Identification of atrial fibrillation using descriptive statistic features and adaptive Neuro-Fuzzy inference system

Mar’atus Solikhah¹, Nuryani Nuryani¹ and Anto Satriyo Nugroho²
¹Physics Department of Graduate Program Sebelas Maret University, Jl. Ir. Sutami 36A Kentingan Jebres Surakarta 57126, INDONESIA
²Center for Information & Communication Technology, Agency for Assessment and Application of Technology, INDONESIA

E-mail: s.maratus@gmail.com, anto.satriyo@bppt.go.id

Abstract. Atrial fibrillation (AF) is one of the most common arrhythmia which can cause a serious problem. Nevertheless, well treated AF might not lead to any further complication. Early detection of AF could be an important preventative step that have to be conducted. In this article, we aim to make an automatic detection of atrial fibrillation. Seven descriptive statistic features have been utilized to detect AF. The features obtained could differ between two condition: normal and AF. Later, we use Adaptive Neuro-Fuzzy Inference System (ANFIS) as a classification method. Sugeno-type fuzzy inference system along with Gaussian type of fuzzy are utilized to classify the condition. The proposed method is applied on MIT-BIH Atrial Fibrillation database. The performance obtained from this proposed method might be considered for a medical application.

1. Introduction
Cardiac is an important organ for the human body which supplies blood throughout the body tissues. Cardiac disorders could greatly affect the quality of human life, even the most fatal could lead to death. World mortality rates data show that cardiac disease is the leading cause of death in the world [1]. One common type of cardiac disease is atrial fibrillation. Atrial fibrillation (AF), a type of arrhythmia, is a chaotic electrical activity in the atrium due to irregular electrical signals that trigger rapid and irregular heartbeat [2][3]. The prevalence of atrial fibrillation is quite high and indicates a progressive increase each year [4]. AF can also increase the risk of stroke and death [5].

Early detection of (AF) is needed especially with regard to paroxysmal and asymptomatic cases. This detection could be done by conducting ECG recording analysis. ECG is a device which commonly used in the measurement of cardiac electrical activity. An AF-based detection system could be made to facilitate an automated early detection thus reducing the risk of more serious AF complications. A study of arrhythmia detection has been conducted by researchers, including the research conducted by [6][7]. More specifically, one of the arrhythmia type that still be an interesting topic for researchers is AF. A Swarm Fuzzy Inference System (SFIS) based AF detection has been proposed by [8]. This approach used electrocardiographic P-wave as a feature. An atrial fibrillation detection which exploit statistic feature has also been done in many studies and shows a good performance. This features are mean, standard deviation [9], variance [10], Mean Absolute Difference (MAD), Median Absolute
Difference (MEAD) [11], Root Mean Square of Successive Difference (RMSSD) [12][9]. This methods show a good performance as well.

In this study, we employ descriptive statistic features of RR interval to make an automated detection system of atrial fibrillation. This features include standard deviation, mean/average, mode, range, minimum and maximum value of RR interval. Later, an ANFIS method are applied to classify between AF and normal condition.

The rest of this paper is organized as follows. The method section would explain about data and method used for AF detection, then followed by results section which explain the result obtained from this study. Finally, the last section of this paper is conclusion.

2. Numerical Methods

The data used in this research is an Atrial Fibrillation database from Massachusetts Institutes of Technology Beth Israel Hospital (MIT-BIH). This contains paroxysmal atrial fibrillation records from 25 patients with about 10 hours duration each and 250 Hz sampling frequency. In this study, we only use 23 data and exclude data record 00735 and 03665 due to unavailability of recorded signal data.

In general, the detection system of the proposed method is constructed as three main steps: pre-processing, feature extraction and classification.

2.1. Pre-processing

In the first step of pre-processing, the data of each patient is segmented into 3 different segments: 30 seconds, 60 seconds and 90 seconds. This segment is, then, given a target for each that differ between AF and non AF condition. The segment is considered as AF if it contains more than 30% of AF, otherwise it is denoted as normal. After the data are segmented, the RR interval of each segment is calculated. RR interval is the time interval between two consecutive R peaks in the ECG waveform. For the R peak, we collect the data provided by Physionet [13]. To obtain a better result, we remove outliers from RR intervals data before performing feature extraction. Here we utilize hampel filter for this outlier removal [14].

