Automated Detection of Atrial Fibrillation using R-R intervals and multivariate based classification.

Alan Kennedy*[a], Dewar D. Finlay[a], Daniel Guldenring[a], Raymond R. Bond[b], James McLaughlin[a], Kieran Moran[c].

[a]Nanotechnology and Integrated BioEngineering Centre, University of Ulster, Northern Ireland, UK.

[b]Computer Science Research Institute, University of Ulster, Northern Ireland, UK.

[c]Dublin City University, Ireland.

Corresponding Author:
Alan Kennedy
NIBEC Building
University of Ulster, Jordanstown
Newtownabbey
Co. Antrim
BT37 0QB

Abstract

Automated detection of AF from the electrocardiogram (ECG) still remains a challenge. In this study we investigated two multivariate based classification techniques, Random Forests (RF) and k-nearest neighbor ($k-nn$), for improved automated detection of AF from the ECG. We have compiled a new database from ECG data taken from existing sources. R-R intervals were then analyzed using four previously described R-R irregularity measurements: (1) The coefficient of sample entropy ($CoSEn$) (2) The coefficient of variance (CV) (3) Root mean square of the successive differences ($RMSSD$) and (4) median absolute deviation (MAD). Using outputs from all four R-R irregularity measurements RF and $k-nn$ models were trained. RF classification improved AF detection over $CoSEn$ with overall specificity of 80.1% vs. 98.3% and positive predictive value of 51.8% vs. 92.1% with a reduction in sensitivity, 97.6% vs. 92.8%. $k-nn$ also improved specificity and PPV over $CoSEn$ however the sensitivity of this approach was considerably reduced (68.0%).

1 Introduction

The automated detection and management of patients with atrial fibrillation (AF) is becoming a priority in healthcare systems globally [1], [2]. Patients suffering from AF have a 5-fold increase in the likelihood of suffering a stroke event [3], making early detection and intervention a priority. Typically, when patients are suspected of having AF, a 12-lead electrocardiogram (ECG) is recorded and this remains the gold standard for AF detection [4]. However, some AF patients suffer short and intermittent episodes which are difficult to confirm using the 12-lead ECG, given the relatively short duration of the recording (10 seconds). Patients who are suspected of having AF, which is not confirmed on the 12-lead ECG, are usually referred for extended (24-72 hour) ECG monitoring in an effort to detect the suspected AF events [4]. This extended continuous monitoring is usually performed using a Holter monitoring system although recently a number of new devices have emerged that specifically focus on the detection of AF. These include disposable patch systems [5] and iPhone based ECGs [6]. Due to the extended nature of continuous ECG recordings and the amount of associated data generated it becomes more challenging to perform human interpretation. Ideally, AF would be automatically detected via a software based algorithm, alleviating the need for time consuming manual review of ECGs by clinicians. Accurate automated detection of AF from the ECG still remains a challenge due to sustained ectopic beats (Figure 1), motion artefact and with some devices, T-wave over sensing [7]. AF is confirmed on the ECG as the absence of the P-wave, an irregular R-R interval and in some patients a fibrillatory wave on the baseline ECG [8]. Most ambulatory systems rely on the analysis of the R-R interval alone for the detection of AF. This is due to the fact that the P-wave, and, in particular, the fibrillatory wave have a low signal-to-noise ratio during ambulation. In this study we aim to determine if the automated detection of AF from R-R intervals can be improved with Random Forests (RF) and k-nearest neighbour ($k-nn$) classification models.

