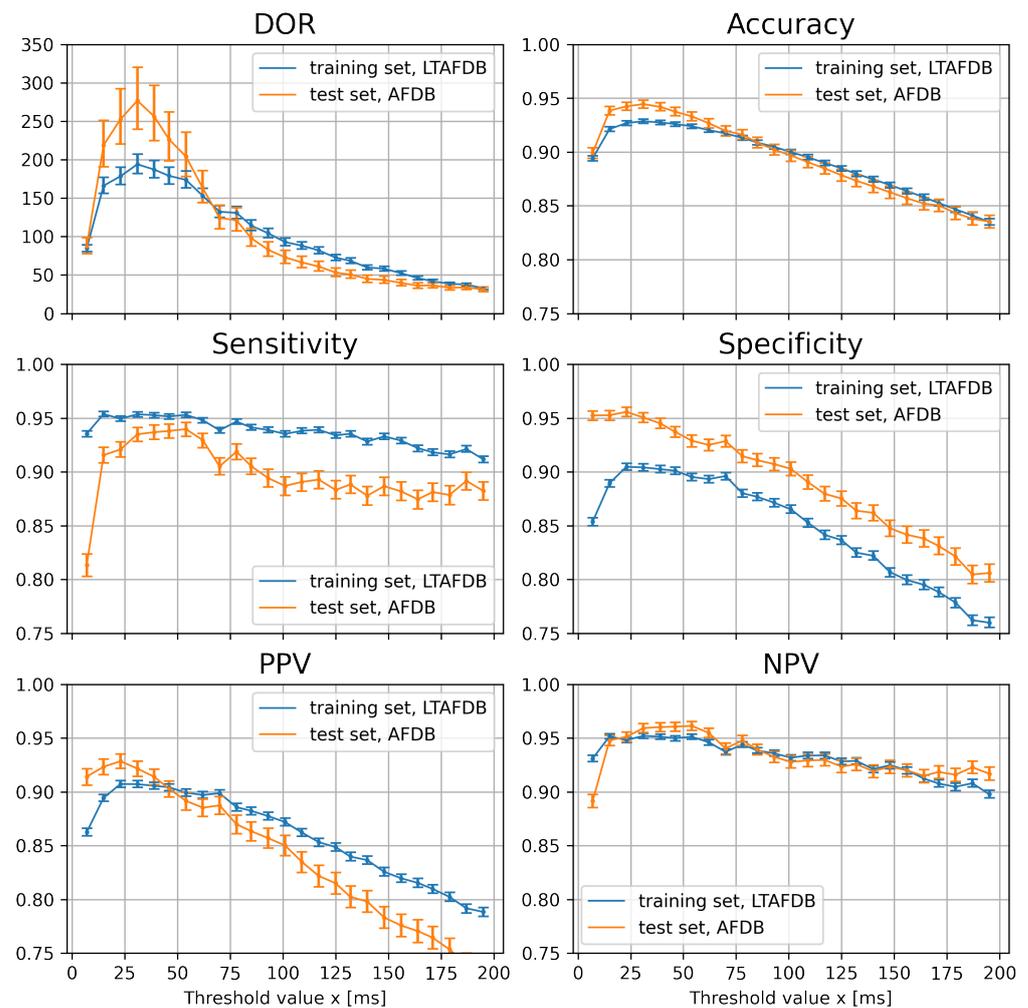
**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Results of Validation on MIT–BIH AF Database

We used the optimal cutoff values of pRRx parameters from the training set long-term Atrial Fibrillation Database (LTAfDB) to detect AF in the test set from the Massachusetts Institute of Technology–Beth Israel Hospital (MIT–BIH) Atrial Fibrillation Database (AFDB) [35,49]. Data preparation was the same as in the training set (Figure 1), and a nonparametric bootstrap with 5000 samples was used to estimate classification results’ distribution.

### Appendix A.1. Comparison between Training Set and Test Set Results

A comparison of the results of AF detection in the training set and the test set is presented in Figure A1. The classification metrics include the diagnostic odds ratio (DOR), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In the test set, the highest DOR (median 276) and accuracy (median 0.944) are achieved for pRR31 (in training set 194 and 0.928, respectively). Peak values of DOR, accuracy, PPV, NPV, and specificity are higher in the test set AFDB, while sensitivity is higher in the training set LTAfDB in the whole range of threshold values  $x$ .



