






Article

# Detection of Atrial Fibrillation Episodes in Long-Term Heart Rhythm Signals Using a Support Vector Machine

Robert Czabanski <sup>1</sup>, Krzysztof Horoba <sup>2,\*</sup>, Janusz Wrobel <sup>2</sup>, Adam Matonia <sup>2</sup>,  
Radek Martinek <sup>3</sup>, Tomasz Kupka <sup>2</sup>, Michal Jezewski <sup>1</sup>, Radana Kahankova <sup>3</sup>,  
Janusz Jezewski <sup>2</sup> and Jacek M. Leski <sup>1</sup>

<sup>1</sup> Department of Cybernetics, Nanotechnology and Data Processing, Silesian University of Technology, PL44100 Gliwice, Poland; robert.czabanski@polsl.pl (R.C.); michal.jezewski@polsl.pl (M.J.); jleski@polsl.pl (J.L.)

<sup>2</sup> Łukasiewicz Research Network–Institute of Medical Technology and Equipment, PL 41800 Zabrze, Poland; januszw@itam.zabrze.pl (J.W.); adamm@itam.zabrze.pl (A.M.); tomekk@itam.zabrze.pl (T.K.); janusz.jezewski@itam.zabrze.pl (J.J.)

<sup>3</sup> Department of Cybernetics and Biomedical Engineering, VSB–Technical University of Ostrava, 708 00 Ostrava-Poruba, Czech Republic; radek.martinek@vsb.cz (R.M.); radana.kahankova@vsb.cz (R.K.)

\* Correspondence: krzysztof.horoba@itam.zabrze.pl

Received: 19 December 2019; Accepted: 27 January 2020; Published: 30 January 2020



**Abstract:** Atrial fibrillation (AF) is a serious heart arrhythmia leading to a significant increase of the risk for occurrence of ischemic stroke. Clinically, the AF episode is recognized in an electrocardiogram. However, detection of asymptomatic AF, which requires a long-term monitoring, is more efficient when based on irregularity of beat-to-beat intervals estimated by the heart rate (HR) features. Automated classification of heartbeats into AF and non-AF by means of the Lagrangian Support Vector Machine has been proposed. The classifier input vector consisted of sixteen features, including four coefficients very sensitive to beat-to-beat heart changes, taken from the fetal heart rate analysis in perinatal medicine. Effectiveness of the proposed classifier has been verified on the MIT-BIH Atrial Fibrillation Database. Designing of the LSVM classifier using very large number of feature vectors requires extreme computational efforts. Therefore, an original approach has been proposed to determine a training set of the smallest possible size that still would guarantee a high quality of AF detection. It enables to obtain satisfactory results using only 1.39% of all heartbeats as the training data. Post-processing stage based on aggregation of classified heartbeats into AF episodes has been applied to provide more reliable information on patient risk. Results obtained during the testing phase showed the sensitivity of 98.94%, positive predictive value of 98.39%, and classification accuracy of 98.86%.

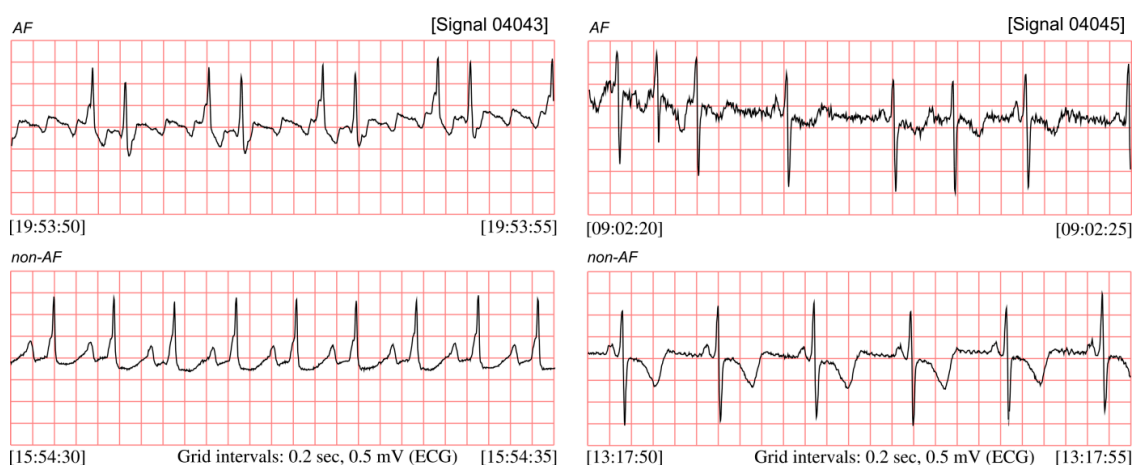
**Keywords:** support vector machine (SVM); heart rate variability (HRV); HRV features; atrial fibrillation (AF); AF detection

## 1. Introduction

Atrial fibrillation (AF) is the most common heart arrhythmia, which occurs when the atria contracts quickly and irregularly at rates of 400 to 600 per minute. These contractions are independent from ventricles, which themselves operate at much lower rate. AF symptoms often include palpitations, irregular heartbeat, shortness of breath, chest pains and others, but they can be also asymptomatic and is then called silent AF. The frequency of AF occurrence is strictly correlated with the patient's age [1,2]. The prognosis indicates that the AF occurrence within the period of the next 20–30 years will

double, mainly due to the longer life span of the population. The AF detection is important, since this heart arrhythmia is a well-known risk factor for occurrence of ischemic stroke, even six times higher than among patients without the arrhythmia [3].

Figure 1 presents the ECG signals from the MIT-BIH Atrial Fibrillation database (MIT-BIH AF) published on PhysioNet [4–6], comprising both segments with the AF episodes and non-AF segments. AF episodes occur irregularly and may last from a few heartbeats to hours, which significantly hinder the possibility to diagnose the silent AF by means of occasionally performed ambulatory ECG recordings. It implies that the longer the recording, the higher chance to detect the silent AF episodes [7,8]. The most efficient techniques of long-term monitoring are: Holter monitor, continuous telemetry [9–12], or implementable devices with internal memory [13–15]. However, visual analysis of long 24-h recording requires a lot of time and efforts from the cardiologists, thus the methods for automated detection of atrial fibrillation are needed to improve the objectivity of interpretation. When based on ECG, the efficient automated AF detection requires a high quality signal. It may not be ensured by the long-term monitoring techniques which usually comprise periods of daily physical activity of the patient which distort the ECG signal.



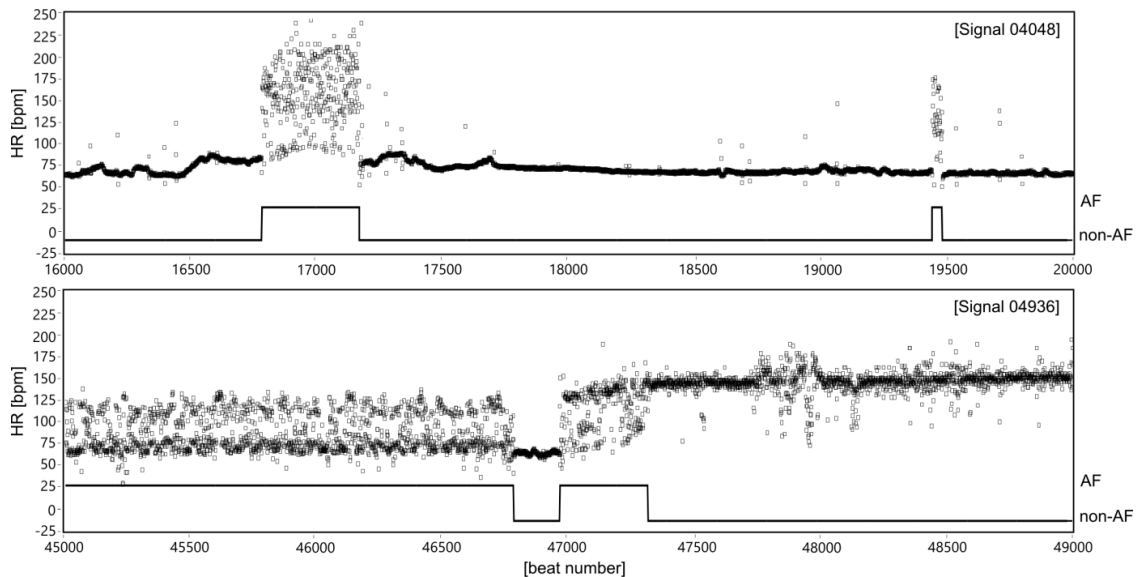
**Figure 1.** ECG signals (04043 and 04045) taken from the MIT-BIH AF database, with recognized segments of atrial fibrillation and the non-AF ones.

The AF episode is manifested in ECG by significant changes of duration of the beat-to-beat (RR) intervals [16–19], see Figure 1. However, the RR intervals irregularity caused by AF occurrences is much more easy to observe after converting RR intervals into the instantaneous heart rate (HR) signal (Figure 2). The presented HR signals confirm that AF episodes occur very accidentally, and they can last a few seconds (signal 04048), but also expand to long lasting episodes (signal 04936).

In the light of above facts an efficient automated method for AF detection should be based on estimation of RR irregularity or equivalent i.e., HR irregularity observed in long-term recording [20,21]. Moreover, such approach enables to involve the various recording methods which can provide signals in which the heartbeats can be detected. Beside electrocardiogram, such signals include photoplethysmogram [22–24] or seismocardiogram [25]. Using a photoelectric sensor is attractive in case of home telecare as long-term recording should be accomplished by instrumentations being minimally troublesome and inconvenient to the patient [26,27]. It may be a smart monitor in a form of a wrist bracelet with a specialized reflective optical sensor to perform the heart rate monitoring using the method previously developed by the authors [28].

The general concept of the methods most commonly used for automated detection of AF episodes relies on determination of features estimating the RR interval changes, and then application of the statistical analysis or more advanced classifier to differentiate between AF episode and normal sinus rhythm segments, basing on the information on RR irregularity. The feature set is composed most commonly of different statistical measures (mean or median HR, root mean square of successive RR

differences, turning point ratio). It can also include normalized RR intervals [29,30] or normalized RR differences [31], Shannon entropy [19] or coefficient of sample entropy [15,32]. Other form to present the RR irregularity are: the density histogram of the difference between successive RR intervals [33,34], map that plots RR intervals versus change of RR intervals [35], mapping the RR-interval time series to binary symbolic sequences [36,37] or Markov score of RR interval [16].



**Figure 2.** Two HR signals expressed in beats per minute (bpm) with clinically recognized AF segments of different characteristics of HR changes in relation to normal sinus rhythm (non-AF). The AF segments are marked using the experts' annotations provided for particular records in the MIT-BIH AF database.

In the simplest approach to AF classification the Receiver Operating Characteristics (ROC) curve has been used to find the optimal threshold values for the input features providing the best classification performance [30,35,36,38]. The statistical test (Kolmogorov-Smirnov) was used in [33] to check if the density histograms of the test data differ from the standard density ones prepared as a template of AF episodes. In order to differentiate between AF and non-AF patterns the various classification methods have been applied: Neyman-Pearson detector [31], Random Forest (RF) model and k-nearest neighbors classifier [32], Support Vector Machine (SVM) with promising results reported in [39–41], as well as artificial neural network [42], also with interval transition matrices as an input [43].

In [39] SVM approach was used for classification of the 30-s segments of ECG and 300-beat sequences of RR intervals. Two parameters of Stationary Wavelet Transform (peak-to-average power ratio and log-energy entropy) were used for raw ECG-based approach, while five features were extracted from HR signal. The efficiency of AF detection achieved by the feature-based classification of the RR sequences was tested against the algorithm based on raw ECG. Higher sensitivity was ensured by the HR-based approach, while ECG-based algorithm provided improved specificity and classification accuracy. The classifier based on SVM with radial basis function was proposed in [40], with two features as the inputs: the average of RR differences and the standard deviation of differences in a defined duration. The same SVM classifier was employed in [41]. The input set comprised more RR interval features: median heart rate, minimum RR interval, mean RR interval, various entropy measures, and difference irregularity measure.

The features estimating the RR variability are calculated in a sliding window comprising an established number of consecutive RR intervals (or HR values). Since there is no standard for the window length, many works have aimed to find the optimal length, providing the best classification performance. Some works assumed that AF episodes of less than 30 s duration are not clinically significant, which led to higher optimal number of heartbeats: 100 [33], 128 [35] and 150 [37]. Other authors claim that longer windows tend to miss short AF episodes, and thus they applied significantly

shorter windows: 30 [32], 12 [15] or even 8 beats [44]. It is obvious that different window length reported as the optimal value depends on the method used for automated AF detection. Another important aspect of using the sliding window for AF detection is how many beats it is shifted. Shifting the window every heartbeat results in one beat resolution of the AF classification. Then each heartbeat (RR interval), usually corresponding to the middle of the window, is classified as AF or non-AF. In such case, determination of classification performance is evident as each automatically classified beat can be related to the reference one, basing on the expert annotations. Otherwise, additional condition has to be applied—the window is labeled as AF episode only if the number of clinically annotated AF beats within the window exceeds a predefined threshold, usually 0.5 like in [35,45]. However, it is obvious that the threshold value affects the classification performance. The threshold has been included into the input feature set and tuned for optimum sensitivity and specificity in [38]. However, it should be noted that in such approach, the reference information is modified to achieve the best classification performance of the automated method tested, which seems to be rather doubtful.