2.2. Feature extraction

The features are obtained by calculating statistics parameters i.e. standard deviation, mean/average, modes, range, minimum and maximum for each segment. The standard deviation ($s$) and average ($\bar{x}$) define as follows, which $n$ is the number of data element and $x_i$ is $i$-th element of the data.

\begin{align}
    s &= \left( \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \\
    \bar{x} &= \frac{1}{n} \sum_{i=1}^{n} x_i
\end{align}

(1) \hspace{1cm} (2)

2.3. Classification

The classification method used in this proposed study is Adaptive Neuro-Fuzzy Inference System (ANFIS), a Takagi-Sugeno-type fuzzy inference system with hybrid optimization. The structure of this method is similar to multilayer feed forward neural network but without weight. Basically, the architecture of ANFIS are constructed from five layer of nodes, two layers (i.e. the first and fourth layer) are adaptive nodes, while the others are fixed nodes. The classification process divided into two phase, training and testing. All the parameters in the training phase are tuned using a hybrid algorithm [15].
The output of classification then used to calculate the sensitivity \( (Se) \), specificity \( (Spe) \) and accuracy \( (Acc) \) value. These parameters are used to analyse the performance of the system made. These parameters are describe as follows.

\[
Se (%) = \frac{TP}{TP + FN} \times 100
\]

\[
Spe (%) = \frac{TN}{TN + FP} \times 100
\]

\[
Acc (%) = \frac{TP + TN}{TP + TN + FP + FN} \times 100
\]

TP (True Positive) is the number of AF detected as AF, while FP (False Positive) is the number of AF detected as normal. FN (False Negative) and TN (True Negative), respectively, is the number of normal detected as AF and normal detected as normal.

3. Results and Discussion

The RR intervals of AF has higher variability than normal as shown in Figure 1. The left and right side of the figure is the RR intervals during AF and normal respectively. The data fluctuation of RR intervals during AF are relatively high compared to the normal condition. For normal, there is no significant change in each RR intervals. This condition occurs due to characteristic of AF itself, which the RR interval values vary [2][16].

![Figure 1. RR intervals of AF and normal](image)

The variability of RR intervals could be indicated by the descriptive statistics value. In this proposed method, we evaluate seven descriptive statistic value namely standard deviation, mean/average, mode, range, minimum and maximum value of RR interval as features. Table 1 shows the features obtained from the feature extraction. Standard deviations and ranges indicate the distribution of the data In AF conditions, the standard deviation and range are higher than normal, then the AF data distribution is larger than normal. In contrast, the mean, median, mode and mean values of AF are lower because for AF it has an average length of RR intervals that is relatively shorter than normal.
Table 1. Descriptive statistics features during AF and normal for 90s segment

| Features              | Atrial Fibrillation (mean ± std, ms) | Normal (mean ± std, ms) |
|-----------------------|--------------------------------------|-------------------------|
| Standard deviation    | 31.81 ± 13.84                        | 18.67 ± 16.01           |
| Mean                  | 165.57 ± 33.70                       | 197.75 ± 43.78          |
| Median                | 165.57 ± 34.73                       | 197.87 ± 45.42          |
| Mode                  | 153.87 ± 135.00                      | 192.61 ± 145.00         |
| Range                 | 158.45 ± 65.66                       | 103.78 ± 79.22          |
| Max                   | 265.16 ± 69.99                       | 248.57 ± 60.97          |
| Min                   | 105.72 ± 21.10                       | 144.78 ± 56.30          |

The features obtained above are then used as inputs for the classification process. Table 2 shows the performance of the system created. This process is divided into two stages, namely the training and testing phase. At the training phase, the system performs learning to obtain the parameters used in the testing phase. The performance of each phase be known by calculating the sensitivity, specificity and accuracy value. In general, sensitivity shows how well a system can detect AF conditions, whereas specificity indicates the ability of the system to detect normal condition. Overall, system performance can be shown from its accuracy value.