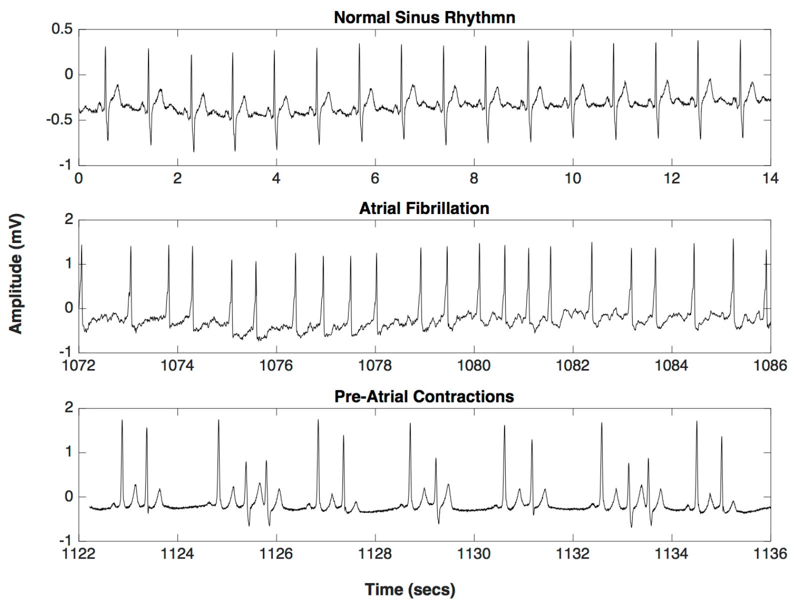

Figure 1. An example of normal sinus rhythm (top row), atrial fibrillation (middle row) and sustained pre-atrial contractions (bottom row).

2 Method

2.1 Study Data

To facilitate our experiments, we developed a new database from a range of different sources. The database structure is outlined in the Table 1. Automated QRS detection was performed on lead II using the open source GQRS algorithm [9], this method of QRS detection was chosen over the manual human annotations as it allows for the automated detection of QRS complex's, which as described previously is highly desirable. The GQRS detection algorithm was chosen based on its improved performance over other available QRS detection algorithms [10]. As can be seen in Table 1 the database ($AFDB_{comb}$) contained 322 records consisting of a total of 4232410 automatically detected ECG beats of which 17% were labeled as AF based on the reference rhythm annotations provided with the existing databases. The database also contained a significant number of pre-atrial contractions (4490) and pre-ventricular contractions (101213), the main sources of false positives for automated AF algorithms [7]. The 322 records were then randomly split into a training dataset of 249 patients (75%) and a testing dataset of the remaining 73 patients (25%).

Database	# of Records	# NSR beats	# AF beats	# PVC	# PAC
THEW	73	0	201085	0	0
MIT-BIH Arrhythmia	48	98366	11442	7136	2546
MIT-BIH AF	23	635803	486650	/	/
MIT-BIH LT	7	693120	0	64095	0
MIT-BIH NSR	18	1742213	0	26	0
MIT-BIH SVA	78	185096	0	9943	0
SPIT	75	173495	5140	20013	1944
Totals	322	3528093	704317	101213	4490
AF Prevalence		83%	17%		

Table 1. The number of AF and non AF beats detected using the GQRS algorithm. Also the number of PVC and PAC beats within each database taken from provided human beat annotations.

2.2 R-R interval pre-processing

A simple three-point median filter was applied to the R-R interval data prior to analysis. This filter is commonly implemented in an effort to remove sporadic ectopic beats from the R-R series before AF detection is attempted [11]. The filter is defined as below:

$$RR_{mf} = median\{RR(n-2), RR(n-1), RR(n)\} \tag{1}$$

Where RR_{mf} is the output of the median filter and RR is the time series of R-R intervals.

2.3 R-R irregularity measurements

For this study we implemented four R-R irregularity measurements (1) the coefficient of variance (2) the root mean square of the successive differences (3) the coefficient of sample entropy and (4) the median absolute deviation. All measurements were implemented with moving segment window of 30 R-R intervals.

2.3.1 The coefficient of variance

The coefficient of variance (CV) was calculated as follows:

$$CV = \frac{RR_{\sigma}}{RR_{\mu}} \tag{2}$$

Where RR_{σ} is the standard deviation of the R-R interval and RR_{μ} is the mean R-R interval. Episodes of AF are expected to have a greater value of the CV than those of NSR [12].

2.3.2 The root mean square of the successive differences

The root mean square of the successive difference ($RMSSD$) is calculated as follows:

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}((RR)_{i+1} - (RR)_i)^2} \tag{3}$$

Where i is the R-R interval and N is the length of the segment window. Since AF exhibits higher variability in the R-R interval than NSR, the $RMSSD$ is expected to be higher during AF than NSR [13] .