In order to avoid short false positive AF episodes or short artifact of classified AF the post-processing correction was applied, like dedicated mechanism called AF alarm enhancer [16]. It is the hysteresis counter that begins (or ends) an episode if established number of consecutive analyzed RR segments have been classified as AF (or non-AF). Other post-processing method was based on median filtering [45].

In [16], after combining R-R interval Markov score with two P-wave measurements: the location expressed by P-R interval duration, and the morphology defined as similarity between two consecutive P-waves, the sensitivity did not change, whereas specificity and positive predictive values increased slightly.

A novel deep learning has been adopted for automated detection of AF in the long-term ECG recordings. This classifier learns directly from the RR intervals and therefore there is no need to extract the features. The model based on deep Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) was used in [46], and combining with the Convolutional- and Recurrent-Neural Networks to extract high level features was proposed in [45]. Although a high classification performance has been reported, the computational complexity of deep learning model is much higher than traditional feature-based classifier. In this paper, we describe the method for automated AF detection which assigns the vector of parameters quantitatively describing the HR signal into two classes representing the absence or presence of atrial fibrillation. As estimation of HR variability is also important part of the Fetal Heart Rate (FHR) analysis [47–49], the indices widely used for FHR variability description have been considered as potentially useful for AF detection. The detection method presented in this paper was derived from the machine learning principles. Our classification routine was performed by means of the Lagrangian Support Vector Machine (LSVM) [50]—the state-of-the-art classifier based on the linearly convergent learning algorithm. The efficient LSVM learning procedure was obtained from the reformulation of the Quadratic Programming (QP) optimization problem of the Support Vector Machine (SVM) [51]. Additional aggregation stage has been applied to provide more reliable information on risk for the patient. The performance of the proposed AF detection method was examined using the MIT-BIH Atrial Fibrillation database, which includes 25 ten-hour long ECG recordings.

## 2. Materials and Methods

Automated detection of the atrial fibrillation episodes proposed in this work starts with extraction of sixteen HR irregularity features composing the classifier input vector. Then the LSVM classifier is applied to mark a given heartbeat as AF or non-AF one. Final step is aggregation of the classified beats into AF episodes.

### 2.1. HR Irregularity Features

Considering on-line detection of AF and limited computational power of the developed mobile monitor, we applied a simple linear classifier which recognizes the AF heartbeats basing on easily accessible information about heart rhythm and HR features [52,53]. Apart from the HR value, other

four input features have been selected in a series of preliminary investigations carried out among larger feature set [54]. Having the information on heartbeats detected, the instantaneous heart rate values  $HR_i$  (expressed in beats per minute) are calculated according to the formula:

$$HR_i[\text{bpm}] = \frac{60000}{RR_i[\text{ms}]}, \quad (1)$$

where:  $RR_i$  is the  $i$ -th interval between two consecutive heart beats expressed in milliseconds.

Next, the features are determined in symmetrical moving window comprising 21 of  $HR_i$  values:

- $MED_i = \text{median}\{HR_{i-k}, \dots, HR_{i+k}\};$
- $MAD_i = \text{median}\{x_{i-k}, \dots, x_{i+k}\},$  where  $x_i = |HR_i - MED_i|;$
- $QNT_i$ —represents the quantile of order 0.7 estimated over 21 values of heart rate;
- $PRP_i$ —is the ratio of number of HR values between thresholds level of 120 to 160 bpm, to total number.

where:  $i$  is the number of consecutive heartbeats to be classified, and  $k = 1 \dots 10$ .

The values of the additional parameters: window width  $N$  set to 21, quantile order set to 0.7 and HR thresholds of 120 and 160 were determined as a result of previously performed experiments [54].

For the new classification method, the input vector has been significantly expanded. It additionally comprises seven measures obtained from classical analysis of HR variability used in adults' electrocardiography. This analysis includes exclusively sinus excitation, i.e., generated by the sinus-atrial node. Thus, it concerns only sinus rhythm variability, and any other types of excitation are excluded and replaced with artificially generated beats. Corrected in this way the series of changes in the subsequent RR intervals become the basis for the determination of heart rate variability measures. The most commonly used quantitative analysis methods can be divided into time, frequency, time-frequency and non-linear methods. In the presented work, the indices describing the HR variability were used in an unusual way as a set of features allowing the detection of atrial fibrillation episodes. The four selected features, obtained in statistical analysis in time domain within the same moving window, are as follows:

- The mean heart rate:

$$\overline{HR}[\text{bpm}] = \frac{1}{N} \sum_{i=1}^N HR_i. \quad (2)$$

- Standard deviation of instantaneous heart rate values:

$$\text{STD\_HR}[\text{bpm}] = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\overline{HR} - HR_i)^2}. \quad (3)$$

- Root Mean Square of Successive Differences (RMSSD) which measures the variability within a data set—RR intervals—according to the following equation:

$$\text{RMSSD}[\text{ms}] = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (RR_{i+1} - RR_i)^2}. \quad (4)$$

- Percentage of differences between the RR intervals that exceed the value of 50 ms, denoted as pNN50 [%]:

$$\text{pNN50} = \frac{\sum_{i=1}^{N-1} A_i}{N-1} * 100\%, \quad (5)$$

where:

$$A_i = \begin{cases} 1 & \text{when } |RR_{i+1} - RR_i| > 50 \\ 0 & \text{when } |RR_{i+1} - RR_i| \leq 50 \end{cases} \quad (6)$$

In addition, three non-linear features of HRV analysis were applied in the form of:

- Poincare graph, which is a graphical representation of the current interval  $RR_i$  plotted against subsequent one  $RR_{i+1}$ . Using the ellipse fitting technique, in each moving window comprising 21 heartbeats, two standard deviations are determined from the points: perpendicular to the regression line (SD1) and along the line (SD2). The SD1 describes the short-term variability of the heart rhythm, while the SD2 refers to the long-term HR variability.
- Turning Points Ratio (TPR) measures the randomness of fluctuations within a data set, by calculating the ratio of the number of turning points to the maximum number of possible turning points. Turning point is found if both the preceding and succeeding points are either greater or lower. It is expected in random data set of arbitrary length  $N$ , that the number of possible turning points is  $(2N - 4)/3$ , with a standard deviation of  $\sqrt{(16N - 29)/90}$ .

A separate group of features used for the detection of AF episodes are parameters commonly used in fetal heart rate analysis [55,56]. It turns out that in perinatal medicine quite different features are used to describe the FHR variability, mainly short-term (beat-to-beat) [57]. For the detection of AF episodes, four widely known short-term coefficients (indices) have been selected [58,59]. They are characterized by high sensitivity to changes in subsequent values of RR intervals and thus they potentially may be useful for AF detection [60–62]:

- The Yeh's index (DI\_Yeh) whose determination starts with calculation of the auxiliary values  $d_i$  representing the ratio of the difference between two successive RR intervals to their sum:

$$d_i = \frac{RR_i - RR_{i+1}}{RR_i + RR_{i+1}}. \quad (7)$$

Then, for the analyzed signal fragment, the DI\_Yeh index is defined as the standard deviation from the obtained coefficients  $d_i$ :

$$DI\_Yeh [ms] = \sqrt{\frac{1}{N-2} \sum_{i=1}^{N-1} (d_i - \bar{d})^2}, \quad (8)$$

where:  $\bar{d} = \frac{1}{N-1} \sum_{i=1}^{N-1} d_i$ ,  $N$ —number of beats set to 21.

- The Zugaib's variability index (STV\_Zug) has been defined as an average of the absolute values of the differences between successive  $D_i$  values and their median value:

$$STV\_Zug [ms] = \frac{1}{N-1} \sum_{i=1}^{N-1} |D_i - Med|, \quad (9)$$

where: Med—median value for the  $D_i$  series,  $N$ —number of beats set to 21. The  $D_i$  value represents the ratio of the absolute value of the difference between the heart intervals RR to their sum:

$$D_i = \frac{|RR_{i+1} - RR_i|}{RR_{i+1} + RR_i}, \quad (10)$$

- The Huey's index (STV\_Huey) was defined as the sum of absolute values of differences of subsequent instantaneous HR values for which the sign of difference was changed:

$$\text{STV\_Huey [bpm]} = \sum_{i=2}^{N-1} k \cdot |\text{HR}_{i+1} - \text{HR}_i| \quad (11)$$

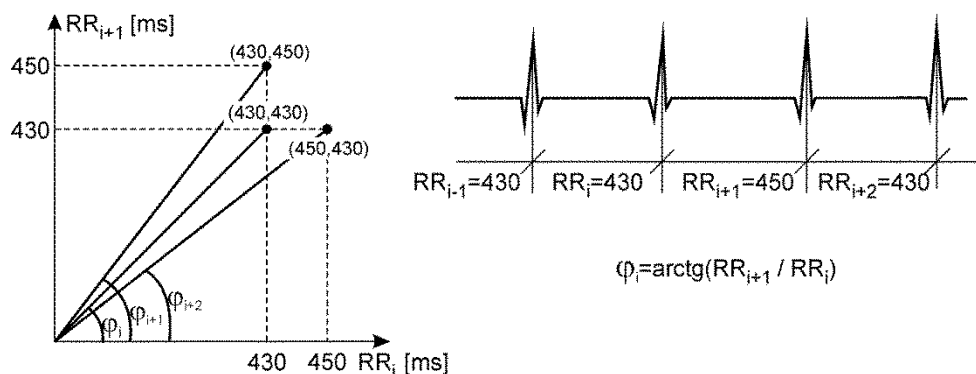
where:

$$k = \begin{cases} 1 & \text{for } (\text{HR}_{i-1} - \text{HR}_i) \cdot (\text{HR}_i - \text{HR}_{i+1}) < 0 \\ 0 & \text{for } (\text{HR}_{i-1} - \text{HR}_i) \cdot (\text{HR}_i - \text{HR}_{i+1}) \geq 0. \end{cases} \quad (12)$$

- The definition of de Haan's index (STI\_Haan) is based on a polar coordinate system whose both axes refer to RR intervals expressed in milliseconds, and points represent the pairs of subsequent intervals ( $\text{RR}_{i-1}$ ,  $\text{RR}_i$ ), as shown in Figure 3. STI\_Haan is determined as the interquartile range of the angles  $\varphi_i$  between the lines connecting the point with origin of the coordinate system, and the X axis, designated for subsequent periods  $\text{RR}_i$ :

$$\text{STI\_Haan} = \text{IQR}(\varphi_i), \quad (13)$$

where:  $i = 1, 2 \dots N, N$ —number of beats.



**Figure 3.** Distribution of successive  $\text{RR}_i$  intervals in the polar coordinate system, illustrating the definitions of the de Haan's index describing the short-term HR variability.

## 2.2. LSVM Classifier

The proposed method for automated recognition of AF episodes is based on a machine learning approach. To achieve high accuracy of AF detection, we applied the classification routine that originates from the Statistical Learning Theory (SLT) [63]. The SLT is the base for the machine learning methods which are characterized by a high generalization ability, meaning the high efficiency when evaluating previously unknown data i.e., data that have not been used when designing the classifier (also called as classifier training or learning). One of the major achievements of SLT is the Structural Risk Minimization (SRM) principle, which states that the quality of machine learning depends both on the empirical data and the complexity of the model. The most-known practical implementation of the SRM is the Support Vector Machine (SVM) methodology [51,64,65]. The SVM allows for finding the hyperplane in the input feature space which divides the considered classes with the widest margin of separation. The input data that are used to define the margin are called the support vectors. The original SVM algorithm was formulated as a linearly constrained quadratic optimization problem. Consequently, the learning procedure of high computational complexity was obtained [66–68]. As the low computational cost of the detection method is of our special interest, in the proposed solution the Lagrangian Support Vector Machine (LSVM) [50] was applied. Its learning replaces the quadratic programming with the linearly convergent iterative algorithm which results in significant reduction of the computational complexity and higher efficiency when compared to the original SVM [50].