In this study, we also conducted variations in segment length. These three segment variations show good results with performance values, both sensitivity, specificity and accuracy, above 75%. The greater the value of sensitivity, specificity and accuracy, then the system is said to have good performance. Of the three variations, the best results are obtained on the segment variation of 90 seconds with the sensitivity specificity and accuracy value of 81.45%, 90.81% and 87.38%, respectively. From this result, the system created considered to be used in medical applications with good performance.

Table 2. Performance of the proposed method for each segment

| Segment (second) | Training | Testing |
|------------------|----------|---------|
|                  | Sensivity (%) | Specificity (%) | Accuracy (%) | Sensitivity (%) | Specificity (%) | Accuracy (%) |
| 30               | 75.23     | 87.62   | 83.97       | 75.20          | 88.05           | 84.28        |
| 60               | 78.88     | 90.89   | 87.36       | 79.74          | 90.64           | 87.38        |
| 90               | 84.38     | 91.32   | 89.11       | 81.45          | 90.81           | 87.83        |

4. Conclusion
A study of atrial fibrillation detection using Adaptive Neuro-Fuzzy Inference System (ANFIS) has been conducted. Seven descriptive statistic features are utilized with three different segment variation. These features show a good result for AF detection. The best performance obtained when we performed 90s segment variation, which have the sensitivity, specificity and accuracy of 81.45%, 90.81% and 87.83% respectively. The performance obtained from this proposed method might be considered for a medical application.

5. References
[1] Colin DM and Dejan L 2006 PLoS Medicine 3 2011 – 30.
[2] Wolters K 2011 ECG Interpretation Made Incredibly Easy 5th edition (China: Lippicott-Williams & Wilkins)
[3] Amy ET 2015 The Journal of American Medical Association 313 10
[4] Chugh SS, Havmoeller R, Narayanan K, Singh D, Rienstra M, Benjamin EJ, Gillum RF, Kim YH, McAnulty Jr, Zheng ZJ, Forouzanfar MH, Naghavi M, Mensah GA, Ezzati M and Murray CJ 2014 *Circulation* **129** 837 – 47
[5] Crandall MA, Horne BD, Day JD, Anderson JL, Muhlestein JB, Crandall BG, Weiss JP, Osborne JS, Lappé DL and Bunch TJ 2009 *Pacing Clin Electrophysiol* **32** 981 – 6
[6] Mar’atus S, Nuryani and Darmanto 2015 *Jurnal Fisika dan Aplikasinya* **11** 36 – 40
[7] Kandala NVPSR and Ravindra D 2017 *Computers in Biology and Medicine* **87** 271 – 84
[8] Nuryani N, Bambang H, Iwan Y, Anik L, Eka A and Kemas F 2015 *Procedia Computer science* **72** 154 – 61
[9] Sung CT, Pei WH, Chi SH, Shih MS, Yen HL, Jiann SS, Dar ML, An YW and Jiann SJ 2017 *Scientific Reports* **7** 1 – 7
[10] Nuryani N, Mar’atus S, Anto SN, Adfal A and Eka A 2016 *Journal of Physics: Conference Series* **776** 012105
[11] Olli L, Tero H, Zuhair I, Sami N, Timo K, Antti S, Tuomas K, Tuija V, Juhani A, Mikko P and tero K 2018 *IEEE Journal of Biomedical and Health Informatics* **22** 108 – 18
[12] Jinseok L, Bersain AR, David DMM, Oscar M and Ki HC 2013 *IEEE Transaction on Biomedical Engineering* **60** 203 – 6
[13] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK and Stanley HE 2000 *Circulation* **101** 215 – 20
[14] Ronald KP, Yrjö N, Jaakko A and Moncef G 2016 *EURASIP Journal on Advances in Signal Processing* **2016**
[15] Nazmy TM, El-Messiry H and Al-Bokhity B 2010 *Journal of Theoretical and Applied Information Technology* **12** 1 – 6
[16] H Thomas M 2005 *Electrocardiography for The Family Physician: The Essentials* (Florida: Brown Walker Press)