2.3.3 The coefficient of sample entropy

The coefficient of sample entropy ($CoSEn$) was first described by Lake and Moorman and is based on the concept of sample entropy [14] which is calculated as follows:

$$SampEn = -\ln(cp) = -\ln\left(\frac{A}{B}\right) = -\ln(A) + \ln(B) \tag{4}$$

Where A is the number of matches at m and B is the number of matches at $m+1$, within a tolerance of r. The coefficient of sample entropy [15] is the natural negative logarithm of the conditional probability that R-R intervals that match at length m will matches at length $m+1$.

$$CoSEn = SampEn + \ln(2r) - \ln(mean(RR)) \tag{5}$$

The $CoSEn$ is expected to be greater during AF than normal sinus rhythm (NSR), again due to the irregularity of the R-R interval during AF episodes.

2.3.4 The median absolute deviation

The median absolute deviation (MAD) was popularized by Hampel who attributed its discovery to Gauss [16]. MAD is a robust measurement of the variability in a numerical series of data. MAD is calculated as follow:

$$MAD = median(\left|RR_i - median(RR_{seg})\right|) \tag{6}$$

Where RR_i is the R-R interval and RR_{seg} is all the R-R intervals within the segment window. The R-R intervals during an episode of AF have a larger variability therefore AF is expected to have a greater value of MAD than NSR [17].

2.4 Multivariate based classification methods

For this study we focused on two main classification methods, RF and $k-nn$. Both supervised machine learning models were trained using the four R-R irregularity measurements as input features and the human rhythm annotations used as an output reference. These classification approaches were chosen in an effort to incorporate the best aspects of the existing algorithms in a way that allows for better discrimination between NSR and AF.

2.4.1 Random Forests

RF is an ensemble machine learning method for classification, in this case classifying rhythm as either NSR or AF. RF creates many classification trees from subsets of the data created using the bootstrap method with replacement. To determine the optimal number of trees to grow for the RF classifier the out-of-bag classification error was calculated over a number of investigated forests sizes. The optimal number of trees grown was found to be 30.

2.4.2 K Nearest Neighbour

$k - nn$ is a method of pattern recognition whereby new R-R intervals are classified as NSR or AF based on the majority vote of the number of closest surrounding neighbors in feature space. A standard 10-fold cross validation was performed on the training dataset to determine an optimal value for k, which in this case was 17.

2.5 Statistical Analysis

2.5.1 Algorithm Training

For all four implemented R-R irregularity measurement receiver operating characteristic (ROC) curves were created and the area under the ROC curve (AUC) calculated. From each ROC curve the optimal detection thresholds were defined as the minimum Euclidean distance from perfect classification. Defined thresholds were then taken and applied to the testing dataset.

2.5.2 Algorithm Testing

The performance on the four R-R irregularity measurement and two classification models were assessed using overall beat-by-beat sensitivity, specificity and positive predictive value (PPV). As well as this individual beat-by-beat accuracy was calculated and confidence intervals created using bootstrap with replacement over 1000 bootstrap trials. Finally, significant differences in accuracy calculations across all six detection approaches on the testing dataset were determined using the Wilcoxon signed rank test ($\alpha = 0.05$).

2.6 Results

2.6.1 Algorithm Training

When assessing the performance of the four R-R irregularity measurements on the 249 patients of the $AFDB_{comb}$ training dataset (results presented in Table 2), $CoSEn$ performed best with AUC of 0.92 followed by RMSSD with AUC of 0.91. The ROC plots shown in Figure 2 are provided to allow comparison between the four different automated AF detectors on the $AFDB_{comb}$ and $MIT - BIH\ AFDB$. Outputs of each R-R irregularity measurement were then used as input features to develop the RF and $k - nn$ classification models. Optimal detection thresholds were chosen from the ROC plot of the $AFDB_{comb}$ (Figure 2a) and defined as the minimum Euclidean distance from perfect classification.