Let us consider a training set  $\mathcal{L}$ , which contains  $N_{\text{TRN}}$  vectors  $\mathbf{x}_0(1), \mathbf{x}_0(2), \dots, \mathbf{x}_0(N_{\text{TRN}}) \in \mathbb{R}^t$  of  $t$  parameters quantitatively describing the HR signal, and the corresponding output value  $y_0(1), y_0(2), \dots, y_0(N_{\text{TRN}}) \in \{-1, 1\}$  defining the absence (non-AF)  $y_0(n) = -1$  or the presence  $y_0(n) = 1$  of the AF episode. The linear SVM classification problem of  $\mathcal{L}$  can be formulated as the constrained minimization:

$$\min_{\mathbb{R}^N \times \mathbb{R}^N} f_{\text{SVM}}(\mathbf{w}, \boldsymbol{\xi}) = \frac{\mathbf{w}^T \mathbf{w}}{2} + \gamma \mathbf{1}^T \boldsymbol{\xi}, \quad (14)$$

subject to the condition:

$$\mathbf{D}(\mathbf{X}_0 \mathbf{w} - \mathbf{1} w_0) + \boldsymbol{\xi} \geq \mathbf{1}, \quad (15)$$

and:

$$\boldsymbol{\xi} \geq \mathbf{0}, \quad (16)$$

where  $\mathbf{w} \in \mathbb{R}^t$  and  $w_0 \in \mathbb{R}$  are the parameters of two bounding planes:

$$\begin{cases} \mathbf{x}^T \mathbf{w} - w_0 = +1, \\ \mathbf{x}^T \mathbf{w} - w_0 = -1. \end{cases} \quad (17)$$

separating the training data with the margin  $\frac{2}{\|\mathbf{w}\|}$ ,  $\gamma \geq 0$  is a constant that controls the trade-off between model simplicity and model matching to the training data,  $\mathbf{1} \in \mathbb{R}^N$  denotes the vector with all entries equal to one,  $\boldsymbol{\xi} \in \mathbb{R}^N$  is the vector of the slack (error) variables, that allow the classes to be bounded with the maximum "soft" margin i.e., with the minimum sum of deviations of training errors and maximum margin for the correctly classified vectors,  $\mathbf{D} = \text{diag}(y_0(1), y_0(2), \dots, y_0(N)) \in \mathbb{R}^{N \times N}$  is a diagonal matrix with class labels along its diagonal,  $\mathbf{X}_0 = \begin{bmatrix} \mathbf{x}_0^T(1) & \mathbf{x}_0^T(2) & \dots & \mathbf{x}_0^T(N) \end{bmatrix}^T \in \mathbb{R}^{N \times t}$  is the matrix of training input data.

In contrast to SVM, the Lagrange support vector machine maximizes the margin between the separating planes with respect to both orientation ( $\mathbf{w}$ ) and location of the planes ( $w_0$ ). Moreover, in the LSVM criterion function the sum of the slack variables  $\mathbf{1}^T \boldsymbol{\xi}$  (14) is replaced with the sum of squares  $\boldsymbol{\xi}^T \boldsymbol{\xi}$  making the constraint (16) redundant. Consequently, the linear LSVM is defined as minimization problem of the functional:

$$\min_{\mathbb{R}^t \times \mathbb{R} \times \mathbb{R}^N} f_{\text{LSVM}}(\mathbf{w}, b, \boldsymbol{\xi}) = \frac{1}{2}(\mathbf{w}^T \mathbf{w} + w_0^2) + \frac{\gamma}{2} \boldsymbol{\xi}^T \boldsymbol{\xi}, \quad (18)$$

subject to the constraint (17). Moreover, the dual problem of (20):

$$\min_{\mathbb{R}_+^N} f_{\text{dLSVM}}(\boldsymbol{\lambda}) = \frac{1}{2} \boldsymbol{\lambda}^T \mathbf{Q} \boldsymbol{\lambda} - \mathbf{1}^T \boldsymbol{\lambda}, \quad (19)$$

where:  $\mathbf{Q} = \frac{1}{\gamma} + \mathbf{H} \mathbf{H}^T \in \mathbb{R}^{N \times N}$ ,  $\mathbf{H} = \mathbf{D} \begin{bmatrix} \mathbf{X}_0 & -\mathbf{1} \end{bmatrix} \in \mathbb{R}^{N \times (t+1)}$  and  $\mathbf{I} \in \mathbb{R}^{N \times N}$  is the identity matrix, has the non-negativity constrain only  $\boldsymbol{\lambda} \in \mathbb{R}_+^N$ .

The solution can be determined based on the Karush-Kuhn-Tucker necessary and sufficient optimality conditions [50]. This leads to a linearly convergent iterative scheme which constitutes the LSVM method:

$$\boldsymbol{\lambda}^{(k+1)} = \mathbf{Q}^{-1} \left( \mathbf{1} + \left( \mathbf{Q} \boldsymbol{\lambda}^{(k)} - \mathbf{1} - \alpha \boldsymbol{\lambda}^{(k)} \right)_+ \right), \quad (20)$$

where:  $k$  is the iteration index and  $\left( \mathbf{Q} \boldsymbol{\lambda}^{(k)} - \mathbf{1} - \alpha \boldsymbol{\lambda}^{(k)} \right)_+ \in \mathbb{R}^N$  is the vector with all of its negative components set to zero.



The above algorithm is convergent for any starting point if:

$$0 < \alpha < \frac{2}{\gamma}. \quad (21)$$

The parameters of the bounding planes that separate the classes can be recovered from the solution of the dual problem by using the following formulas:

$$\begin{cases} \mathbf{w} = \mathbf{X}_0^T \mathbf{D} \lambda, \\ w_0 = -\mathbf{1}^T \mathbf{D} \lambda, \\ \xi = \frac{\Delta}{\gamma}. \end{cases} \quad (22)$$

The LSVM approach reduces significantly the time necessary to perform calculations for the optimal  $(\mathbf{w}, w_0)$  while preserving high classification efficiency of the original SVM learning.

The basic LSVM is a linear classifier, thus to handle the non-linearly separable data the so-called “kernel trick” is required. It is based on the premise that the complex non-linear classification problem will be linearly separable in some feature space of higher dimensionality and involves the non-linear transformation of input data in the high-dimensional space. The linear separating plane  $\mathbf{x}^T \mathbf{w} - w_0 = 0$  is then replaced by the non-linear surface:

$$K(\mathbf{x}_e^T, \mathbf{x}_{0e}^T) \mathbf{D} \lambda = 0, \quad (23)$$

where:  $\mathbf{x}_e = \left[ \mathbf{x}^T; -1 \right]^T$ ,  $\mathbf{x}_{0e} = \left[ \mathbf{x}_0^T; -1 \right]^T$ , and  $K$  is the kernel function. Redefinition of the dual problem (19) by using:

$$\mathbf{Q} = \frac{\mathbf{I}}{\gamma} + \mathbf{D}K(\mathbf{X}_{0e}, \mathbf{X}_{0e}^T)\mathbf{D}, \quad (24)$$

which makes the LSVM iterative schema (20) valid for any positive semidefinite kernel  $K$  [50]. In the proposed approach we used the radial (Gaussian) kernel:

$$K(\mathbf{x}, \mathbf{y}) = \exp(-\chi \|\mathbf{x} - \mathbf{y}\|^2), \text{ where } \chi > 0. \quad (25)$$

### 2.3. Performance Evaluation

The performance (generalization ability) of the AF classification was evaluated by the classification accuracy (CA), defined as the percentage of correctly classified cases in the testing set (data which was not used during classifier training). As the AF detection process is a kind of diagnostic test giving negative or positive results, we also measured the classification quality using sensitivity (Se), specificity (Sp), positive (PPV) and negative (NPV) predictive value, calculated for the testing data set using a confusion matrix. Since evaluation of the classification efficiency is difficult when analyzing all the prognostic measures simultaneously, we calculated also the F-Score (FS), defined as a harmonic mean of Se and PPV:

$$FS [\%] = \frac{2 \cdot Se \cdot PPV}{Se + PPV}. \quad (26)$$

### 2.4. Heartbeat Aggregation

The aggregation of the classified heartbeats should lead to removal of accidental changes of heartbeat status, and thus to obtain more reliable information on AF episodes. This process is controlled by two parameters: the window width and percentage threshold. Each heartbeat status is validated in symmetrical window by checking if the number of the heartbeats with the same status exceeds the percentage threshold. We defined the percentage threshold for the AF status, as the number of heartbeats classified as AF to the number of all beats in the analyzed window. If the threshold is exceeded the AF status remains unchanged, otherwise is set to non-AF. The optimal values of the

control parameters (window width and the percentage threshold) have been found to ensure the best AF detection performance expressed by the maximal value of the F-Score.

### 2.5. Material

To verify the effectiveness of the proposed AF detection method we have used the MIT-BIH Atrial Fibrillation database (MIT-BIH AF) [4,5], which includes 25 ECG signals, each of 10 h in duration. Of these, 23 ECG signals are accompanied by time markers of detected QRS complexes, while two signals are represented only by information on heart rate, which however is enough for this study. The database contains a total of 1,221,534 heartbeats, with 519,788 annotated as AF. However, when using the window of 21-beat width shifted with one beat, first and last ten heartbeats in each signal were excluded as the HR features were not determined for those heartbeats in incomplete window. For all 25 signals 500 beats were excluded. Finally, our research material consisted of 1,221,034 heartbeats of which 519,664 were related to AF episodes.

The aim of our research was to achieve the highest quality of classification (maximum FS value) of the all MIT-BIH AF database. In the subsequent experiments we considered each of the  $N_{ALL} = 1,221,034$  heartbeats as an independent event. Such a large number of feature vectors to be processed makes the application of LSVM based classifier difficult. For example, the LSVM training requires matrices of  $N_{TRN} \times N_{TRN}$  dimension, where  $N_{TRN}$  is the size of a training set. Hence, when applying only half of the heartbeats from MIT-BIH AF database ( $N_{TRN} = 610,517$ ) for training, only one of these matrices would require approximately 2982 GB of the RAM memory (when stored as double-precision floating-point values). At the same time the efficiency of the LSVM classifier depends to a large extent on the choice of the training data [69]. For this reason, one of the main objectives of our study was to find a training set of the smallest possible size that still would guarantee a satisfactory quality of AF detection. At the first stage, we investigated the ability to distinguish between the AF and non-AF episodes for all data by applying the LSVM classifier trained using heartbeats extracted (drawn randomly) from a single record only. In this way, we were able to specify which signals are most useful for the LSVM classifier training, i.e., leading to the highest classification quality of the all database, as well as to determine the size of the training set necessary to achieve the satisfactory level of the F-Score values. On this basis, we conducted learning by randomly selecting training data from the all database and from a selected group of signals (characterized by the highest values of prognostic measures).

The percentage of the heartbeats marked as AF episodes varies (Table 1). For example, signal 00735 contains only 0.83% of AF episodes ( $N_{AF} = 332$ ), while signals 07162 and 07859 consist exclusively of AF episodes ( $N_{AF} = 39,277$  and  $N_{AF} = 60,245$ , respectively). Except only one signal 06995, there are large disproportions between the numbers of beats representing the AF absence and presence (see  $N_{AF}/N_{SIG}$  in Table 1, where  $N_{SIG}$  is the total number of the heartbeats in a given signal), which may adversely affect the LSVM training [69]. To avoid the problem of poor classification efficiency, being the result of highly imbalanced data, the same number of cases from the minority and the majority class was randomly drawn from a given signal (without replacement) to maintain an equal size of both classes in the training data [70–72]. Also, as the generalization ability of a classifier is of crucial importance, only 50% of the heartbeats from the minority class of given signal were used during LSVM training. All the remaining heartbeats (AF and non-AF episodes) were used as a testing set to estimate the classification quality.

In order to explain the way of selecting the training data from a given signal, let us consider an example of a training set that was formed based on heartbeats originating from the signal 00735. As the AF is the minority class in this signal, firstly 166 (50% of the minority class) heartbeats annotated as AF were randomly selected as the training data. Secondly, the selected AF data were completed with 166 heartbeats that were randomly drawn from the non-AF heartbeats of the signal 00735. Finally, these 332 heartbeats were enclosed in the training, while the remaining 1,220,702 in the testing set. From each signal 50 different training sets were generated at random. As there are no non-AF episodes in the signals 07162 and 07859 (Table 1), we could not use the data extracted from these signals for classifier training.

**Table 1.** The size of the analyzed classes described by the number of AF ( $N_{AF}$ ), non-AF ( $N_{nAF}$ ) episodes, as well as by the percentage of AF episodes ( $N_{AF}$ ) to the total number of the heartbeats ( $N_{SIG}$ ) in a given signal.

|                  | Signal |        |        |        |        |        |        |        |        |        |        |        |        |
|------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|                  | 00735  | 03665  | 04015  | 04043  | 04048  | 04126  | 04746  | 04908  | 04936  | 05091  | 05121  | 05261  | 06426  |
| $N_{AF}$         | 332    | 11057  | 525    | 14634  | 813    | 3293   | 30873  | 5810   | 39681  | 138    | 33759  | 934    | 53115  |
| $N_{nAF}$        | 39880  | 41687  | 43459  | 47260  | 39100  | 39546  | 16979  | 55929  | 13944  | 36634  | 16101  | 44579  | 2019   |
| $N_{AF}/N_{SIG}$ | 0.83%  | 20.96% | 1.19%  | 23.64% | 2.04%  | 7.69%  | 64.52% | 9.41%  | 74.00% | 0.38%  | 67.71% | 2.05%  | 96.34% |
|                  | Signal |        |        |        |        |        |        |        |        |        |        |        |        |
|                  | 06453  | 06995  | 07162  | 07859  | 07879  | 07910  | 08215  | 08219  | 08378  | 08405  | 08434  | 08455  | All    |
| $N_{AF}$         | 445    | 27505  | 39277  | 60245  | 40035  | 6758   | 33118  | 14194  | 11478  | 45083  | 2310   | 44252  | 519664 |
| $N_{nAF}$        | 34371  | 27663  | 0      | 0      | 16538  | 29820  | 10217  | 45078  | 34016  | 13752  | 37519  | 15279  | 701370 |
| $N_{AF}/N_{SIG}$ | 1.28%  | 49.86% | 100.0% | 100.0% | 70.77% | 18.48% | 76.42% | 23.95% | 25.23% | 76.63% | 5.80%  | 74.33% | 42.56% |

### 3. Results

#### 3.1. LSVM Classifier Performance

For the purpose of the LSVM classification the class labels +1 (−1) were assigned to the feature vectors that represent the presence (absence) of the AF episode. The input data were scaled to the range  $[-1, +1]$  as recommended in [50]. To guarantee the convergence of the LSVM learning, the parameter  $\alpha$  was set to  $1.9/\gamma$ . The stop condition was an execution of the maximum number of 100 iterations or  $\|\lambda^{(k)} - \lambda^{(k-1)}\| \leq 10^{-5}$ .

