

Reliable PPG-based Algorithm in Atrial Fibrillation Detection

Shih-Ming Shan¹, Sung-Chun Tang², Pei-Wen Huang¹, Yu-Min Lin¹
Wei-Han Huang¹, Dar-Ming Lai³, An-Yeu (Andy) Wu¹

¹Graduate Institute of Electronics Engineering, National Taiwan University, Taipei, Taiwan

²Stroke Center and Department of Neurology, National Taiwan University Hospital, Taipei, Taiwan

³Division of Neurosurgery, Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan

¹{shan, pwhuang, yumin, weihan, andywu}@access.ee.ntu.edu.tw; ²tangneuro@gmail.com, ³002867@ntuh.gov.tw

Abstract—Atrial Fibrillation (AF) is the most common type of arrhythmia. Since AF is a risk factor for stroke, automatic detection of AF is an important public health issue. Currently, the most useful and accurate tool for diagnosing AF is electrocardiography (EKG). On the other hand, PPG-based AF detection desires exploration. Photoplethysmogram (PPG) is an alternative technique to obtain the heart rate information by pulse oximetry. Convenience makes PPG promising in identifying arrhythmia like AF. The aim of this study is to investigate the potential of analyzing PPG waveforms to identify patients with AF. With the extracted features from multiple parameters, including interval and amplitude of PPG signals, patients were classified into AF and non-AF by support vector machine (SVM). The receiver operating characteristic curve (ROC) and statistical measures were applied to evaluate model performances. Among 468 patients' signals recorded in clinic environments, we achieve ROC area under curve, sensitivity and accuracy of 0.971, 0.942, and 0.957, respectively. The result suggests that the PPG-based AF detection algorithm is a promising pre-screening tool to help doctors monitoring patient with arrhythmia.

Keywords—Atrial Fibrillation, Photoplethysmogram, Multiscale entropy, Shannon entropy, Support vector machine

I. INTRODUCTION

Atrial Fibrillation (AF) [1] is the most common type of arrhythmia, which is an abnormal heart rhythm characterized by rapid and irregular heart beating. Since AF is a risk factor for stroke, automatic detection of AF is an important issue of public health. A major challenge in AF detection is paroxysmal AF, of which AF occurs occasionally and then returns to normal shortly. The paroxysmal and asymptomatic nature of AF makes short-term monitoring insufficient to identify all types of AF. Currently, the most useful and accurate tool for diagnosing AF is electrocardiography (EKG). In contrary to normal sinus rhythm (NSR), a typical EKG signal in AF shows no P waves and an irregular ventricular rate. However, each EKG monitoring device has its limitations or drawbacks, such as inconvenience, the need of at least 2 channels or costly.



Fig. 1. Concept of PPG-based AF detection.

Photoplethysmogram (PPG) is a technique to obtain the oxygen saturation of blood and heart rate by pulse oximetry.

PPG can be measured from fingertips, wrists, or earlobes. Compared to EKG examination, PPG examination is more convenient and suitable for long-term monitoring. The concept of PPG AF detection is shown in Fig. 1. Therefore, PPG has high potential to identify paroxysmal AF. The related works [2] and [3] adopted PPG signals to detect occurrence of AF. However, state-of-the-art researches only utilized pulse-pulse interval (PPI) as a parameter. In addition, there is still performance gap between [3] and EKG-based algorithm.

Therefore, a long-term PPG-based AF detection framework is proposed in this work. Compared to the related works [2] [3], the proposed framework is based on the following: 1) Jointly analysis of PPG parameters 2) More traditional and non-linear features added to improve the performance of classification. 3) Reliable classification to avoid over-fitting and class imbalance. The comparisons are summarized in Table I. Experiment results shows that the performance gap between PPG-based detection and EKG-based detection is eliminated by proposed framework. As a result, the PPG-based AF detection is a promising pre-screening tool to help doctors monitoring patient with arrhythmia.

TABLE I. COMPARISON TABLE

	Jointly Analysis of Parameters	Variety of Features	Feature Selection	Classification Algorithms
Related work [3]	No	3	No	Basic
Proposed PPG Framework	Yes	More Considered	Yes	More Reliable

The rest of the paper is organized as follows. Section II introduces the related works of PPG-based AF detection. Section III presents the proposed PPG-based AF detection framework. Section IV presents the experimental setups and results. Section V is the conclusion.