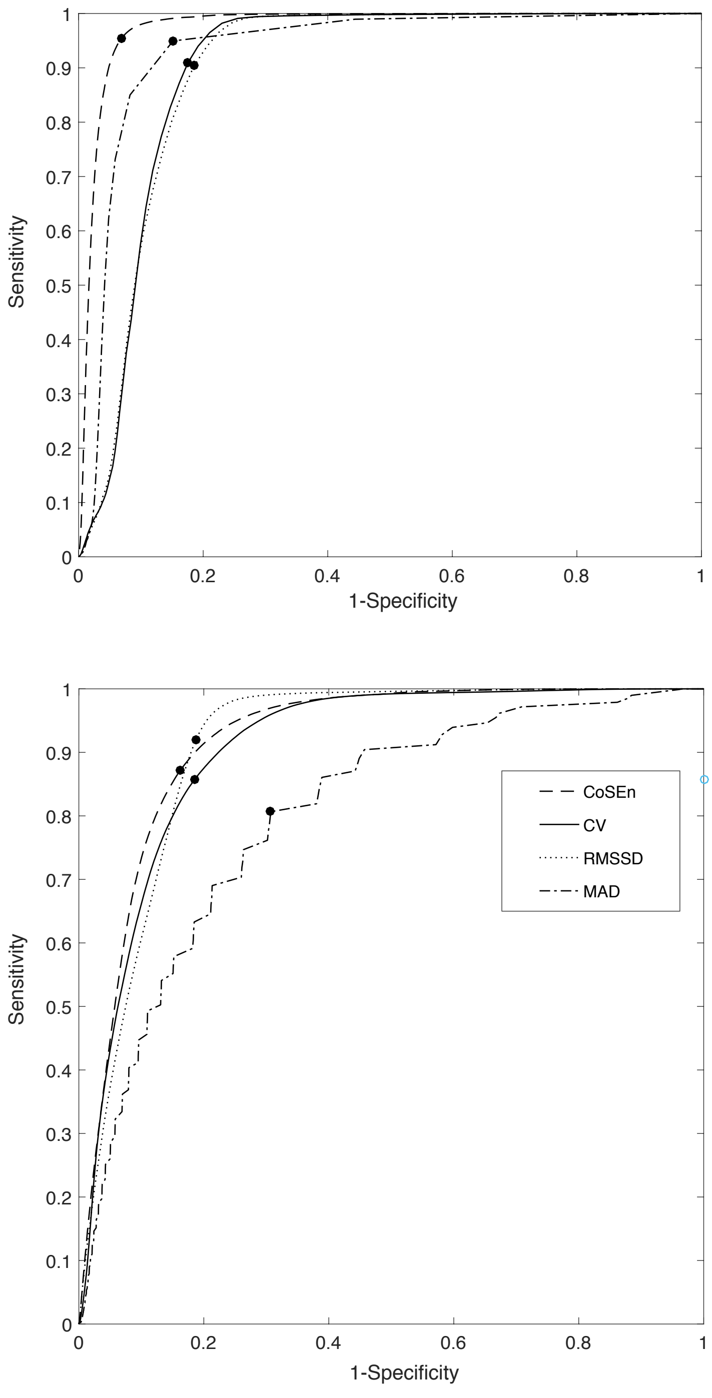

Figure 2. ROC curves for AF detection for the four implemented R-R irregularity measurement on (a) $AFDB_{comb}$ training set and the (b) $MIT-BIH\ AFDB$. Where (o) is the optimal detection thresholds.

Database	CoSEn	CV	RMSSD	MAD
$MIT-BIH\ AFDB$	0.97	0.88	0.89	0.92
$AFDB_{comb}\ training$	0.92	0.90	0.91	0.80

Values represent area under the ROC curve

Table 2. Area under the ROC curve values for the $MIT-BIH\ AFDB$ and the $AFDB_{comb}$ training dataset used in this study. The performance of the $CoSEn$ and MAD algorithms are reduced on the $AFDB_{comb}$ dataset compared to the $MIT-BIH\ AFDB$.

2.6.2 Algorithm and Classification Model Testing

From the 73 patients of the $AFDB_{comb}$ testing dataset (results presented in Table 3), RF classification improved AF detection over $CoSEn$ with overall specificity of 80.1% vs. 98.3% and positive predictive value of 51.8% vs. 92.1% with a reduction in sensitivity, 97.6% vs. 92.8%. $k-nn$ also improved specificity and PPV over $CoSEn$ however the sensitivity of this approach was considerably reduced (68.0%). When assessing R-R irregularity approach's individually (Figure 3), the $CoSEn$ provided the best AF detection accuracy compared to the next best algorithm $RMSSD$ (median accuracy, 96.9% vs. 92.7%, $p < 0.001$).