The Monte Carlo Cross Validation procedure was used to assess the classification performance [70,73]. In each experiment, the MIT-BIH AF database was 50 times randomly divided into separate training and testing sets. The mean values and standard deviations of performance measures for all 50 trails are presented as the final results.

At the beginning five training sets chosen from the five signals (25 training sets in total) of the smallest size, i.e., data extracted from signals 05091, 00735, 06453, 04015, and 04048, were used to find LSVM classifier parameters  $\gamma$  and  $\chi$ . Their values were searched within the set  $\{10^{-5}, 4 \cdot 10^{-5}, 7 \cdot 10^{-5}, 10^{-4}, 4 \cdot 10^{-4}, 7 \cdot 10^{-4}, \dots, 7 \cdot 10^4, 10^5\}$ . Parameters providing the highest mean F-Score calculated for all the 25 testing sets ( $\gamma = 10$ , and  $\chi = 4$ ) were selected and used in all numerical experiments performed.

Table 2 shows the results of classification of testing data using the training sets which were extracted separately from each of the available signals. One can notice that the highest classification sensitivity of the classification  $Se = 99.12 \pm 0.29\%$  was obtained using the training data extracted from the signal 05091, and the highest specificity  $Sp = 97.42 \pm 0.17\%$  by using the training data selected from the signal 04126. However, the highest classification quality  $FS = 93.39 \pm 0.22\%$ , as well as  $CA = 94.22 \pm 0.19\%$ , were obtained with the training data extracted from the signal 08405.

The best FS does not apply to the highest relative size of the training set ( $N_{TRN}/N_{ALL} = 2.25\%$ , signal 06995), but to the signal 08405 ( $N_{TRN}/N_{ALL} = 1.13\%$ ), where only 5.72% of the heartbeats (on average) from the all database were classified incorrectly. Not the size of the training set is of crucial importance, but the occurrence of the input vectors containing the quantitative parameters of HR variability description which allow for separating the AF and non-AF heartbeats with the widest separation margin (support vectors), guarantying the best classification quality.

To allow a comparison with the results reported in the literature, in the last column of the Table 2 values of the  $FS_{ALL}$  are presented. They were calculated by using all heartbeats from the MIT-BIH AF database ( $N_{ALL} = 1,221,034$ ). It is necessary to emphasize that the positive bias of the classifier efficiency being the result of incorporating the classification results of the training data, is insignificant here as the mean difference between  $FS_{ALL}$  and FS is equal to 0.09%. It is due to the very small size of the training data up to a maximum of 2.25% of the size of the MIT-BIH AF database (signal 06995, see Table 2, column 2).

During our next experiment, we investigated the LSVM generalization ability when training with balanced (of an equal size of AF and non-AF classes) data sets extracted from the mixed signals. The maximum size of training data for that experiment was determined basing on analysis of performance measures obtained for particular signals, as listed in Table 2. We have assumed satisfactory classification quality as  $FS > 90\%$ . Since the effectiveness of the LSVM classifier increases with the number of training data, among training data of different size which ensured  $FS > 90\%$ , we selected the maximum size being equal to 1.39% of total number of heartbeats (signal 04746). It refers to 16,979 heartbeats (see Table 1). Finally, 17,000 training vectors were randomly selected from each of the following signals set:

- mixed two signals of the highest FS (08405) and Se (05091)—marked as the Training Data no. 2—TD<sub>2</sub>,
- TD<sub>2</sub> with additional signal of the highest Sp (04126)—TD<sub>3</sub>,
- all the signals—TD<sub>4</sub>.

**Table 2.** The results of MIT-BIH AF database evaluation using the LSVM classifier ( $\gamma = 10$ ,  $\chi = 4$ ) trained with the balanced data separately extracted from each of the signal. The second column provides the percentage of the size of the training data ( $N_{TRN}$ ) to the number of all heartbeats ( $N_{ALL}$ ) in the MIT-BIH AF database. Columns 3–8 shows the results for testing data only (not used during the learning phase), the last column presents the F-Score values ( $FS_{ALL}$ ) calculated using all heartbeats from the MIT-BIH AF database. The best results are in boldface.

| Signal       | $N_{TRN}/N_{ALL}$<br>[%] | Performance Measure [%] |                     |                     |                     |                     |                     |                     |
|--------------|--------------------------|-------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|              |                          | CA                      | Se                  | Sp                  | PPV                 | NPV                 | FS                  | $FS_{ALL}$          |
| 00735        | 0.03                     | 93.60 ± 0.53 *          | 97.54 ± 0.52        | 90.68 ± 1.10        | 88.60 ± 1.15        | 98.03 ± 0.39        | 92.85 ± 0.54        | 92.85 ± 0.54        |
| 03665        | 0.91                     | 81.42 ± 0.42            | 60.55 ± 1.04        | 96.84 ± 0.08        | 93.41 ± 0.13        | 76.87 ± 0.46        | 73.47 ± 0.77        | 73.81 ± 0.76        |
| 04015        | 0.04                     | 76.17 ± 1.63            | 48.14 ± 4.20        | 96.93 ± 0.46        | 92.09 ± 0.82        | 71.66 ± 1.62        | 63.12 ± 3.55        | 63.15 ± 3.55        |
| 04043        | 1.20                     | 91.47 ± 1.93            | 95.61 ± 1.84        | 88.41 ± 2.95        | 85.99 ± 2.99        | 96.48 ± 1.44        | 90.52 ± 2.01        | 90.63 ± 1.98        |
| 04048        | 0.07                     | 89.56 ± 2.69            | 87.37 ± 7.56        | 91.18 ± 1.59        | 88.06 ± 1.46        | 91.01 ± 4.75        | 87.53 ± 3.88        | 87.54 ± 3.88        |
| <b>04126</b> | 0.27                     | 77.81 ± 1.22            | 51.32 ± 3.02        | <b>97.42 ± 0.17</b> | <b>93.64 ± 0.30</b> | 73.02 ± 1.20        | 66.25 ± 2.56        | 66.39 ± 2.54        |
| 04746        | 1.39                     | 91.55 ± 0.15            | 98.80 ± 0.08        | 86.20 ± 0.27        | 84.08 ± 0.26        | 98.99 ± 0.07        | 90.85 ± 0.14        | 90.99 ± 0.14        |
| 04908        | 0.48                     | 92.70 ± 0.72            | 96.64 ± 0.39        | 89.79 ± 1.32        | 87.53 ± 1.42        | 97.31 ± 0.29        | 91.85 ± 0.74        | 91.90 ± 0.74        |
| 04936        | 1.14                     | 87.53 ± 0.51            | 91.68 ± 0.36        | 84.47 ± 0.94        | 81.34 ± 0.91        | 93.22 ± 0.26        | 86.20 ± 0.48        | 86.36 ± 0.48        |
| <b>05091</b> | 0.01                     | 90.29 ± 0.53            | <b>99.12 ± 0.29</b> | 83.75 ± 1.07        | 81.89 ± 0.94        | <b>99.23 ± 0.25</b> | 89.68 ± 0.49        | 89.68 ± 0.49        |
| 05121        | 1.32                     | 87.38 ± 0.66            | 96.21 ± 0.41        | 80.87 ± 1.16        | 78.79 ± 1.02        | 96.66 ± 0.35        | 86.63 ± 0.62        | 86.78 ± 0.61        |
| 05261        | 0.08                     | 83.01 ± 2.89            | 64.54 ± 7.01        | 96.70 ± 0.23        | 93.50 ± 0.49        | 78.78 ± 3.28        | 76.17 ± 5.10        | 76.19 ± 5.09        |
| 06426        | 0.17                     | 91.84 ± 0.60            | 89.91 ± 0.89        | 93.26 ± 1.05        | 90.83 ± 1.27        | 92.59 ± 0.58        | 90.36 ± 0.66        | 90.38 ± 0.66        |
| 06453        | 0.04                     | 92.30 ± 0.83            | 95.70 ± 2.44        | 89.78 ± 1.57        | 87.45 ± 1.55        | 96.63 ± 1.74        | 91.36 ± 1.01        | 91.36 ± 1.01        |
| 06995        | 2.25                     | 91.07 ± 0.61            | 86.59 ± 1.44        | 94.37 ± 0.30        | 91.89 ± 0.41        | 90.54 ± 0.92        | 89.15 ± 0.82        | 89.40 ± 0.80        |
| 07879        | 1.35                     | 91.16 ± 0.27            | 93.12 ± 0.51        | 89.72 ± 0.29        | 86.98 ± 0.32        | 94.65 ± 0.38        | 89.95 ± 0.31        | 90.10 ± 0.31        |
| 07910        | 0.55                     | 88.31 ± 0.87            | 89.04 ± 1.23        | 87.77 ± 1.10        | 84.35 ± 1.23        | 91.55 ± 0.89        | 86.63 ± 0.98        | 86.71 ± 0.97        |
| 08215        | 0.84                     | 89.19 ± 0.46            | 96.44 ± 1.00        | 83.84 ± 0.84        | 81.53 ± 0.72        | 96.96 ± 0.81        | 88.35 ± 0.50        | 88.46 ± 0.49        |
| 08219        | 1.16                     | 85.99 ± 1.11            | 72.70 ± 2.54        | 95.80 ± 0.55        | 92.75 ± 0.90        | 82.64 ± 1.35        | 81.49 ± 1.70        | 81.76 ± 1.67        |
| 08378        | 0.94                     | 91.24 ± 0.27            | 97.58 ± 0.29        | 86.56 ± 0.47        | 84.29 ± 0.45        | 97.98 ± 0.23        | 90.45 ± 0.28        | 90.55 ± 0.27        |
| <b>08405</b> | 1.13                     | <b>94.22 ± 0.19</b>     | 96.11 ± 0.47        | 92.82 ± 0.28        | 90.81 ± 0.30        | 97.00 ± 0.35        | <b>93.39 ± 0.22</b> | <b>93.47 ± 0.22</b> |
| 08434        | 0.19                     | 92.81 ± 0.39            | 97.37 ± 0.51        | 89.44 ± 0.64        | 87.23 ± 0.66        | 97.87 ± 0.40        | 92.02 ± 0.42        | 92.04 ± 0.42        |
| 08455        | 1.25                     | 92.99 ± 0.21            | 97.73 ± 0.19        | 89.49 ± 0.39        | 87.28 ± 0.40        | 98.16 ± 0.15        | 92.21 ± 0.22        | 92.32 ± 0.21        |

\* mean ± standard deviation.

In each of these cases the classification performance was calculated, as in previous experiments, for 50 different training/testing data divisions. The obtained classification results are presented in Table 3. As reference, the classification results using the training data extracted from the signal 08405 (TD<sub>1</sub>), providing the highest FS value during the previous experiment, are shown as well.

The balanced training vectors (TD<sub>2</sub>, TD<sub>3</sub>), being extracted from those signals that provided the best quality of the AF detection during our previous experiment, did not improve the classification efficiency. In fact, lower classification quality was obtained when comparing to the classification results based on training data extracted from the signal 08405 only (TD<sub>1</sub>). However, by applying the training data that was randomly drawn from all signals (TD<sub>4</sub>) we achieved the highest quality  $FS = 97.26 \pm 0.04\%$  ( $FS_{ALL} = 97.30 \pm 0.04\%$ ), and the highest accuracy  $CA = 97.42 \pm 0.04\%$  ( $CA_{ALL} = 97.44 \pm 0.04\%$ ). It is worth to emphasize that we were able to get these results using only 1.39% of all MIT-BIH AF

heartbeats as the training data. This confirms very high classifier generalization ability. The highest FS ( $FS_{ALL}$ ) value that has been noted among the fifty various divisions was equal to 97.34% (97.38%).