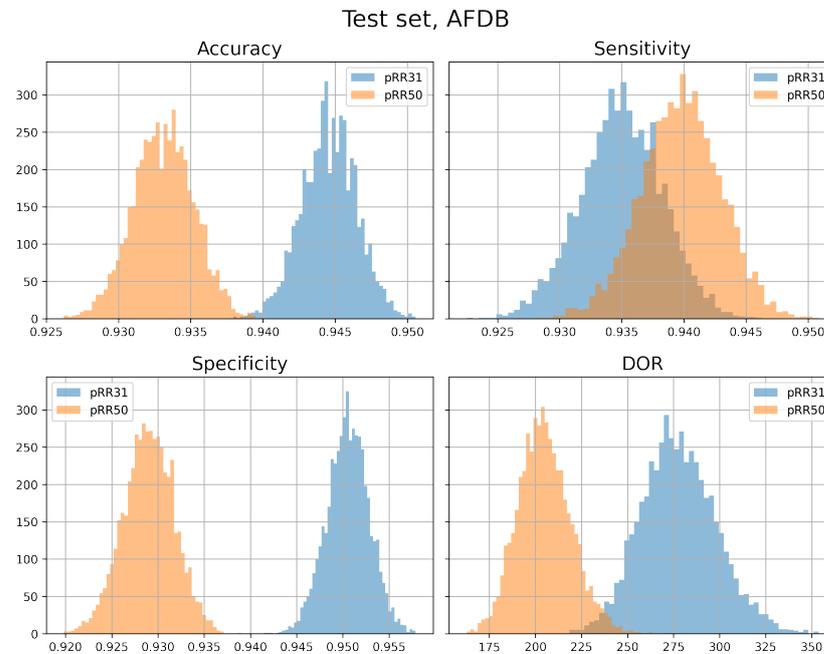
**Figure A1.** Medians and 95% confidence intervals of AF detection results in the training set Long-Term Atrial Fibrillation Database (LTAfDB) and the test set MIT-BIH Atrial Fibrillation Database (AFDB): diagnostic odds ratio (DOR, **top left**), accuracy (**top right**), sensitivity (**middle left**), specificity (**middle right**), positive predictive value (PPV, **bottom left**), and negative predictive value (NPV, **bottom right**).

The difference between performance in the training set and the test set can be a result of one or more factors, including:

- Distinct patient populations in two databases:
  - LTAfDB: 84 subjects;
  - AFDB: 25 subjects.
- Various sampling frequencies:
  - LTAfDB: 128 Hz;
  - AFDB: 250 Hz.
- The different lengths of the ECG recording:
  - LTAfDB: 24 h (night and day);
  - AFDB: 10 h (unspecified time of day).
- Unequal proportions of 60-s segments with AF and SR:
  - LTAfDB: 32,769 SR, 32,141 AF (49.5% AF);
  - AFDB: 8836 SR, 5490 AF (38.3% AF).

### Appendix A.2. Comparison between pRR31 and pRR50

The histograms in Figure A2 show the distribution of four classification metrics (accuracy, sensitivity, specificity, and DOR) in AF detection in the test set using pRR31 and pRR50. For accuracy, specificity, and DOR, the results are higher for pRR31 than for pRR50. For sensitivity, there is an overlap between the distributions for pRR31 and pRR50, but the average is higher for pRR50.



**Figure A2.** Histograms of atrial fibrillation detection results in the test set MIT–BIH Atrial Fibrillation Database (AFDB) for pRR31 and pRR50 (the percentages of successive RR intervals differing by at least 31 ms and 50 ms, respectively). DOR - diagnostic odds ratio.

Since not all the metrics had normal distribution (Shapiro–Wilk test) both for pRR31 and pRR50, we used the Wilcoxon test to compare the performance of pRR31 and pRR50. In all cases, the differences between the classification results cannot be explained by randomness.

### Appendix A.3. Conclusions from the Validation

The results presented in the Appendix A confirm the study's findings. For AF detection in 60-s ECG, the most effective threshold value  $x$  in pRR $x$  is not 50 ms (pRR50). Both in the training set LTAfDB and in the test set AFDB, pRR31 achieves the best performance, measured by accuracy and DOR. This phenomenon is not limited to a single ECG database; thus, pRR50 should not necessarily be a default pRR $x$  parameter used in AF detection.