II. RELATED WORKS

In previous studies [2] [3], AF detection using the PPG signals measured from iPhone were demonstrated. In the training phase, EKG was used to build the model. First, the parameter, R-R interval (RRI), was extracted from the EKG database. Second, selected features were extracted from RRI series. The models were built by finding the threshold values of the features that provided the largest area under the receiver operating characteristic (ROC) curves in binary classification

of AF and NSR. In the testing phase, the PPG signals measured by the studies were used to test their models. First, the parameter, the interval, was extracted from the PPG signal. Second, the same features from the training phase, were extracted from interval series. Then, these features were used to test the models. They achieved high accuracy in prediction and proved that PPG-based AF detection is a promising tool. The flowchart of the related works is shown in Fig. 2.

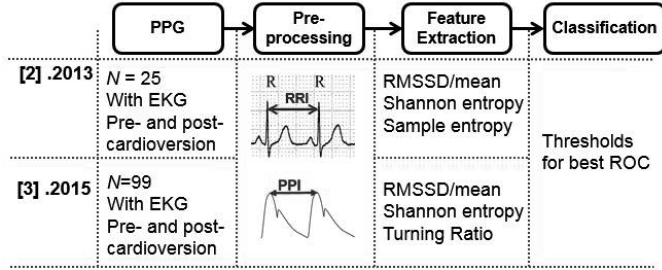


Fig. 2. Flowchart of related works.

III. PROPOSED AF DETECTION FRAMEWORK

In this study, a long-term PPG-based AF detection framework is proposed, as shown in Fig. 3. Input is the raw data and the output is the results of classification. Between the input and output are three different kinds of blocks. In pre-processing, parameters series were extracted from raw data. Then, the parameter series were used for feature extraction. Finally, the selected features were used in classification. The details of each block will be illustrated successively in the following sub-sections.

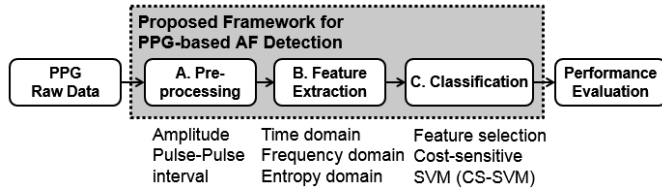


Fig. 3. Flowchart of proposed AF detection framework.

A. Pre-processing

The traditional Pan-Tompkins peak detection algorithm [4] and [5] were used to find the peaks of EKG and PPG signals, respectively. Both amplitude (AMP), and pulse-pulse interval (PPI) were extracted from PPG signals as parameters. And R-R intervals (RRI) were extracted from EKG signals for comparison as well. For example, if a patient's heart rate is 80, there's around 160 PPIs in the 2 minute record of PPG. The PPI series of length 160 is the output of the pre-processing block.

Since EKG is the most used and reliable tool for doctors to identify AF. Each patient's EKG signals were visually checked and labeled as AF or non-AF. The label is used as the golden outcome of the AF detection in this paper.

B. Feature Extraction

Except common linear features, three types of non-linear features, including multiscale entropy, Shannon entropy and turning point ratio, are applied. The details of each feature are illustrated as follows.

1) *Linear features* [6]: We use some of the most popular linear features in heart rate variety (HRV) researches, according to [6]. Linear features includes time domain features and frequency domain features.

Time-domain analysis measures the variation in heart rate over time or the intervals between successive normal cardiac cycles. Mean and standard deviation (SD) represent overall distribution. NN counts ratio (pNNR) and root mean square successive difference (RMSSD) is based on interval differences:

$$RMSSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - x_{i-1})^2} \quad (1)$$

Utilizing both SD and RMSSD, the features contain both short-term and long-term variations. Furthermore, to balance the variation in mean value, normalized SD and normalized RMSSD is calculated by dividing mean value.