**Table 3.** The results of MIT-BIH AF database evaluation using the LSVM classifier ( $\gamma = 10$ ,  $\chi = 4$ ) trained with the balanced dataset ( $N_{TRN} = 17\,000$ ), extracted from: the signal of highest FS (TD<sub>1</sub>), mixed signals of the highest Se and FS (TD<sub>2</sub>), mixed signals that were characterized by the highest Se, Sp and FS (TD<sub>3</sub>), and all the signals (TD<sub>4</sub>). The best results obtained for testing data only and for all data are in boldface.

| Training Data Selection | Performance Measure [%] |                     |                     |                     |                     |                     |
|-------------------------|-------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                         | Se                      | Sp                  | PPV                 | NPV                 | FS                  | CA                  |
| Testing data only       |                         |                     |                     |                     |                     |                     |
| TD <sub>1</sub>         | 96.11 ± 0.47 *          | 92.82 ± 0.28        | 90.81 ± 0.30        | 97.00 ± 0.35        | 93.39 ± 0.22        | 94.22 ± 0.19        |
| TD <sub>2</sub>         | 95.63 ± 0.76            | 92.43 ± 0.29        | 90.31 ± 0.31        | 96.64 ± 0.56        | 92.89 ± 0.36        | 93.79 ± 0.30        |
| TD <sub>3</sub>         | 82.53 ± 0.94            | 95.19 ± 0.17        | 92.68 ± 0.22        | 88.08 ± 0.56        | 87.31 ± 0.53        | 89.82 ± 0.37        |
| <b>TD<sub>4</sub></b>   | <b>98.12 ± 0.08</b>     | <b>97.31 ± 0.09</b> | <b>96.42 ± 0.11</b> | <b>98.60 ± 0.06</b> | <b>97.26 ± 0.04</b> | <b>97.65 ± 0.04</b> |
| All the database        |                         |                     |                     |                     |                     |                     |
| TD <sub>1</sub>         | 96.17 ± 0.46            | 92.89 ± 0.27        | 90.93 ± 0.30        | 97.03 ± 0.34        | 93.47 ± 0.22        | 94.28 ± 0.19        |
| TD <sub>2</sub>         | 95.70 ± 0.74            | 92.52 ± 0.28        | 90.46 ± 0.31        | 96.68 ± 0.55        | 93.00 ± 0.36        | 93.87 ± 0.29        |
| TD <sub>3</sub>         | 82.81 ± 0.92            | 95.25 ± 0.16        | 92.82 ± 0.21        | 88.21 ± 0.55        | 87.53 ± 0.52        | 89.96 ± 0.37        |
| <b>TD<sub>4</sub></b>   | <b>98.14 ± 0.08</b>     | <b>97.34 ± 0.09</b> | <b>96.47 ± 0.11</b> | <b>98.61 ± 0.06</b> | <b>97.30 ± 0.04</b> | <b>97.68 ± 0.04</b> |

\* mean ± standard deviation.

One of the basic goals of our work was also to verify if the application of the fetal heart rate variability features, apart from the classical parameters of ECG signal variability, improves the automated recognition of AF episodes. Hence, we investigated how the quality of LSVM classification is affected by the exclusion of FHR variability parameters from the analyzed feature vectors. Table 4 shows the results of the AF classification after removing Yeh's, Zugaib's, Huey's and de Haan's indices, while maintaining the same divisions of the research data into training and testing sets.

**Table 4.** The results of MIT-BIH AF database evaluation when FHR variability features were excluded from recognition of the AF episodes. The LSVM classifier ( $\gamma = 10$ ,  $\chi = 4$ ) was trained with the balanced dataset ( $N_{TRN} = 17\,000$ ), extracted from: the signal of highest FS (TD<sub>1</sub>), mixed signals of the highest Se and FS (TD<sub>2</sub>), mixed signals that were characterized by the highest Se, Sp and FS (TD<sub>3</sub>), and all the signals (TD<sub>4</sub>). The best results obtained for testing data only and for all data are in boldface.

| Training Data Selection | Performance Measure [%] |                     |                     |                     |                     |                     |
|-------------------------|-------------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                         | Se                      | Sp                  | PPV                 | NPV                 | FS                  | CA                  |
| Testing data only       |                         |                     |                     |                     |                     |                     |
| TD <sub>1</sub>         | 91.04 ± 1.75 *          | 93.61 ± 0.34        | 91.34 ± 0.41        | 93.42 ± 1.20        | 91.18 ± 0.91        | 92.52 ± 0.71        |
| TD <sub>2</sub>         | 91.90 ± 1.30            | 92.64 ± 0.40        | 90.21 ± 0.48        | 93.95 ± 0.91        | 91.04 ± 0.72        | 92.32 ± 0.58        |
| TD <sub>3</sub>         | 82.08 ± 1.13            | 95.17 ± 0.20        | 92.62 ± 0.25        | 87.81 ± 0.67        | 87.03 ± 0.64        | 89.61 ± 0.45        |
| <b>TD<sub>4</sub></b>   | <b>97.97 ± 0.09</b>     | <b>97.02 ± 0.10</b> | <b>96.04 ± 0.12</b> | <b>98.48 ± 0.07</b> | <b>96.99 ± 0.05</b> | <b>97.42 ± 0.04</b> |
| All the database        |                         |                     |                     |                     |                     |                     |
| TD <sub>1</sub>         | 91.16 ± 1.73            | 93.68 ± 0.34        | 91.44 ± 0.40        | 93.48 ± 1.19        | 91.29 ± 0.90        | 92.61 ± 0.70        |
| TD <sub>2</sub>         | 92.03 ± 1.28            | 92.73 ± 0.39        | 90.37 ± 0.47        | 94.02 ± 0.90        | 91.12 ± 0.71        | 92.43 ± 0.57        |
| TD <sub>3</sub>         | 82.37 ± 1.12            | 95.23 ± 0.19        | 92.76 ± 0.25        | 87.95 ± 0.66        | 87.25 ± 0.63        | 89.76 ± 0.45        |
| <b>TD<sub>4</sub></b>   | <b>97.99 ± 0.09</b>     | <b>97.04 ± 0.10</b> | <b>96.10 ± 0.12</b> | <b>98.49 ± 0.07</b> | <b>97.02 ± 0.05</b> | <b>97.44 ± 0.04</b> |

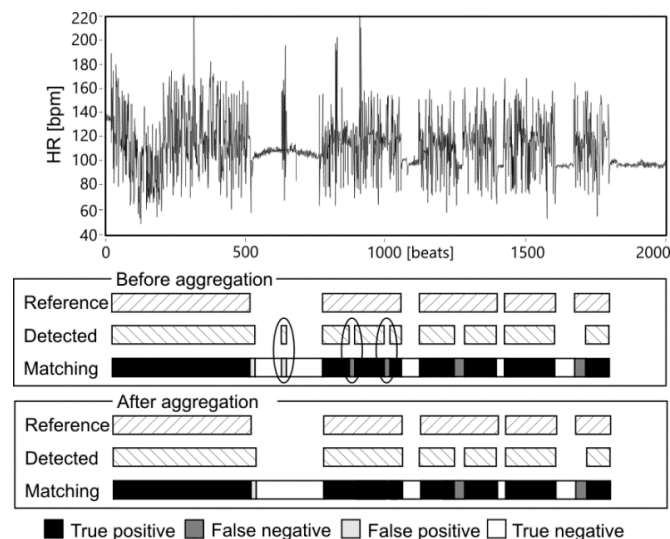
\* mean ± standard deviation.

Similarly, as in previous experiments, we assessed both the generalization ability of the LSVM classifier (estimated based on testing sets only) and the classification quality of the all MIT-BIH database. It may be observed that the absence of the FHR variability features resulted in a lower quality of classification. The mean value of the difference between the FS ( $FS_{ALL}$ ) values calculated for all considered training data

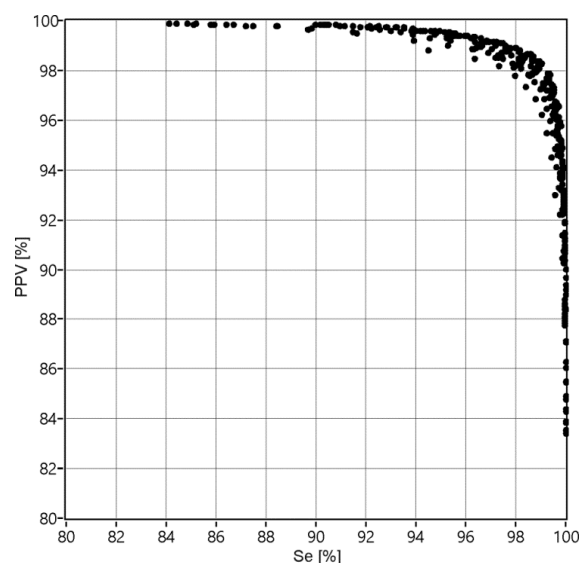
was equal to 1.15 (1.14) percentage point, with the minimum 0.27 (0.27) for  $TD_4$  and the maximum 2.21 (2.18) for  $TD_1$ . We can conclude from these results that the introduction of the FHR variability features improves the quality of automated detection of AF episodes based on the LSVM classifier.

### 3.2. Optimal Beats Aggregation

The results obtained after LSVM classification (for division of the higher  $FS_{ALL} = 97.38\%$ , that has been noted among the fifty various divisions in the training data  $TD_4$ ) were used for final AF detection—aggregation of classified heartbeats into AF episodes. Reference AF episodes and the detected ones are expected to overlap each other, but usually they overlapped partially, leading to the cases shown in Figure 4. The TP, TN, FP and FN cases were used to calculate the values of Se, PPV and FS during the process of finding the optimal parameters of the validation window width and the percentage threshold. During that process the window width was changed from 10 to 190 beats with step of 10 beats, while the threshold from 5 to 95% with 5% step. Figure 5 presents how considered values affected the Se and PPV, where each of 361 points refers to a given pair of window width and threshold values.

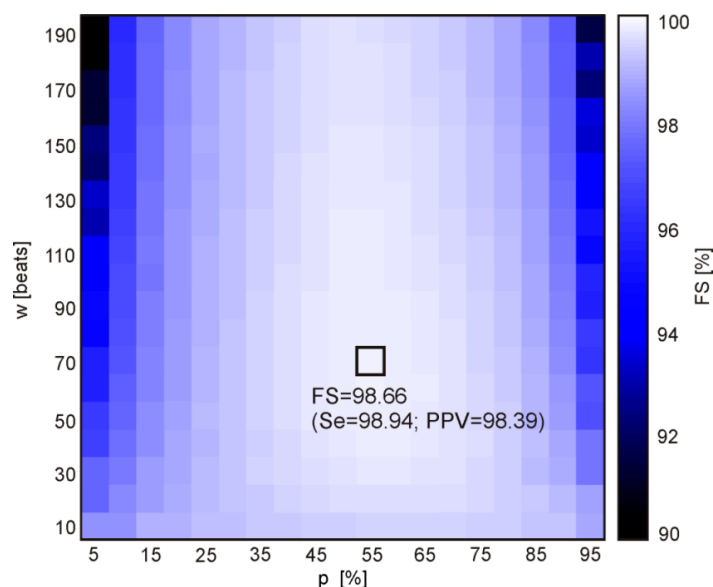


**Figure 4.** An example of matching the reference and detected AF episodes before and after aggregation. Accidental changes of AF status eliminated in the aggregation stage are marked.



**Figure 5.** The values of Se and PPV obtained for 361 combinations of the applied validation criteria: window width  $w = \{10, 20, \dots, 190\}$  and percentage threshold  $p = \{5, 10, \dots, 95\}$ .

The best performance expressed by the maximum value  $FS_{\max} = 98.66\%$ , relating to the  $Se = 98.94\%$  and  $PPV = 98.39\%$ , was obtained for the window width of 70 beats and the percentage threshold of 55%. It can be noted from Figure 6 that the performance increases with an increase of the threshold percentage from 5 to 55%, and then that control parameter does not affect the performance anymore. Considering the number of heartbeats to be aggregated, it is clear that the window should comprise at least 50 beats. Increasing the window width above this value causes slight improvement of the performance, up to optimal value of 70 heartbeats.



**Figure 6.** The distances of points defined by  $(Se(w, p), PPV(w, p))$  to the point of the highest values (100%, 100%), obtained for the 361 combinations of the following validation criteria: window width  $w = \{10, 20, \dots, 190\}$  and percentage threshold  $p = \{5, 10, \dots, 95\}$ . The brightest rectangle indicates the maximum FS (the highest algorithm performance) determined for a given combination  $(w, p)$ .

### 3.3. AF Detection Performance

The values of the performance measures:  $Se$ ,  $Sp$ ,  $PPV$ ,  $NPV$ ,  $CA$  and  $FS$  have been calculated with and without aggregation stage (Table 5). All the performance measures increased after optimized aggregation.