## References

- Hindricks, G.; Potpara, T.; Dagres, N.; Arbelo, E.; Bax, J.J.; Blomström-Lundqvist, C.; Boriani, G.; Castella, M.; Dan, G.A.; Dilaveris, P.E. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS) The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. *Eur. Heart J.* **2021**, *42*, 373–498. [[CrossRef](#)] [[PubMed](#)]
- Tonko, J.B.; Wright, M.J. Review of the 2020 ESC Guidelines for the Diagnosis and Management of Atrial Fibrillation—What Has Changed and How Does This Affect Daily Practice. *J. Clin. Med.* **2021**, *10*, 3922. [[CrossRef](#)] [[PubMed](#)]
- Roten, L.; Goulouti, E.; Lam, A.; Elchinova, E.; Nozica, N.; Spirito, A.; Wittmer, S.; Branca, M.; Servatius, H.; Noti, F. Age and Sex Specific Prevalence of Clinical and Screen-Detected Atrial Fibrillation in Hospitalized Patients. *J. Clin. Med.* **2021**, *10*, 4871. [[CrossRef](#)]

4. Ble, M.; Benito, B.; Cuadrado-Godia, E.; Pérez-Fernández, S.; Gómez, M.; Mas-Stachurska, A.; Tizón-Marcos, H.; Molina, L.; Martí-Almor, J.; Cladellas, M. Left Atrium Assessment by Speckle Tracking Echocardiography in Cryptogenic Stroke: Seeking Silent Atrial Fibrillation. *J. Clin. Med.* **2021**, *10*, 3501. [[CrossRef](#)]
5. Krisai, P.; Roten, L.; Zeljkovic, I.; Pavlovic, N.; Ammann, P.; Reichlin, T.; Auf der Maur, E.; Streicher, O.; Knecht, S.; Kühne, M.; et al. Prospective Evaluation of a Standardized Screening for Atrial Fibrillation after Ablation of Cavotricuspid Isthmus Dependent Atrial Flutter. *J. Clin. Med.* **2021**, *10*, 4453. [[CrossRef](#)]
6. Boriani, G.; Palmisano, P.; Malavasi, V.L.; Fantecchi, E.; Vitolo, M.; Bonini, N.; Imberti, J.F.; Valenti, A.C.; Schnabel, R.B.; Freedman, B. Clinical Factors Associated with Atrial Fibrillation Detection on Single-Time Point Screening Using a Hand-Held Single-Lead ECG Device. *J. Clin. Med.* **2021**, *10*, 729. [[CrossRef](#)]
7. Electrophysiology Task Force of the European Society of Cardiology the North American Society of Pacing. Heart Rate Variability. Standards of measurement, physiological interpretation, and clinical use. *Circulation* **1996**, *93*, 1043–1065. [[CrossRef](#)]
8. Sagnard, A.; Guenancia, C.; Mouhat, B.; Maza, M.; Fichot, M.; Moreau, D.; Garnier, F.; Lorgis, L.; Cottin, Y.; Zeller, M. Involvement of Autonomic Nervous System in New-Onset Atrial Fibrillation during Acute Myocardial Infarction. *J. Clin. Med.* **2020**, *9*, 1481. [[CrossRef](#)] [[PubMed](#)]
9. Rizwan, A.; Zoha, A.; Mabrouk, I.B.; Sabbour, H.M.; Al-Sumaiti, A.S.; Alomainy, A.; Imran, M.A.; Abbasi, Q.H. A Review on the State of the Art in Atrial Fibrillation Detection Enabled by Machine Learning. *IEEE Rev. Biomed. Eng.* **2020**, *14*, 219–239. [[CrossRef](#)] [[PubMed](#)]
10. Khan, A.A.; Lip, G.Y.H.; Shantsila, A. Heart rate variability in atrial fibrillation: The balance between sympathetic and parasympathetic nervous system. *Eur. J. Clin. Investig.* **2019**, *49*, e13174. [[CrossRef](#)] [[PubMed](#)]
11. Krasteva, V.; Christov, I.; Naydenov, S.; Stoyanov, T.; Jekova, I. Application of Dense Neural Networks for Detection of Atrial Fibrillation and Ranking of Augmented ECG Feature Set. *Sensors* **2021**, *21*, 6848. [[CrossRef](#)] [[PubMed](#)]
12. Kareem, M.; Lei, N.; Ali, A.; Ciaccio, E.J.; Acharya, U.R.; Faust, O. A review of patient-led data acquisition for atrial fibrillation detection to prevent stroke. *Biomed. Signal Process. Control* **2021**, *69*, 102818. [[CrossRef](#)]