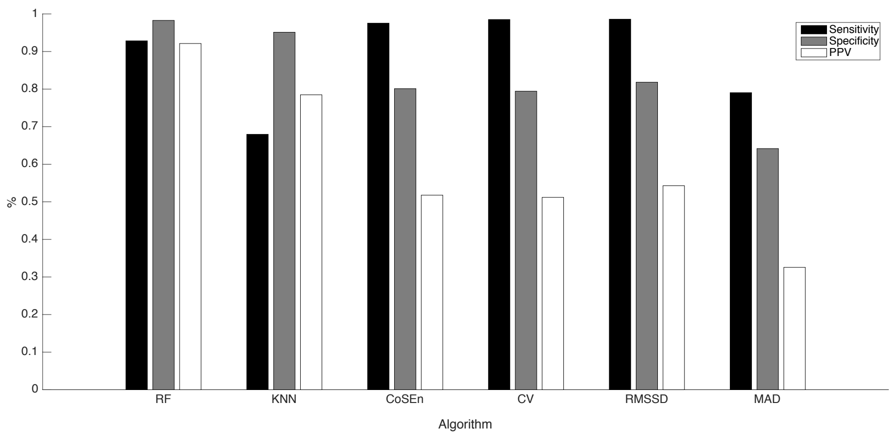

Figure 3. The individual accuracy of the six investigated detection algorithms. The RF model improved the median accuracy but the difference is very small however the 5% and 95% bootstrap confidence intervals are much smaller for RF.

The individual patient accuracy calculations, shown in Figure 3, demonstrated a small improvement in median accuracy for AF detection from RF of 99.1% compared to 96.9% achieved by $CoSEn$. Also of interest from these results is the clear reduction in bootstrap confidence intervals. To determine the performance of the AF algorithms on recordings containing ectopic beats, records from the testing dataset containing PVCs and PACs were isolated and median AF detection accuracy calculated to establish performance.

As shown in Table 4, the RF model appeared to be more effective on both records with a PVCs with median accuracy of 98.6% compared to 85.7% achieved by $CoSEn$. Also was true with the records containing PACs with median accuracy of 99.1% compared to 94.0% achieved by CoSEn. Although there are a limited number of recordings containing annotated PVCs and PACs in the testing dataset of this study, results from these isolated records appear to indicate the RF model is more effective in reducing false positives due to ectopic beats. However further investigation is required.

	$\boldsymbol{CoSEn}$	$\boldsymbol{RF}$
PACs (n = 11)	0.94	0.99
PVCs (n = 36)	0.86	0.99

Values represent median accuracy

Table 4. The median accuracy of $CoSEn$ and RF algorithm on recordings containing annotated PVCs and PACs, the two major sources of false positives from AF algorithms, from the testing database.

2.7 Discussion

Recently a number of studies have demonstrated improvement performance of automated AF detection on the MIT-BIH AFDB [13], [18]–[20] . As highlighted in Figure 2, new databases should be explored in order to develop more robust methods of automated AF detection. The MIT-BIH database is commonly used for reporting new algorithm performance and remains the most used database for algorithm testing and comparison. In this study we assembled a large database which may be more appropriate for automated AF algorithm training and testing. Performance of algorithm training on the training dataset is compared to the MIT-BIH database in Figure 2. The performance of CoSEn and MAD were significantly reduced when employing $AFDB_{comb}$ during development in comparison to $MIT - BIH\ AFDB$. This in turn highlights that results obtained using the $MIT - BIH\ AFDB$ may be optimistic and not truly representative. During algorithm testing, the best performing R-R irregularity method for diagnosing AF was the $CoSEn$, which provided a significant improvement in median accuracy over the next best method the $RMSSD$. This observation falls in line with the work described by Langley et al. [21] who demonstrated the improved accuracy of $CoSEn$ over CV and RMSSD for short R-R interval recordings. When Moorman and Lake first described the principle of the CoSEn [14] they also demonstrated its improved performance over CV in classifying AF from NSR. Petrenas et al. [22] investigated the performance of $CoSEn$ for AF detection from the R-R interval and found that for small segment window lengths (<30 beats) the sensitivity, specificity and accuracy of AF detection was significantly reduced in recordings with high signal noise, also the authors demonstrated a reduction in the performance of $CoSEn$ during PACs confirming observations made in [7]. The RF model described in this study improved the median accuracy over $CoSEn$ (99.1% vs. 96.9%, $p < 0.001$). Also, as shown in Table 4, the RF model provides improved accuracy from records containing PVCs and PACs. However, due to further testing of the developed classification models is required to effectively establish their performance in the presence of PVCs and PACs. This is important to note as PVCs and PACs occur frequently with the prevalence increasing with age [23] and are a major source of false positives from automated AF algorithms.

