Arrhythmia Classification of Electrocardiogram Recorded Data with Random Forest Method

Sutrisno Salomo Hutagalung, Dwi Esti Kusumandari, Yosafat Vincent Saragih, Jessica Tania, Arjon Turnip
Implementation Unit for Instrumentation Development, Indonesian Institute of Sciences, Indonesia.

Email: sutrisnosalomohutagalung@gmail.com

Abstract. Arrhythmia is a condition, that our heart beat rhythm change irregularly. The doctor do manual classification process to analyze and diagnose the heart beat rhythm from ECG record. We proposed Random Forest method as classification method, to solve the problem. To cut the preprocessing time, we use WFDB library. INCART arrhythmia database is provided as our training and testing data to build the classification model. The feature use is QRS amplitude, QRS amplitude Forward, QRS amplitude Backward, RR interval, Heart Rate Variance (HRV) on the QRS point, backward and forward. We using Scikit Learn for build our classification model, and tested using Scikit Learn and Weka. To classify our object data, we using FOGD-based QRS detector, and to provide our object dataset, Bitalino machine are used. The result still under consideration and need to validate by physician or cardiologist.

1. Introduction
Heart is a vital organ of human. The main function of heart, is to pump and distribute blood to entire body, so all human organs get sufficient blood supply that suit for their function. But unhealthy lifestyle; imbalance eating habit and lack of physical activity which benefit our health, and psychological problem like stress and depression, bring bad influence for heart function. The heart work improperly, that can be identified with irregular change of heart beat rhythm. The change of heartbeat irregularly, in medical term known as arrhythmia, can be lead into more fatal consequence for patient. Heartbeat that too slow or too fast in normal human body condition, can be sign for more complex cardiovascular illness. To help diagnose process of arrhythmia, doctor use ECG or Electro Cardiogram record from ECG machine. But this process has still done with manual procedure, that consume time [1].

To solve the time consumption problem in analyze the change of heart beat rhythm, today, the classification system that integrated with ECG machine are develop, with hope, the classification system can give the result automatically, as same as the ECG signal that received from electrodes that attached to patient chest. But a new problem comes, that the feature selection of ECG signal affect the accuracy of classification result. Some researcher using RR and morphological based feature [1][2][3][4], some of research using new
approaching like wavelet based [5][6][7]. The model of classification and the number of data which use in training and testing phase, also affect the classification result by the system. From the previous research result [8][3][9], some model give satisfied result, with accuracy more than 80%. But the accuracy that give by the system is pretty suspicious, so we need more parameters of per-class accuracy that can justified the result validity. Some of the result didn’t give good result for overall dataset [1][2], but if we analyze for accuracy class-by-class, the accuracy result pretty good. In this point, we propose a heart beat rhythm classification model to solve the problem upside. Random Forest method is choosen as classification model, because it can applied quickly. We also try to analyze the available features used for classification model, and choose the best features for the classification model.

2. Method
Heart is human organ that has function to pump and distribute blood to entire body[10]. When heart do contraction and relaxation, the heart do one beat. This contraction controlled by sinoatrial node, which has function to send signal that command heart muscle to contract. If the heart muscle do improper contraction in normal cycle, the irregular beat created. This condition named as Arrythmia.

Heart works by contracting heart muscle, which made rhythm while distribute blood to entire body. Start from sinoatrial node as the trigger for contraction, than spread to entire heart muscle [11]. This activity produce electric signal that propagate with the same way as signal movement. The result from the propagation is, the electric current change in the skin surface. This change receive by ECG electrodes and recorded. The position of ECG electrodes is illustrated from the Figure 1 beside.

![ECG electrode position](image)

**Figure 1.** ECG electrode position [1].

From this process, the change of the electric current from electric signal, received by electrodes, converted into digitalized ECG record, following by several process. The electric signal which received are preprocess by an amplifier with optical isolation. Than pass into high-pass filter and low pass-filter. Than the signal converted into digital signal that can we seen.

Random Forest is a classification and regression model based from tree. Random Forest consist of many decission tree model and receive a input. A Random Forest model use some tree model, which receive vector of features from subsample of dataset as input, and return class type as the output. Each tree in the model give a different result, and voting process use to determine data class as the output of the model [12] Random Forest are choosen, due the law of large number that applied in the model. The law reduce overfitting probability. Feature
selection from the data picked randomly, give better result classification. One of a researcher are using this method [13] for classify beat type.

3. Design And Application
The flowchart of the process can be seen at Figure 2. For the training and testing process of the system, we are using INCART dataset provided from Physionet.org. Because we only use one lead, we choose V1 lead. We divide the training and testing dataset into two part, the training dataset (record I01, I03, I05, I07, I09, I11, I13, I15, I17, I19, I21, I23, I25, I27, I29, I11, I13, I15, I17, I19, I31, I33, I35, I37, I39, I41, I43, I45, I47, I49, I51, I53, I55, I57, I59, I61, I63, I65, I67, I69, I71, I73, I75) and testing dataset (record I02, I04, I06, I08, I10, I12, I14, I16, I18, I20, I22, I24, I26, I28, I30, I12, I14, I16, I18, I20, I32, I34, I36, I38, I40, I42, I44, I46, I48, I50, I52, I54, I56, I58, I60, I62, I64, I66, I68, I70, I72, I74). The ECG feature that we use in our model are QRS Amplitude, QRS Amplitude from one QRS complex behind, QRS Amplitude from one QRS complex forward, RR interval, heart rate of the RR interval, heart rate of the one RR interval behind and heart rate of the one RR interval forward.