13. Marinucci, D.; Sbröllini, A.; Marcantoni, I.; Morettini, M.; Swenne, C.A.; Burattini, L. Artificial Neural Network for Atrial Fibrillation Identification in Portable Devices. *Sensors* **2020**, *20*, 3570. [[CrossRef](#)] [[PubMed](#)]
14. Michel, P.; Ngo, N.; Pons, J.F.; Delliaux, S.; Giorgi, R. A filter approach for feature selection in classification: Application to automatic atrial fibrillation detection in electrocardiogram recordings. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 130. [[CrossRef](#)] [[PubMed](#)]
15. Buś, S.; Jędrzejewski, K.; Guzik, P. Using Minimum Redundancy Maximum Relevance Algorithm to Select Minimal Sets of Heart Rate Variability Parameters for Atrial Fibrillation Detection. *J. Clin. Med.* **2022**, *11*, 4004. [[CrossRef](#)]
16. Parsi, A.; Glavin, M.; Jones, E.; Byrne, D. Prediction of paroxysmal atrial fibrillation using new heart rate variability features. *Comput. Biol. Med.* **2021**, *133*, 104367. [[CrossRef](#)]
17. Biton, S.; Gendelman, S.; Ribeiro, A.H.; Miana, G.; Moreira, C.; Ribeiro, A.L.P.; Behar, J.A. Atrial fibrillation risk prediction from the 12-lead electrocardiogram using digital biomarkers and deep representation learning. *Eur. Heart J.-Digit. Health* **2021**, *2*, 576–585. [[CrossRef](#)]
18. Ewing, D.J.; Neilson, J.M.; Travis, P. New method for assessing cardiac parasympathetic activity using 24 hour electrocardiograms. *Heart* **1984**, *52*, 396–402. [[CrossRef](#)]
19. Bigger Jr, J.T.; Kleiger, R.E.; Fleiss, J.L.; Rolnitzky, L.M.; Steinman, R.C.; Miller, J.P. Components of heart rate variability measured during healing of acute myocardial infarction. *Am. J. Cardiol.* **1988**, *61*, 208–215. [[CrossRef](#)]
20. Mietus, J.E.; Peng, C.K.; Henry, I.; Goldsmith, R.L.; Goldberger, A.L. The pNNx files: Re-examining a widely used heart rate variability measure. *Heart* **2002**, *88*, 378–380. [[CrossRef](#)]
21. Torres, B.; Naranjo Orellana, J. Use of pNNx statistics in the evaluation of heart rate variability at rest and during exercise. *Arch. Med. Deporte* **2010**, *27*, 255–259.
22. Saiz-Vivo, J.; Corino, V.D.; de Melis, M.; Mainardi, L.T. Unsupervised Classification of Atrial Fibrillation Triggers Using Heart Rate Variability Features Extracted from Implantable Cardiac Monitor Data. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; pp. 426–429. [[CrossRef](#)]
23. Saiz-Vivo, J.; Corino, V.D.A.; Hatala, R.; de Melis, M.; Mainardi, L.T. Heart Rate Variability and Clinical Features as Predictors of Atrial Fibrillation Recurrence After Catheter Ablation: A Pilot Study. *Front. Physiol.* **2021**, *12*, 672896. [[CrossRef](#)] [[PubMed](#)]
24. Liu, Y.; Chen, J.; Bao, N.; Gupta, B.B.; Lv, Z. Survey on atrial fibrillation detection from a single-lead ECG wave for Internet of Medical Things. *Comput. Commun.* **2021**, *178*, 245–258. [[CrossRef](#)]
25. Datta, S.; Puri, C.; Mukherjee, A.; Banerjee, R.; Dutta Choudhury, A.; Singh, R.; Ukil, A.; Bandyopadhyay, S.; Pal, A.; Khandelwal, S. Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier. In Proceedings of the 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017. [[CrossRef](#)]
26. Jiménez Serrano, S.; Yagüe Mayans, J.; Simarro-Mondéjar, E.; Calvo, C.; Castells Ramon, F.; Roig, J. Atrial Fibrillation Detection Using Feedforward Neural Networks and Automatically Extracted Signal Features. In Proceedings of the 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017. [[CrossRef](#)]
27. Teijeiro, T.; Garcia, C.A.; Castro, D.; Félix, P. Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG Records. In Proceedings of the 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017. [[CrossRef](#)]