This study describes an improvement in the performance of automated AF detection algorithms that are based on R-R interval based detection methods, through implementation of a RF classification model. R-R interval irregularity is the most accessible ECG characteristic for AF detection [8]. It must be appreciated though that the R-R interval based approach does have limitations and these are rooted in the that features which are specific to the function of the atria such as P-wave analysis, the P-R interval or fibrillatory wave are not accounted for. However, analysis of these features is not without issue as they may be difficult to accurately record during ambulation due to the relatively low magnitude of the P-wave and fibrillatory wave on the ECG [24], [25].

2.8 Conclusions

Of conventional approaches the $CoSEn$ performed better than CV, $RMSSD$ and MAD in AF detection from the $AFDB_{comb}$ database. The $k-nn$ classification model improved specificity and PPV value over $CoSEn$ but with a substantial cost in sensitivity. The RF classification model also improved specificity and PPV with a smaller reduction in sensitivity.

2.9 Acknowledgments

This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation Action under Grant Agreement no. 643491. PATHway: Technology enabled behavioural change as a pathway towards better self-management of CVD (www.pathway2health.eu). This work has also been supported by the Department of Employment and Learning Northern Ireland, the Institution of Engineering and Technology and the Health Informatics Society of Ireland.

References

[1] S. S. Chugh, R. Havmoeller, K. Narayanan, D. Singh, M. Rienstra, E. J. Benjamin, R. F. Gillum, Y. H. Kim, J. H. McAnulty, Z. J. Zheng, M. H. Forouzanfar, M. Naghavi,

G. A. Mensah, M. Ezzati, and C. J. L. Murray, "Worldwide epidemiology of atrial fibrillation: A global burden of disease 2010 study," *Circulation*, vol. 129, no. 8, pp. 837–847, 2014.

[2] J. S. Healey, S. J. Connolly, M. R. Gold, C. W. Israel, I. C. Van Gelder, A. Capucci, C. P. Lau, E. Fain, S. Yang, C. Bailleul, C. A. Morillo, M. Carlson, E. Themeles, E. S. Kaufman, and S. H. Hohnloser, "Subclinical atrial fibrillation and the risk of stroke," *N Engl J Med*, vol. 366, no. 2, pp. 120–129, 2012.

[3] T. Sanna, H.-C. Diener, R. S. Passman, V. Di Lazzaro, R. A. Bernstein, C. A. Morillo, M. M. Rymer, V. Thijs, T. Rogers, F. Beckers, K. Lindborg, and J. Brachmann, "Cryptogenic Stroke and Underlying Atrial Fibrillation," *N. Engl. J. Med.*, vol. 370, no. 26, pp. 2478–2486, 2014.

[4] K. Harris, D. Edwards, and J. Mant, "How can we best detect atrial fibrillation?," *J. R. Coll. Physicians Edinb.*, vol. 42 Suppl 1, pp. 5–22, 2012.

[5] P. M. Barrett, R. Komatireddy, S. Haaser, S. Topol, J. Sheard, J. Encinas, A. J. Fought, and E. J. Topol, "Comparison of 24-hour Holter monitoring with 14-day novel adhesive patch electrocardiographic monitoring," *Am. J. Med.*, vol. 127, no. 1, pp. 95–e11, 2014.

[6] N. Lowres, L. Neubeck, G. Salkeld, I. Krass, A. J. McLachlan, J. Redfern, A. A. Bennett, T. Briffa, A. Bauman, C. Martinez, and others, "Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies. The SEARCH-AF study," *Thromb Haemost*, vol. 111, no. 6, pp. 1167–1176, 2014.

[7] S. Maddox, M. Hoskins, M. Lloyd, A. Mengistu, S. Rangaraju, L. Henriquez, B.