**Table 5.** The AF detection performance obtained without and with aggregation process, which was optimized for the window width of 110 beats and percentage threshold equal to 55%, calculated for all 1,221,033 reference annotated heartbeats.

| Performance Measures [%] | No Aggregation | With Aggregation |
|--------------------------|----------------|------------------|
| Se                       | 98.10          | 98.94            |
| Sp                       | 97.50          | 98.80            |
| PPV                      | 96.67          | 98.39            |
| NPV                      | 98.57          | 99.21            |
| CA                       | 97.75          | 98.86            |
| FS                       | 97.38          | 98.66            |

## 4. Discussion

The method for automated detection of the episodes of atrial fibrillation in long-term ECG records has been described in this paper, that represents a new approach derived from the machine learning principles—the Lagrangian Support Vector Machine (LSVM). The performance of the proposed method was evaluated on the MIT-BIH Atrial Fibrillation database, which has already been widely used enabling

to compare our results with those reported earlier. On this research material the LSVM classifier, fed with sixteen features describing the heart rate variability, ensured the sensitivity 98.10%, specificity 97.50%, positive predictive value 96.67%, classification accuracy 97.75 and F-Score 97.38%. After aggregation stage those performance measures increased to 98.94%, 98.80%, 98.39%, 98.86% and 98.66 respectively. Especially, significant increase of the PPV value was noted a lower number of false AF detections. Obtained performance is higher than that provided by previously developed classification method based on linear classifier, where  $Se = 95.42\%$  and  $PPV = 94.97\%$  [53]. Thus, the proposed more advanced method has better ability to detect the true occurrences of AF and provides lower number of false arrhythmias. It should be emphasized that both in case of simple linear classifier and advanced LSVM approach the aggregation stage significantly improves the efficiency of AF episodes detection. During this study the HR features have been determined in 21 beats wide window [54]. Nevertheless, the results obtained so far by the authors are better than those provided by other automated AF detection methods reported earlier, evaluated using the MIT-BIH AF database (Table 6).

The most obvious feature to measure the RR irregularity seems to be the difference between successive intervals RR. In [33] the standard density histogram of RR differences was prepared as a template—using the annotated AF episodes, and then the similarities between the density histograms of the test data and the standard density histogram were estimated using the standard coefficient of variation (CV test) and Kolmogorov-Smirnov (K-S) statistical test. For the optimal threshold of the test output found by ROC, the K-S test showed  $Se = 94.4\%$ ,  $Sp = 97.2\%$ , and  $PPV = 96.1\%$ , for the window length of 100 intervals. Detection of AF episodes based on density histogram of RR differences was also developed by Huang et al. [74]. The proposed more advanced analysis included two steps: AF event detection using the delta RR interval distribution difference curve and AF event classification. Using the ROC curves for determining the threshold of the K-S test, the authors have achieved the higher  $Se$  and  $Sp$  (96.1% and 98.1%, respectively) for the MIT-BIH AF database. The algorithm described in [34] has been based on the extraction of simple geometric features determined from the histogram of RR prematurity, computed as the percentage variation from the current heart rate and the differences between two successive RR intervals. The feature set included: number of nonempty bins, main distribution width, difference between mean and median and geometric test of bimodality. The score system was introduced to finally classify ten-second segment as non-AF or AF period. Using the MIT-BIH AF database, the RR prematurity algorithm provided the sensitivity of 91% and PPV of 92%, while for the RR differences  $Se = 92\%$ , and  $PPV = 78\%$ . The map that plots RR intervals versus RR differences was proposed in [35]. For reference, a window was labeled as true AF episode if 1/2 of intervals in the window were annotated as AF. Threshold value of discriminative parameter—nonempty cell—was determined by ROC, and led to sensitivity 95.8% and specificity 96.4% for the optimal window length of 128 intervals.

Another linear transformation of RR intervals to differentiate between AF episodes and normal sinus rhythm was described in [44]. The proposed algorithm starts with preprocessing (estimating the RR trend and filtering the ectopic beats), then two functions to measure the RR irregularity are calculated, and finally fusion of these signals is used for detection of AF episodes relying on the fixed threshold. Based on the distribution of the fusion signal output for AF and non-AF beats, the optimal detection threshold with identical values of sensitivity and specificity was set. Using the MIT-BIH AF the authors reported very high values of sensitivity (97.1%) and specificity (98.3%). For that approach a very short window of 8 beats was found as optimal. The authors underline that the proposed method can be matched to detect very short episodes, but it is at the expense of lower specificity.

The next approach applying only RR interval was based on the variance of normalized RR intervals over ten-second sliding window [29]. According to the authors, the normalization improves the detection performance. The authors used the morphology independent QRS detector to compute RR intervals and variance, and then they smoothed the resulting classifications, using simple majority voting scheme over 600 beat windows, for further robustness. However, the tests carried out on the MIT-BIH AF database showed that the proposed algorithm has sensitivity of 96% but specificity only of 89%, which is sufficient



for AF screening only. The more advanced normalization of RR intervals by an affine transformation was proposed in [30]. Interval irregularity was represented by the sparseness of normalized interval probability distribution which was measured by the normalized entropy calculated in the window. The authors used three lengths of the window (30, 50 and 70 beats) to show their influence on the normalization. The ROC analysis enabled them to find the threshold value for the entropy classifier output, that ensured the following values of Se, Sp, PPV and CA: 96.39%, 96.38%, 95.19%, 96.38%.

**Table 6.** An overview of published results of existing AF detection methods using the MIT-BIH Atrial Fibrillation Database.

| Method                       | Features                        | Window            | Key Techniques   | Results        |                |       |                |
|------------------------------|---------------------------------|-------------------|--|----------------|----------------|-------|----------------|
|                              |                                 |                   |  | Se             | Sp             | PPV   | CA             |
| Tateno et al. 2001 [33]      | RR difference                   | 100 beats         | Histogram, Kolmogorov-Smirnov test, ROC.                                       | 94.4           | 97.2           | 96.1  | –              |
| Huang et al. 2011 [74]       | RR difference                   | 23 beats          | Histogram, SD analysis, Kolmogorov-Smirnov test.                               | 96.1           | 98.1           | –     | –              |
| Petrucci et al. 2005 [34]    | RR difference, RR prematurity   | 60 s              | Geometric measures of histogram, Score system.                                 | 92             | –              | 92    | –              |
| Lian et al. 2011 [35]        | RR interval, RR difference      | 128 beats         | Mapping RR intervals versus RR differences, thresholds.                        | 95.8           | 96.4           | –     | –              |
| Petrenas et al. 2015 [44]    | RR interval                     | 8 beats           | Thresholds   | 97.1           | 98.3           | –     | –              |
| Logan et al. 2005 [29]       | RR interval                     | 600 beats         | Variance of normalized RR interval, simple majority voting.                    | 96             | 89             | –     | –              |
| Islam et al. 2016 [30]       | RR interval                     | 70 beats          | Normalization of RR intervals by an affine transformation.                     | 96.39          | 96.38          | 95.19 | 96.38          |
| Babaeizadeh et al. 2009 [16] | RR interval P-wave measurements | –                 | Stationary first-order Markov process, decision tree.                          | 94             | 99             | 98    | –              |
| Zhou et al. 2014 [36]        | RR interval                     | 127 beats         | Mapping the RR sequence into symbolic one, Shannon entropy, ROC.               | 96.89          | 98.25          | 97.62 | 97.67          |
| Zhou et al. 2015 [75]        | RR interval                     | 127 beats         | Online version of [36]   | 97.37          | 98.44          | 97.89 | 97.99          |
| Cui et al. 2017 [37]         | RR interval                     | 150 beats         | Mapping the RR sequence into symbolic one, dissimilarity index.                | 97.04          | 97.96          | –     | 97.78          |
| Dash et al. 2009 [38]        | RR difference                   | 128 beats         | Turning points, RMSSDD, Shannon entropy, ROC.                                  | 94.4           | 95.1           | –     | –              |
| Lake et al. 2011 [15]        | RR difference                   | 12 beats          | Coefficient of sample entropy (CoSEn), ROC.                                    | 91             | 94             | –     | –              |
| Kennedy et al. 2016 [32]     | RR difference                   | 30 beats          | Random forest, k-nearest neighbor.   | 97.6           | 98.3           | 92.1  | –              |
| Andersen et al. 2017 [39]    | RR interval<br>ECG features     | 300 beats<br>30 s | Sample entropy, Shannon entropy, CoSEn, SVM.                                   | 96.81<br>94.27 | 96.20<br>98.84 | –     | 96.45<br>96.68 |
| Kumar et al. 2018 [76]       | ECG features                    | 1000 samples      | Wavelet transform, Random forest.  | 95.8           | 97.8           | –     | 96.8           |
| Nuryani et al. 2015 [40]     | RR difference                   | –                 | SVM with radial basis function.  | 95.81          | 98.44          | –     | 97.50          |
| Colloca et al. 2013 [41]     | RR difference<br>Median HR      | 30                | Entropy, SVM with radial basis function  | 99.20          | –              | 59.33 | 86.60          |
| Faust et al. 2018 [46]       | –                               | 100               | Deep Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM).        | 98.51          | 98.32          | –     | 98.67          |
| Andersen et al. 2019 [45]    | –                               | 31                | Deep learning combining with the convolutional- and Recurrent-Neural Networks. | 98.98          | 96.95          | 95.76 | 97.80          |
| Wrobel et al. 2018 [53]      | HR irregularity features        | 21                | Linear classifier  | 95.42          | 96.12          | 94.97 | 95.62          |
| Proposed method 2019         | HR irregularity features        | 21                | LSVM   | 98.94          | 98.39          | 98.86 | 98.66          |

The sequence of RR interval is assumed to be controlled by a stationary first-order Markov process characterized by a transition probability matrix as it was proposed for the first time by Moody and Mark [4] for automated detection of the AF episodes. As Markov score reflects the relative likelihood of RR intervals sequence in AF episode versus no-AF one, it can be compared to the fixed threshold applied to classify the sequences [16]. In that work the duration statistics with combining all records into one provided the values of Se, Sp, and PPV: 94%, 98% and 97%. Furthermore, a possibility of improvement of the AF episode detection by additional information on ECG morphology was investigated. When the RR interval Markov score was completed with the two P-wave measurements: the location (P-R interval variation) and the morphology (similarity between two consecutive P-waves) the Sp and PPV increased to 99%, while the Se remained unchanged. As the author concluded reduction of the false positive cases is a result of detecting valid P-waves on the ECG recording with irregular rhythm other than AF. Nevertheless, the sensitivity which defines the ability to detect true AF occurrences was significantly lower than the value achieved by our method.

In [37] the changes in RR duration during the sequence have been represented as the binary words, where value of 1 corresponds to increase of interval duration, and 0 means no change or a decrease. Then, the testing segment is classified by comparing its information-based dissimilarity index with those obtained for the templates of AF episode and normal sinus rhythm. Parameters of the classification model: the number of bits, window length and the shift for the dissimilarity comparison boundary were optimized to provide the best performance expressed by sensitivity of 97.04%, specificity of 97.96% and classification accuracy of 97.78%. Another approach based on mapping the RR sequence into symbolic one was proposed in [36]. The detection proceeds in three stages: the initial, where a RR interval sequence is pre-processed with nonlinear and integer filters, the second, where the information of the RR interval changes is converted into symbolic sequence, and final, where the Shannon entropy is calculated to discriminate whether or not the sequence relates to AF episode. Optimal discrimination threshold of Shannon Entropy was obtained by ROC analysis. The RR sequences of 127 beats were processed. The following value of Se, Sp, PPV and CA were: 96.89% 98.25% 97.62% 97.67%, while for the online version of the algorithm: 97.37%, 98.44%, 97.89% and 97.99% [75].

The entropy concept, referring to the disorder or uncertainty of a process, was used in many methods for automated detection of AF episodes, usually being included in the feature set, but also as the only measure of RR irregularity. Three statistics describing randomness, variability and complexity of the RR interval time series were proposed in [38]. The turning points ratio, root mean square of successive RR differences and Shannon entropy were employed to characterize the atrial fibrillation. Using the thresholds and data segment of 128 beats determined by ROC the sensitivity of 94.4% and specificity of 95.1% were achieved for the signals from the MIT-BIH Atrial Fibrillation Database. The optimized sample entropy measure, called coefficient of sample entropy (CoSEn), being able to detect very short AF episodes (even 12 beats) was proposed by Lake and Moorman [15]. This feature estimated the probability that short templates will match with other segments within the analyzed RR interval time series. That process was controlled by two parameters: the template length and the tolerance matching, whose optimal values were established by ROC analysis. The authors found the cutoff CoSEn value, which differentiate between AF and normal sinus rhythms, to provide a sensitivity of 91% and a specificity of 94%. In [32] the CoSEn was combined with three other features: the coefficient of variance, root mean square of the successive differences, and median absolute deviation. The detection performance of each irregularity measure was assessed individually by ROC analysis, and CoSEn performed best. The above parameters were also used as the input features set for two classifiers: random forest (RF) and k-nearest neighbor. Both classification models significantly improved the Sp and PPV values over CoSEn, but with substantial drop in Se. The best specificity of 98.3% and PPV of 92.1% were provided by RF model, while the sensitivity achieved the best value 97.6% when using CoSEn as the only discriminative feature. Those results were reported for the combined database, with MIT-BIH AF among others. When only MIT-BIH AF database was employed,

the authors noticed significant reduction of detection performance for CoSEn and median absolute deviation, expressed by smaller area under the ROC curve.