28. Clifford, G.D.; Liu, C.; Moody, B.; Li-wei, H.L.; Silva, I.; Li, Q.; Johnson, A.E.; Mark, R.G. AF Classification from a Short Single Lead ECG Recording: The PhysioNet/Computing in Cardiology Challenge 2017. In Proceedings of the 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017; pp. 1–4.
29. Jovic, A.; Bogunovic, N. Evaluating and comparing performance of feature combinations of heart rate variability measures for cardiac rhythm classification. *Biomed. Signal Process. Control* **2012**, *7*, 245–255. [[CrossRef](#)]
30. Moody, G.; Mark, R. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50. [[CrossRef](#)] [[PubMed](#)]
31. Buś, S.; Jędrzejewski, K.; Guzik, P. A Study on Selection of HRV-based Features for Different Classifiers in Atrial Fibrillation Detection. In Proceedings of the 2021 Signal Processing Symposium (SPSympo), Lodz, Poland, 20–23 September 2021; pp. 31–34. [[CrossRef](#)]
32. Fawcett, T. An introduction to ROC analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [[CrossRef](#)]
33. Conroy, T.; Guzman, J.H.; Hall, B.; Tsouri, G.; Couderc, J.P. Detection of atrial fibrillation using an earlobe photoplethysmographic sensor. *Physiol. Meas.* **2017**, *38*, 1906–1918. [[CrossRef](#)]
34. Ramesh, J.; Solatidehkordi, Z.; Aburukba, R.; Sagahyroon, A. Atrial Fibrillation Classification with Smart Wearables Using Short-Term Heart Rate Variability and Deep Convolutional Neural Networks. *Sensors* **2021**, *21*, 7233. [[CrossRef](#)]
35. 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](#)]
36. Petrutiu, S.; Sahakian, A.V.; Swiryn, S. Abrupt changes in fibrillatory wave characteristics at the termination of paroxysmal atrial fibrillation in humans. *EP Eur.* **2007**, *9*, 466–470. [[CrossRef](#)] [[PubMed](#)]
37. Wilcoxon, F. Individual Comparisons by Ranking Methods. *Biom. Bull.* **1945**, *1*, 80–83. [[CrossRef](#)]
38. Mann, H.B.; Whitney, D.R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
39. Spearman, C. The Proof and Measurement of Association between Two Things. *Am. J. Psychol.* **1904**, *15*, 72–101. [[CrossRef](#)]
40. Youden, W.J. Index for rating diagnostic tests. *Cancer* **1950**, *3*, 32–35. [[CrossRef](#)]
41. Butkuvienė, M.; Petrėnas, A.; Sološenko, A.; Martín-Yebra, A.; Marozas, V.; Sörnmo, L. Considerations on performance evaluation of atrial fibrillation detectors. *IEEE Trans. Biomed. Eng.* **2021**, *68*, 3250–3260. [[CrossRef](#)]
42. Chinchor, N.; Sundheim, B.M. MUC-5 evaluation metrics. In Proceedings of the Fifth Message Understanding Conference (MUC-5), Baltimore, MD, USA, 25–27 August 1993.
43. Glas, A.S.; Lijmer, J.G.; Prins, M.H.; Bossel, G.J.; Bossuyt, P.M.M. The diagnostic odds ratio: A single indicator of test performance. *J. Clin. Epidemiol.* **2003**, *56*, 1129–1135. [[CrossRef](#)]
44. Efron, B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [[CrossRef](#)]
45. Nguyen, A.; Ansari, S.; Hooshmand, M.; Lin, K.; Ghanbari, H.; Gryak, J.; Najarian, K. Comparative Study on Heart Rate Variability Analysis for Atrial Fibrillation Detection in Short Single-Lead ECG Recordings. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 526–529. [[CrossRef](#)]
46. Christov, I.; Krasteva, V.; Simova, I.; Neycheva, T.; Schmid, R. Ranking of the most reliable beat morphology and heart rate variability features for the detection of atrial fibrillation in short single-lead ECG. *Physiol. Meas.* **2018**, *39*, 094005. [[CrossRef](#)]
47. Kim, D.; Seo, Y.; Jung, W.R.; Youn, C.H. Detection of Long Term Variations of Heart Rate Variability in Normal Sinus Rhythm and Atrial Fibrillation ECG Data. In Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics, Sanya, China, 27–30 May 2008; pp. 404–408. [[CrossRef](#)]
48. Rubin, J.; Parvaneh, S.; Asif, R.; Conroy, B.; Babaeizadeh, S. Densely Connected Convolutional Networks and Signal Quality Analysis to Detect Atrial Fibrillation Using Short Single-Lead ECG Recordings. In Proceedings of the 2017 Computing in Cardiology (CinC), Rennes, France, 24–27 September 2017. [[CrossRef](#)]
49. Moody, G. A new method for detecting atrial fibrillation using RR intervals. *Comput. Cardiol.* **1983**, *10*, 227–230.