Frequency-domain analysis describes the periodic oscillations of the heart rate signal. The signals are decomposed at different frequencies and amplitudes, providing information of relative intensity in the rhythm of the heart. Researches have found that frequency features are related to sympathetic and para-sympathetic nervous systems. We use FFT-based method and calculated the power in low-frequency (LF), power in high-frequency (HF) and LF/HF according to the definition in HRV researches.

2) *Multiscale entropy (MSE)* [7]: MSE is a non-linear method which has been widely used to evaluate the physiologic control mechanisms, such as heart failure and Alzheimer's disease [8]. In the case of AF detection, AF subjects have more unstable heartbeat, and thus prone to present higher entropy value. There are two steps in MSE. First, the multiple coarse-grained time series are calculated by averaging a successively increasing number of data points in non-overlapping windows:

$$y^\tau = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, 1 \leq i \leq N, 1 \leq j \leq \frac{N}{\tau}, \quad (2)$$

where x_i is the original time series, N is the length of x_i , and τ is the scale factor. In 2 minutes of PPG signal, we don't have enough pulses to calculate high scale MSE. So τ is set to be 3 due to sample entropy's data length dependency. This step generates different time series y^τ . Second, the sample entropy (SampEn) for each series y^τ is calculated as:

$$SampEn = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right], \quad (3)$$

where m is the pattern length and r is the similarity criterion. $B^m(r)$ is the number of two sets of simultaneous data points of length m have distance $< r$. And we set $m=2$ and $r=0.15$ in the experiment. SampEn is the conditional probability that a dataset, having repeated itself within a tolerance r for m points, will also repeat itself for $m+1$ points. Thus, the more complex the signal is, the higher the entropy value will be.

3) Other entropy domain features:

Like related work [3], we applied other entropy domain features to quantify the similarities in the signals. We briefly introduce Shannon entropy and turning point ratio in the following section.

Shannon entropy is a common entropy definition in information theory. Shannon's measure of information is the probability of symbols to represent the amount of uncertainty or randomness of data. However, how to define parameter series into symbols can be flexible. Distribution of parameters differs a lot among patients, so fixing thresholds between symbols is inappropriate. We sort parameter series into a fixed number of groups, which is closer to Shannon entropy's original definition. And outliers were removed to avoid trivial point. At last, we calculate the probability p_i of each bins and apply Shannon entropy:

$$E_shannon = - \sum_{i=1}^N p_i \log(p_i). \quad (4)$$

Turning point ratio (TPR) is proposed based on the nonparametric "Runs Test" to measure the randomness in a time-series [9], and the idea is used in EKG-RRI [10]. Each beat in a RRI series is compared to its two nearest neighbors and is defined a turning point if it is greater or less than two neighbors. TPR is the ratio of turning point to total data length L . TPR is higher when series is more random.

And TPR is applied on PPG parameters in the same way. Furthermore, we also build the modified version of TPR. The purpose is to find the turning point of the trend of series, which is different from the original definition.

For each parameter, we extract all these features as candidate features. All candidate features are listed in Table II.

TABLE II. ALL CANDIDATE FEATURES

Domain	Extracted Feature	Physiological Meaning
Time	Mean value (<i>Mean</i>), Median value (<i>Median</i>)	Average and median
	NN counts ratio (<i>pNNR</i>)	Successive ratio of parameter
	Standard deviation (<i>SD</i>)	Long-term variability
	Root mean square successive difference (<i>RMSSD</i>)	Short-term variability
Frequency	Power in very low-frequency range (<i>VLF</i>)	Low frequency term
	Power in low-frequency range (<i>LF</i>)	Sympathetic nervous systems: tensioned
	Power in high-frequency range (<i>HF</i>)	Parasympathetic nervous systems: relaxed
	Ratio of power in LF and HF (<i>LF/HF</i>)	Balance between the sympathetic and parasympathetic nervous systems
Entropy	Multi-scale entropy (<i>MSE</i>)	Matching patterns within parameter series
	Shannon entropy (<i>E_Shannon</i>)	Overall distribution of parameter series
	Turning point ratio (<i>TPR</i>)	Turning points of neighboring parameter series

C. Classification

The steps of classification consist of feature selection, and classification. The details of each step are illustrated as follows.