Three entropy features: sample entropy, coefficient of sample entropy, Shannon entropy, together with two linear measures: root mean square and normalized root mean square of successive differences constituted the set of RR irregularity measures being tested in [39]. Apart from that HR approach, the authors investigated the ECG-driven approach with two features: peak-to-average power ratio and log-energy entropy, extracted from 2-level stationary wavelet transform coefficients. The support vector machine was used for classification in both approaches. Three different segment lengths were evaluated: 60, 100, 300 beats for HR and 10, 15, 30 s for ECG data. Like in [35], any segment containing at least of 50% AF beats was labeled as true AF when processing ECG. For HR approach this level was reduced to 30%. The longest windows provided the best results for both HR (Se 96.81%, Sp 96.20%, CA 96.45%) and ECG (94.27%, 98.84%, 96.98%, respectively) approaches.

The performance of AF detection using the features extracted exclusively from ECG signals was assessed by Kumar [76]. As the data was taken from MIT-BIH database, the obtained results may be related to those provided by the HR-based methods described here. The proposed classification method employed two features: the log-energy entropy and permutation entropy computed from the sub-band signals obtained using flexible analytic wavelet transform. Using random forest classifier, the authors reported sensitivity of 95.8%, specificity of 97.8% and accuracy of 96.8%.

Two features: the average of RR differences in a defined duration, and the standard deviation of differences in a defined duration, were examined as the inputs of the classifier based on SVM with radial basis function in [40]. The proposed method showed following performance on the MIT-BIH AF database: Se = 95.81%, Sp = 98.44% and CA = 97.50%. The same SVM classifier was employed in [41]. The input set comprised more RR interval features: median heart rate, minimum RR interval, mean RR interval, various entropy measures, and difference irregularity measure. The MIT-BIH AF database was used in that study, but only during the training stage, when very good results have been achieved (sensitivity = 99.07%, PPV = 98.27%, accuracy = 98.84). When testing on a series of 200 signals from the MIT-BIH Arrhythmia database, the best accuracy was 86.60% for the window of 30 beats, sensitivity reached 99.20%, but PPV was only 59.33%.

The newest approach to automated AF detection proposed in [46] and [45] has been based on deep learning algorithm, which aims to develop the classification model by using all available information from the input. In case of AF detection from the ECG signals it means no need for extraction of the feature neither from raw ECG nor from RR interval time series. In those works the RR data from MIT-BIH AF were partitioned using sliding window of 100 beats [46] or 31 beats but shifted with 10 beats [45], and then fed to Recurrent Neural Network with Long Short-Term Memory. In both works very good results were reported: Se = 98.51%, Sp = 98.32%, CA = 98.67% in [46], and 98.98%, 96.95%, 97.80% with PPV of 95.76% in [45], when median filtering was used as post processing to improve the detection performance. It should be noted, however that development of the deep learning algorithm has been enabled by recent advances in parallel computing on Graphics Processing Units. The computational complexity of deep learning model is much higher than traditional feature-based classifier. This limits its application in wearable devices for long term monitoring with online AF detection, like wristband monitor in a form of wrist bracelet.

## 5. Conclusions

Despite serious medical consequences, atrial fibrillation is still an underestimated clinical and diagnostic problem. Recognition of this form of arrhythmia requires a long-term monitoring of the heart rhythm, since very often patients are asymptomatic. Moreover, the AF episodes can occur accidentally and may last from minutes to hours. The objectivity and efficiency of the visual analysis of long-term recordings can be improved by automated AF detection.

The paper proposed a LSVM-based approach with an original training stage which outperforms other automated AF detection methods based on the information on beat-to-beat irregularity proposed

in the literature. Our method ensures a very high efficiency in detection of true AF episodes expressed by sensitivity of 98.94%, and at the same time low number of false episodes, as the positive predictive value reached 98.36%. These results were achieved with post-processing aggregation stage, showing a need for final verification of the classified beats. It also turned out that extending the input feature vector to include parameters describing the heart irregularity and being typically used in the fetal heart rate analysis, had positive effect on classification efficiency. Designing the LSVM-based classifier to deal with such large amount of data like from MIT-BIH AF Database led us to valuable conclusion. Not the size of the training set is of crucial importance, but the occurrence of the input vectors containing the quantitative parameters of HR variability description which allow separating the AF and non-AF heartbeats with the widest separation margin (support vectors), thus guaranteeing the best classification quality.

**Author Contributions:** Conceptualization, R.C., K.H. and J.W.; Data curation, A.M., R.M and R.K.; Methodology, R.M., T.K., R.K. and J.J.; Software, R.C., J.W. and A.M.; Supervision, K.H. and R.C.; Validation, J.W., T.K., M.J. and J.M.L.; Writing-original draft, K.H. and R.C.; Writing-review & editing, K.H., M.J., J.J. and J.M.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This scientific research work was supported by the National Science Centre and the National Centre for Research and Development in Poland under the grants 2017/27/B/ST6/01989 and STRATEGMED2/269343/18/NCBR/2016, as well as from the Ministry of Science and Higher Education funding for statutory activities (BK-2019, BK-2020).

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

## References

- Lau, J.K.; Lowres, N.; Neubeck, L.; Brieger, D.B.; Sy, R.W.; Galloway, C.D.; Albert, D.E.; Freedman, S.B. iPhone ECG application for community screening to detect silent atrial fibrillation: A novel technology to prevent stroke. *Int. J. Cardiol.* **2013**, *165*, 193–194. [[CrossRef](#)]
- Grond, M.; Jauss, M.; Hamann, G.; Stark, E.; Veltkamp, R.; Nabavi, D.; Horn, M.; Weimar, C.; Köhrmann, M.; Wachter, R.; et al. Improved Detection of Silent Atrial Fibrillation Using 72-Hour Holter ECG in Patients With Ischemic Stroke: A Prospective Multicenter Cohort Study. *Stroke* **2013**, *44*, 3357–3364. [[CrossRef](#)] [[PubMed](#)]
- Camm, A.J. Atrial Fibrillation and Risk. *Clin. Cardiol.* **2012**, *35*, S1–S2. [[CrossRef](#)] [[PubMed](#)]
- Moody, G. A new method for detecting atrial fibrillation using RR intervals. *Comput. Cardiol.* **1983**, 227–230.
- 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](#)]
- The MIT-BIH Atrial Fibrillation Database–PhysioNet. Available online: <https://physionet.org/physiobank/database/afdb/> (accessed on 20 May 2018).
- Fitzmaurice, D.A.; Hobbs, F.R.; Jowett, S.; Mant, J.; Murray, E.T.; Holder, R.; Raftery, J.P.; Bryan, S.; Davies, M.; Lip, G.Y. Screening versus routine practice in detection of atrial fibrillation in patients aged 65 or over: Cluster randomised controlled trial. *BMJ* **2007**, *335*, 383. [[CrossRef](#)]
- Rawenwaaij-Arts, C.; Kallee, L.; Hopman, J. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability. Standards of measurement, physiologic interpretation, and clinical use. *Circulation* **1996**, *93*: 1043–1065. *Intern. Med.* **1993**, *118*, 436–447.
- Desteghe, L.; Raymaekers, Z.; Lutin, M.; Vijgen, J.; Dilling-Boer, D.; Koopman, P.; Schurmans, J.; Vanduyndhoven, P.; Dendale, P.; Heidbuchel, H. Performance of handheld electrocardiogram devices to detect atrial fibrillation in a cardiology and geriatric ward setting. *Ep Europace* **2016**, *19*, 29–39.
- Haberman, Z.C.; Jahn, R.T.; Bose, R.; Tun, H.; Shinbane, J.S.; Doshi, R.N.; Chang, P.M.; Saxon, L.A. Wireless Smartphone ECG Enables Large-Scale Screening in Diverse Populations. *J. Cardiovasc. Electrophysiol.* **2015**, *26*, 520–526. [[CrossRef](#)]
- Lee, J.; Reyes, B.A.; McManus, D.D.; Maitas, O.; Chon, K.H. Atrial fibrillation detection using an iPhone 4S. *IEEE Trans. Biomed. Eng.* **2012**, *60*, 203–206. [[CrossRef](#)]

12. Vaid, J.; Poh, M.Z.; Saleh, A.; Kalantarian, S.; Poh, Y.K.C.; Rafael, A.; Ruskin, J. Diagnostic accuracy of a novel mobile application (Cardiio Rhythm) for detecting atrial fibrillation. *J. American Coll. Cardiol.* **2015**, *65*, A361. [[CrossRef](#)]
13. Glotzer, T.V.; Hellkamp, A.S.; Zimmerman, J.; Sweeney, M.O.; Yee, R.; Marinchak, R.; Cook, J.; Paraschos, A.; Love, J.; Radoslavich, G.; et al. Atrial high rate episodes detected by pacemaker diagnostics predict death and stroke: report of the Atrial Diagnostics Ancillary Study of the MObde Selection Trial (MOST). *Circulation* **2003**, *107*, 1614–1619. [[CrossRef](#)] [[PubMed](#)]
14. Hindricks, G.; Pokushalov, E.; Urban, L.; Táborský, M.; Kuck, K.-H.; Lebedev, D.; Rieger, G.; Pürerfellner, H.; on behalf of the XPECT Trial Investigators. Performance of a New Leadless Implantable Cardiac Monitor in Detecting and Quantifying Atrial Fibrillation Results of the XPECT Trial. *Circ. Arrhythmia Electrophysiol.* **2010**, *3*, 141–147. [[CrossRef](#)]
15. Lake, D.E.; Moorman, J.R. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am. J. Physiol. Circ. Physiol.* **2011**, *300*, H319–H325. [[CrossRef](#)] [[PubMed](#)]
16. Babaeizadeh, S.; Gregg, R.E.; Helfenbein, E.D.; Lindauer, J.M.; Zhou, S.H. Improvements in atrial fibrillation detection for real-time monitoring. *J. Electrocardiol.* **2009**, *42*, 522–526. [[CrossRef](#)] [[PubMed](#)]
17. Christov, G.B.I. Automatic detection of atrial fibrillation and flutter by wave rectification method. *J. Med. Eng. Technol.* **2001**, *25*, 217–221. [[CrossRef](#)]
18. Christov, I.; Bortolan, G.; Daskalov, I. Sequential analysis for automatic detection of atrial fibrillation and flutter. *Comput. Cardiol.* **2001**, *28*, 293–296.
19. Hindricks, G.; Piorkowski, C. Atrial fibrillation monitoring: mathematics meets real life. *Circulatory* **2012**, *126*, 791–802. [[CrossRef](#)]
20. Hargittai, S. Is it possible to detect atrial fibrillation by simply using RR intervals? *Comput. Cardiol.* **2014**, *41*, 897–900.
21. Slocum, J.; Sahakian, A.; Swiryn, S. Diagnosis of atrial fibrillation from surface electrocardiograms based on computer-detected atrial activity. *J. Electrocardiol.* **1992**, *25*, 1–8. [[CrossRef](#)]
22. Bonomi, A.; Schipper, F.; Eerikainen, L.; Margarito, J.; Aarts, R.; Babaeizadeh, S.; De Morree, H.; Dekker, L. Atrial Fibrillation Detection Using Photo-plethysmography and Acceleration Data at the Wrist. In Proceedings of the 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada, 11–14 September 2016; pp. 081–339.
23. Lu, S.; Zhao, H.; Ju, K.; Shin, K.; Lee, M.; Shelley, K.; Chon, K.H. Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information? *J. Clin. Monit. Comput.* **2008**, *22*, 23–29. [[CrossRef](#)] [[PubMed](#)]
24. Tamura, T.; Maeda, Y.; Sekine, M.; Yoshida, M. Wearable Photoplethysmographic Sensors—Past and Present. *Electronics* **2014**, *3*, 282–302. [[CrossRef](#)]
25. Hurnanen, T.; Lehtonen, E.; Tadi, M.J.; Kuusela, T.; Kiviniemi, T.; Saraste, A.; Vasankari, T.; Airaksinen, J.; Koivisto, T.; Pankaala, M. Automated Detection of Atrial Fibrillation Based on Time–Frequency Analysis of Seismocardiograms. *IEEE J. Biomed. Heal. Inform.* **2017**, *21*, 1233–1241. [[CrossRef](#)] [[PubMed](#)]
26. Wrobel, J.; Jezewski, J.; Horoba, K.; Pawlak, A.; Czabanski, R.; Jezewski, M.; Porwik, P. Medical Cyber-Physical System for Home Telecare of High-Risk Pregnancy: Design Challenges and Requirements. *J. Med. Imaging Heal. Inform.* **2015**, *5*, 1295–1301. [[CrossRef](#)]
27. Jezewski, J.; Pawlak, A.; Horoba, K.; Wrobel, J.; Czabanski, R.; Jezewski, M. Selected design issues of the medical cyber-physical system for telemonitoring pregnancy at home. *Microprocess. Microsyst.* **2016**, *46*, 35–43. [[CrossRef](#)]
28. Roj, D.; Matonia, A.; Sobotnicka, E.; Wrobel, J. Hardware design issues and functional requirements for smart wristband monitor of silent atrial fibrillation. In Proceedings of the 2017 MIXDES—24th International Conference “Mixed Design of Integrated Circuits and Systems”, Bydgoszcz, Poland, 22–24 June 2017; pp. 596–600.
29. Logan, B.; Healey, J. Robust detection of atrial fibrillation for a long term telemonitoring system. *Comput. Cardiol.* **2005**, *32*, 619–622.
30. Islam, S.; Ammour, N.; Alajlan, N.; Aboalsamh, H. Rhythm-based heartbeat duration normalization for atrial fibrillation detection. *Comput. Boil. Med.* **2016**, *72*, 160–169. [[CrossRef](#)]