Davis, and F. Nahab, “High False Positive Rates of Atrial Fibrillation Detection Among Stroke Patients who Receive Medtronic Implantable Loop Recorders,” *Stroke*, vol. 47, no. Suppl 1, pp. A206–A206, 2016.

[8] S. Babaeizadeh, R. E. Gregg, E. D. Helfenbein, J. M. Lindauer, and S. H. Zhou, “Improvements in atrial fibrillation detection for real-time monitoring,” *J. Electrocardiol.*, vol. 42, no. 6, pp. 522–526, 2009.

[9] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. . Stanley, “PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals,” *Circulation*, vol. 101, pp. 215–220, 2000.

[10] M. Llamedo and J. P. Martinez, “Qrs detectors performance comparison in public databases,” in *Computing in Cardiology Conference (CinC), 2014*, 2014, pp. 357–360.

[11] A. Petr\.enas, V. Marozas, and L. Sörnmo, “Low-complexity detection of atrial fibrillation in continuous long-term monitoring,” *Comput. Biol. Med.*, vol. 65, pp. 184–191, 2015.

[12] K. Tateno and L. Glass, “Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and \$Δ\$RR intervals,” *Med. Biol. Eng. Comput.*, vol. 39, no. 6, pp. 664–671, 2001.

[13] S. Dash, K. H. Chon, S. Lu, and E. A. Raeder, “Automatic real time detection of atrial fibrillation,” *Ann. Biomed. Eng.*, vol. 37, no. 9, pp. 1701–1709, 2009.

[14] D. E. Lake and J. R. Moorman, “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.*, vol. 300, no. 1, pp. H319–H325,

2011.

[15] M. Carrara, L. Carozzi, T. J. Moss, M. de Pasquale, S. Cerutti, M. Ferrario, D. E. Lake, and J. R. Moorman, “Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy,” *Physiol. Meas.*, vol. 36, no. 9, p. 1873, 2015.

[16] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, “Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median,” *J. Exp. Soc. Psychol.*, vol. 49, no. 4, pp. 764–766, 2013.

[17] D. T. Linker, “Accurate, Automated Detection of Atrial Fibrillation in Ambulatory Recordings,” *Cardiovasc. Eng. Technol.*, pp. 1–8, 2016.

[18] X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. Zhang, “Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy.,” *Biomed. Eng. Online*, vol. 13, no. 1, p. 18, 2014.

[19] J. Lee, B. A. Reyes, D. D. McManus, O. Mathias, and K. H. Chon, “Atrial fibrillation detection using an iphone 4S,” *IEEE Trans. Biomed. Eng.*, vol. 60, no. 1, pp. 203–206, 2013.

[20] C. Huang, S. Ye, H. Chen, D. Li, F. He, and Y. Tu, “A novel method for detection of the transition between atrial fibrillation and sinus rhythm,” *IEEE Trans Biomed Eng*, vol. 58, no. 4, pp. 1113–1119, 2011.

[21] P. Langley, M. Dewhurst, L. Y. Di Marco, P. Adams, F. Dewhurst, J. C. Mwita, R. Walker, and A. Murray, “Accuracy of algorithms for detection of atrial fibrillation from short duration beat interval recordings,” *Med. Eng. Phys.*, vol. 34, no. 10, pp. 1441–1447, 2012.

[22] A. Petrėnas, L. Sörnmo, A. Lukoševičius, and V. Marozas, “Detection of occult paroxysmal atrial fibrillation,” *Med. Biol. Eng. Comput.*, vol. 53, no. 4, pp. 287–297, 2015.

[23] P. W. Macfarlane, “Comprehensive Electrocardiology,” *Media*, vol. 33, no. 2, p. 1792, 2011.

[24] A. Petrenas, V. Marozas, G. Jaru??evi??ius, and L. S??rnmo, “A modified Lewis ECG lead system for ambulatory monitoring of atrial arrhythmias,” in *Journal of Electrocardiology*, 2015, vol. 48, no. 2, pp. 157–163.

[25] A. Kennedy, D. D. Finlay, D. Guldenring, R. R. Bond, and J. McLaughlin, “Detecting the Elusive P-Wave: A New ECG Lead to Improve the Recording of Atrial Activity,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 2, pp. 243–249, 2016.