1) *Feature selection*: We have various candidate features in the proposed PPG based framework. But the combination of the features to achieve decent performance is unknown. Performing feature selection has advantages of potential higher accuracy, less computing and finding the dominant features for AF detection. Analysis of variance (ANOVA) can select the feature with p-value < 0.05 in a common way. But Most features have p-value < 0.05, making ANOVA inefficient in this case. We applied wrapper type feature selection to find the decent combination. The famous sequential forward selection algorithm (SFS) is chosen to find the best features heuristically. The selected features are shown in the result part.

2) *Cost sensitive SVM*: SVM is the classification technique adopted in our experiment. Among 468 subjects, only 104 of them are AF and 304 of them are non-AF, and the AF class is more important in pre-screening. Therefore, we encountered class imbalance problem. The "class imbalance problem" occurs due to the ratio between majority and minority class in the dataset and leads to lower sensitivity. It is a very common problem in practice [11]. Furthermore, the minority class is more important in most cases. So a cost-sensitive method is adopted to deal with the class imbalance problem.

For a separable case, SVM finds the decision boundary by solving the following optimization problem:

$$\min_{w,b,\xi} \left(\frac{\|w\|^2}{2} + C \sum_{i=1}^l \xi_i \right), \quad (5)$$

where $\xi_i > 0$ corresponds to the slack variable for the i^{th} sample; $C > 0$ is the cost of misclassifying the samples. The cost-sensitive SVM is implemented by modifying the cost of the minority class (positive class label as ξ_i^+) to $C * W$, where W is the penalty weight > 1 and the cost of the majority class (negative class label as ξ_i^-) to C . The optimization equation is modified as:

$$\min_{w,b,\xi} \left(\frac{\|w\|^2}{2} + CW \sum_{i=1}^{l+} \xi_i^+ + C \sum_{i=1}^{l-} \xi_i^- \right). \quad (6)$$

We performed the grid search on cost (C) and penalty weight (W) to find the best combination for highest accuracy. And we applied five-fold cross validation to avoid over-fitting in learning. The whole SVM procedure is supported by the famous machine learning toolbox LIBSVM [12].

IV. EXPERIMENT SETUPS AND RESULTS

A. Experimental Setups

Raw PPG and EKG data were recorded simultaneously from the intensive care unit (ICU) of stroke in National Taiwan University Hospital from February, 2012 to June, 2014. The first 2 minutes of EKG were visually checked and labeled as AF or non-AF, and EKG as well as PPG were used in the proposed framework. The sampling frequencies of EKG and

PPG are 512 Hz and 128 Hz, respectively. In each minute, we excluded the poor quality signals of less than 40 pulses or more than 150 pulses per minute. After outliers were excluded, the total number of subjects was 468.

To evaluate the performances of each framework, the accuracy, sensitivity, specificity, and receiver operating characteristic (ROC) curve of the frameworks are calculated according to the definitions shown in Fig. 4. The predicted class refers to the prediction results of the frameworks. The actual class refers to AF and non-AF labeled by doctors, which are defined as positive and negative, respectively. It is worthwhile to mention that sensitivity is the most important criteria in a pre-screening tool, because sensitivity means the percentage of AF patients detected in the purposed framework.

		Actual Class		
		+	-	
Predicted Class	+	True positive (TP)	False positive (FP)	$TP/(TP+FP)$ =PPV
	-	False negative (FN)	True negative (TN)	$TN/(TN+FN)$ =NPV
		$TP/(TP+FN)$ =Sensitivity	$TN/(TN+FP)$ =Specificity	$(TP+TN)/total$ =Accuracy

Fig. 4. Confusion matrix.

B. Performance Comparisons

To compare our proposed framework with the related works [3], a basis for comparison is needed. Our experiment data are collected in the ICU, the real clinical scenario. We applied their proposed feature RMSSD, Shannon entropy and TPR of PPI on our database to repeat their performance. Moreover, to compare the performance of both PPG and EKG examination, the EKG-based AF detection framework is applied in the same way. The result is shown in Fig. 5 and Table III.