31. Ghodrati, A.; Murray, B.; Marinello, S. RR interval analysis for detection of Atrial Fibrillation in ECG monitors. In Proceedings of the 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Vancouver, BC, Canada, 20–24 August 2008; pp. 601–604.
32. Kennedy, A.; Finlay, D.D.; Guldenring, D.; Bond, R.R.; Moran, K.; McLaughlin, J. Automated detection of atrial fibrillation using R-R intervals and multivariate-based classification. *J. Electrocardiol.* **2016**, *49*, 871–876. [[CrossRef](#)]
33. Tateno, K.; Glass, L. Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and  $\Delta$ RR intervals. *Med. Boil. Eng.* **2001**, *39*, 664–671. [[CrossRef](#)]
34. Petrucci, E.; Balian, V.; Filippini, G.; Mainardi, L. Atrial fibrillation detection algorithms for very long term ECG monitoring. *Comput. Cardiol.* **2005**, *32*, 623–626.
35. Lian, J.; Wang, L.; Muessig, D. A Simple Method to Detect Atrial Fibrillation Using RR Intervals. *Am. J. Cardiol.* **2011**, *107*, 1494–1497. [[CrossRef](#)] [[PubMed](#)]
36. Zhou, X.; Ding, H.; Ung, B.; Pickwell-MacPherson, E.; Zhang, Y. Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy. *Biomed. Eng. Online* **2014**, *13*, 18. [[CrossRef](#)] [[PubMed](#)]
37. Cui, X.; Chang, E.; Yang, W.-H.; Jiang, B.C.; Yang, A.C.; Peng, C.-K. Automated Detection of Paroxysmal Atrial Fibrillation Using an Information-Based Similarity Approach. *Entropy* **2017**, *19*, 677. [[CrossRef](#)]
38. Dash, S.; Chon, K.H.; Lu, S.; Raeder, E.A. Automatic Real Time Detection of Atrial Fibrillation. *Ann. Biomed. Eng.* **2009**, *37*, 1701–1709. [[CrossRef](#)] [[PubMed](#)]
39. Andersen, R.S.; Poulsen, E.S.; Puthusserypady, S. A novel approach for automatic detection of Atrial Fibrillation based on Inter Beat Intervals and Support Vector Machine. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 2039–2042.
40. Nuryani, N.; Harjito, B.; Yahya, I.; Lestari, A. Atrial fibrillation detection using support vector machine. In Proceedings of the Joint International Conference on Electric Vehicular Technology and Industrial, Mechanical, Electrical and Chemical Engineering (ICEVT & IMECE), Surakarta, Indonesia, 4–5 November 2015; pp. 215–218.
41. Colloca, R.; Johnson, A.E.; Mainardi, L.; Clifford, G.D. A support vector machine approach for reliable detection of atrial fibrillation events. *Comput. Cardiol.* **2013**, 1047–1050.
42. Yang, T.-F.; Devine, B.; Macfarlane, P.W. Artificial neural networks for the diagnosis of atrial fibrillation. *Med. Boil. Eng.* **1994**, *32*, 615–619. [[CrossRef](#)]
43. Artis, S.G.; Mark, R.G.; Moody, G.B. Detection of atrial fibrillation using artificial neural networks. In Proceedings of the Computers in Cardiology, Venice, Italy, 23–26 September 1991; pp. 173–176.
44. Petrénas, A.; Marozas, V.; Sörnmo, L. Low-complexity detection of atrial fibrillation in continuous long-term monitoring. *Comput. Boil. Med.* **2015**, *65*, 184–191. [[CrossRef](#)]
45. Andersen, R.S.; Peimankar, A.; Puthusserypady, S. A deep learning approach for real-time detection of atrial fibrillation. *Expert Syst. Appl.* **2019**, *115*, 465–473. [[CrossRef](#)]
46. Faust, O.; Shenfield, A.; Kareem, M.; San, T.R.; Fujita, H.; Acharya, U.R. Automated detection of atrial fibrillation using long short-term memory network with RR interval signals. *Comput. Boil. Med.* **2018**, *102*, 327–335. [[CrossRef](#)]
47. Wrobel, J.; Matonia, A.; Horoba, K.; Jezewski, J.; Czabanski, R.; Pawlak, A.; Porwik, P. Pregnancy Telemonitoring with Smart Control of Algorithms for Signal Analysis. *J. Med. Imaging Heal. Inform.* **2015**, *5*, 1302–1310. [[CrossRef](#)]
48. Jezewski, J.; Horoba, K.; Roj, D.; Wrobel, J.; Kupka, T.; Matonia, A. Evaluating the fetal heart rate baseline estimation algorithms by their influence on detection of clinically important patterns. *Biocybern. Biomed. Eng.* **2016**, *36*, 562–573. [[CrossRef](#)]
49. Wrobel, J.; Roj, D.; Jezewski, J.; Horoba, K.; Kupka, T.; Jezewski, M. Evaluation of the Robustness of Fetal Heart Rate Variability Measures to Low Signal Quality. *J. Med. Imaging Heal. Inform.* **2015**, *5*, 1311–1318. [[CrossRef](#)]
50. Mangasarian, O.L.; Musicant, D.R. Lagrangian support vector machines. *J. Mach. Learn. Res.* **2001**, *1*, 161–177.
51. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
52. Roj, D.; Wrobel, J.; Matonia, A.; Horoba, K.; Henzel, N. Control and signal processing software embedded in smart wristband monitor of silent atrial fibrillation. In Proceedings of the 2017 MIXDES—24th International Conference “Mixed Design of Integrated Circuits and Systems”, Bydgoszcz, Poland, 22–24 June 2017; pp. 585–590.

53. Wrobel, J.; Horoba, K.; Matonia, A.; Kupka, T.; Henzel, N.; Sobotnicka, E. Optimizing the Automated Detection of Atrial Fibrillation Episodes in Long-term Recording Instrumentation. In Proceedings of the 2018 25th International Conference “Mixed Design of Integrated Circuits and System” (MIXDES), Gdynia, Poland, 21–23 June 2018; pp. 460–464.
54. Henzel, N.; Wrobel, J.; Horoba, K. Atrial fibrillation episodes detection based on classification of heart rate derived features. In Proceedings of the 2017 MIXDES—24th International Conference “Mixed Design of Integrated Circuits and Systems”, Bydgoszcz, Poland, 22–24 June 2017; pp. 571–576.
55. Wróbel, J.; Horoba, K.; Pander, T.; Jeżewski, J.; Czabański, R. Improving fetal heart rate signal interpretation by application of myriad filtering. *Biocybern. Biomed. Eng.* **2013**, *33*, 211–221. [[CrossRef](#)]
56. Jeżewski, M.; Wrobel, J.; Horoba, K.; Gacek, A.; Henzel, N.; Leski, J. The Prediction of Fetal Outcome by Applying Neural Network for Evaluation of CTG Records. In *Computer Recognition Systems 2. Advances in Soft Computing*; Kurzynski, M., Puchala, E., Wozniak, M., Zolnierrek, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 45, pp. 532–541.
57. Jeżewski, J.; Wrobel, J.; Matonia, A.; Horoba, K.; Martinek, R.; Kupka, T.; Jeżewski, M. Is Abdominal Fetal Electrocardiography an Alternative to Doppler Ultrasound for FHR Variability Evaluation? *Front. Physiol.* **2017**, *8*, 305. [[CrossRef](#)] [[PubMed](#)]
58. Czabanski, R.; Jeżewski, M.; Wrobel, J.; Horoba, K.; Jeżewski, J. A Neuro-Fuzzy Approach to the Classification of Fetal Cardiotocograms. In IFMBE Proceedings of the 14th Nordic Baltic Conference on Biomedical Engineering and Medical Physics, Riga, Latvia, 16–20 June 2008; Volume 20, pp. 446–449, ISBN 978-3-540-69366-6.
59. Czabański, R.; Jeżewski, J.; Horoba, K.; Jeżewski, M. Fetal state assessment using fuzzy analysis of fetal heart rate signals—Agreement with the neonatal outcome. *Biocybern. Biomed. Eng.* **2013**, *33*, 145–155. [[CrossRef](#)]
60. Matonia, A.; Jeżewski, J.; Kupka, T.; Horoba, K.; Wrobel, J.; Gacek, A. The influence of coincidence of fetal and maternal QRS complexes on fetal heart rate reliability. *Med. Boil. Eng.* **2006**, *44*, 393–403. [[CrossRef](#)]
61. Jeżewski, J.; Matonia, A.; Kupka, T.; Roj, D.; Czabanski, R. Determination of fetal heart rate from abdominal signals: Evaluation of beat-to-beat accuracy in relation to the direct fetal electrocardiogram. *Biomed. Tech./Biomed. Eng.* **2012**, *57*, 383–394. [[CrossRef](#)]
62. Jeżewski, M.; Czabanski, R.; Horoba, K.; Leski, J. Clustering with Pairs of Prototypes to Support Automated Assessment of the Fetal State. *Appl. Artif. Intell.* **2016**, *30*, 572–589. [[CrossRef](#)]
63. Vapnik, V. *Statistical Learning Theory*; John Wiley & Sons: New York, NY, USA, 1998; ISBN 978-0471030034.
64. Abe, S. *Support Vector Machines for Pattern Classification*; Springer Science and Business Media LLC: London, UK, 2010; ISBN-13: 9781849960977.
65. Steinwart, I.; Christmann, A. *Support Vector Machines*; Springer: New York, NY, USA, 2008; ISBN 978-0-387-77242-4.
66. Joachims, T. *Learning to Classify Text Using Support Vector Machines—Methods, Theory, and Algorithms*; Kluwer Academic Publishers: Norvell, CA, USA, 2002; ISBN 079237679X.
67. Suykens, J.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
68. Tsang, I.W.; Kwok, J.T.; Cheung, P.M. Core vector machines: Fast SVM training on very large data sets. *J. Mach. Learn. Res.* **2005**, *6*, 363–392.
69. Czabanski, R.; Jeżewski, M.; Horoba, K.; Jeżewski, J.; Leski, J. Fuzzy Analysis of Delivery Outcome Attributes for Improving the Automated Fetal State Assessment. *Appl. Artif. Intell.* **2016**, *30*, 556–571. [[CrossRef](#)]
70. Jeżewski, M.; Leski, J.; Czabanski, R. An Attempt to Optimize the Cardiotocographic Signal Feature Set for Fetal State Assessment. *J. Med. Imaging Heal. Inform.* **2015**, *5*, 1364–1373. [[CrossRef](#)]
71. Jeżewski, M.; Leski, J.M. Nonlinear Extension of the IRLS Classifier Using Clustering with Pairs of Prototypes. In *Advances in Intelligent Systems and Computing*; Burduk, R., Jackowski, K., Kurzynski, M., Wozniak, M., Zolnierrek, A., Eds.; Springer: Cham, Switzerland; Heidelberg, Germany; New York, NY, USA, 2013; Volume 226, pp. 121–130. ISBN 978-3-319-00968-1.
72. Dubitzky, W.; Granzow, M.; Berrar, D. *Fundamentals of Data Mining in Genomics and Proteomics*; Springer Science & Business Media: New York, NY, USA, 2007. [[CrossRef](#)]
73. Picard, R.R.; Cook, R.D. Cross-validation of regression models. *J. Am. Stat. Assoc.* **1984**, *79*, 575–583. [[CrossRef](#)]
74. Huang, C.; Ye, S.; Chen, H.; Li, D.; He, F.; Tu, Y. A novel method for detection of the transition between atrial fibrillation and sinus rhythm. *IEEE Trans. Biomed. Eng.* **2010**, *58*, 1113–1119. [[CrossRef](#)]

75. Zhou, X.; Ding, H.; Wu, W.; Zhang, Y. A Real-Time Atrial Fibrillation Detection Algorithm Based on the Instantaneous State of Heart Rate. *PLoS ONE* **2015**, *10*, e0136544. [[CrossRef](#)]
76. Kumar, M.; Pachori, R.B.; Acharya, U.R. Automated diagnosis of atrial fibrillation ECG signals using entropy features extracted from flexible analytic wavelet transform. *Biocybern. Biomed. Eng.* **2018**, *38*, 564–573. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).