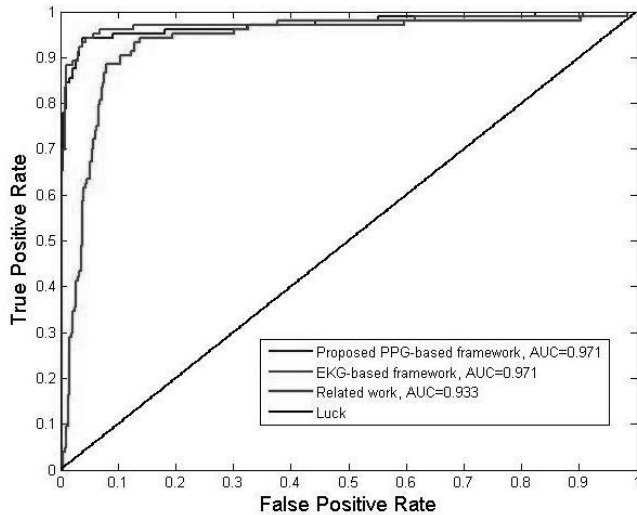


Fig. 5. Comparison of ROC of frameworks.

TABLE III. ACCURACY CRITERIA

	Accuracy	Sensitivity	Specificity	ROC AUC
Proposed PPG Framework	0.957	0.942	0.962	0.971
Related work [3]	0.895	0.904	0.893	0.933
EKG Framework	0.955	0.961	0.953	0.971

The result suggests that the proposed PPG-based framework has better performance than related works and comparable performance to EKG-based framework. So the proposed framework is potential for AF detection.

V. CONCLUSION

A PPG based framework utilizing multiple parameters is proposed to detect AF. Among 468 patients' signals recorded in clinic environments, we achieved performance comparable to EKG-based framework. The result suggests that the PPG-based AF detection algorithm is a promising pre-screening tool to help doctors monitoring patient with arrhythmia.

ACKNOWLEDGMENT

This work has been supported by: National Taiwan University (NTU)-National Taiwan University Hospital (NTUH)-MediaTek Innovative Medical Electronics Research Center, NTUH PC1000.

REFERENCES

- [1] V. Markides and R.J. Schilling, "Atrial fibrillation: classification, pathophysiology, mechanisms and drug treatment," *Heart*, vol. 89, no. 8, pp. 939-943, Aug. 2003.
- [2] J. Lee, B. A. Reyes, D. D. McManus, O. Maitas and K. H. Chon, "Atrial Fibrillation Detection Using an iPhone 4S," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 203-206, Jan. 2013.
- [3] J. W. Chong, N. Esa, D. D. McManus and K. H. Chon, "Arrhythmia Discrimination Using a Smart Phone," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 3, pp. 815-824, May 2015.
- [4] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230-236, March 1985.
- [5] M. Di Rienzo, P. Castiglioni, and G. Parati, "Arterial blood pressure processing," *Wiley Encyclopedia of Biomedical Engineering*, Apr. 2006.
- [6] Gernot Ernst, "Heart Rate Variability" Springer Science & Business Media, 2013, Chapter 4.
- [7] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E*, vol. 71, pp. 1-17, 2005.
- [8] P. H. Tsai, C. Lin, J. Tsao, "Empirical mode decomposition based detrended sample entropy in electroencephalography for Alzheimer's disease," *Journal of Neuroscience Methods*, vol. 210, pp. 230-237, Sep. 2012.
- [9] W. A. Wallis and G. H. Moore, "A Significance Test for Time Series Analysis," *Journal of the American Statistical Association*, vol. 36, pp. 401-409, 1946.
- [10] S. Dash, E. Raeder, S. Merchant and K. Chon, "A statistical approach for accurate detection of atrial fibrillation and flutter," in *Proc. Annual Computers in Cardiology Conference (CinC)*, Sep. 2009, pp. 137-140.
- [11] C. P. Chan and S. Stolfo, "Toward scalable learning with non-uniform class and cost distributions," in *Proc. Int'l Conf. Knowledge Discovery and Data Mining*, Aug. 1998, pp. 164-168.
- [12] C. C. Chang and C. J. Lin, "LIBSVM : a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, Vol.2, no. 27, pp. 1-27, 2011.