We also try the system using seven dataset provided by Technical Implementation Unit for Instrumentation Development, Indonesian Institute of Sciences, Indonesia. The dataset taken from 8 (eight) peoples, who can be assumed have normal condition. We using Bitalino ECG machine in Figure 3 to record our participant heart condition. To preprocess the heart of our patient, the method proposed in [14] are used. We want to investigate if the method are suitable for Bitalino unit.
Figure 3. Bitalino ECG unit.

The tester do two types of activities, from each of the activities, we record their heart activities using our Bitalino unit. Due ethical reason, all of our tester are male, because the placement of the electodes are in the chest. The tester activities are seating for 1 (one) minute and walking for 1(one) minute. The time interval between the activities are 30 (thirty) seconds.

4. Result and Discussion
We analyze our model using Weka software. From the result of Weka, there is data imbalance, where class –0 or Normal are dominated the dataset, both for training and testing. Table 1 and 2 show the disparity of each class.

| Class  | Total   | Percentage of population in dataset |
|--------|---------|-------------------------------------|
| Normal | 90.760  | 89.711%                             |
| VEB    | 8.900   | 8.797%                              |
| SVEB   | 1.283   | 1.268%                              |
| Fusion | 224     | 0.221%                              |
| Pace   | 2       | 0.0019%                             |
| Unknown| 0       | 0%                                  |

| Class   | Total   | Percentage of population in dataset |
|---------|---------|-------------------------------------|
| Normal  | 83.270  | 85.897%                             |
| VEB     | 12.757  | 13.159%                             |
| SVEB    | 856     | 0.883%                              |
| Fusion  | 54      | 0.055%                              |
| Pace    | 4       | 0.004%                              |
| Unknown | 64      | 0.066%                              |

The impact of the imbalance, the feature doesn’t has better class separation. The N or Normal class has bigger accuracy score, than another class. The result is, the lowest accuracy for Fusion or F class. From some publication, Paced and Unknown class are ignored from classification [2][5]. HR features has better performance to separate each class, due the result give more information about activity change of heart. The result of the classification for the testing set are described in Table 3.
Table 3. Testing result using testing dataset.

| Class    | TP   | FP   | P     | R     | F-M   |
|----------|------|------|-------|-------|-------|
| Normal   | 0.988| 0.75 | 0.988 | 0.988 | 0.914 |
| VEB      | 0.908| 0.013| 0.914 | 0.911 | 0.897 |
| SVEB     | 0.625| 0.005| 0.531 | 0.574 | 0.572 |
| Fusion   | 0    | 0    | 0     | 0     | 0     |
| Pace     | 0    | 0    | NA    | NA    | 0.5   |
| Unknown  | 0    | 0    | NA    | NA    | 0.5   |

After we conduct the classification test, we do classification test from subject dataset. For the result, we noted that this result is a preliminary result and not verified by cardiologist. We assumed all subject in healty condition. The result can be described at Table 4.

Table 4. Subject data classification result.

| Sample | Activity | Total Peak | N  | SVEB | VEB  | F  | Q  | U  |
|--------|----------|------------|----|------|------|----|----|----|
| 1      | Sit      | 89         | 88 | 0    | 1    | 0  | 0  | 0  |
| 2      | Walking  | 69         | 69 | 0    | 0    | 0  | 0  | 0  |
| 3      | Sit      | 80         | 80 | 0    | 0    | 0  | 0  | 0  |
| 4      | Walking  | 85         | 85 | 0    | 0    | 0  | 0  | 0  |
| 5      | Sit      | 82         | 82 | 0    | 0    | 0  | 0  | 0  |
| 6      | Walking  | 94         | 94 | 0    | 0    | 0  | 0  | 0  |
| 7      | Sit      | 91         | 91 | 0    | 0    | 0  | 0  | 0  |
| 8      | Walking  | 91         | 91 | 0    | 0    | 0  | 0  | 0  |
| 9      | Sit      | 79         | 68 | 6    | 5    | 0  | 0  | 0  |
| 10     | Walking  | 64         | 62 | 0    | 2    | 0  | 0  | 0  |
| 11     | Sit      | 71         | 63 | 1    | 7    | 0  | 0  | 0  |
| 12     | Walking  | 67         | 67 | 0    | 0    | 0  | 0  | 0  |
| 13     | Sit      | 78         | 77 | 0    | 1    | 0  | 0  | 0  |
| 14     | Walking  | 74         | 52 | 2    | 20   | 0  | 0  | 0  |
| 15     | Sit      | 55         | 55 | 0    | 0    | 0  | 0  | 0  |
| 16     | Walking  | 67         | 56 | 4    | 7    | 0  | 0  | 0  |

5. Conclusion
From the result of our experiment, we can conclude that the Random Forest has a good accuracy result. The overall accuracy is 97.1403539002114%, Normal (N) 98.9576078191426%, Supraventricular Ectopic Beat (SVEB) 88.51610803010111%, Ventricular Ectopic Beat (VEB) 62.73364485981309% and Fusion (F), Paced using Pacemaker (Q) and Unknown/Non-Beat (U) is 0.0%. For further research, better feature engineering are need to provide better accuracy. We also suggest for better classification model research, using deep-learn based classifier as proposed like [4] to
improve the result of accuracy and cut the time for feature engineering. We also suggest the research for better signal preprocessing to provide rich and excellent signal extraction from patient.

References

[1] E. J. da S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, “ECG-based heartbeat classification for arrhythmia detection: A survey,” Comput. Methods Programs Biomed., vol. 127, pp. 144–164, 2016.

[2] P. deChazal, M. O‘Dwyer, and R. B. Reilly, “Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features,” IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196–1206, 2004.

[3] Y. C. Yeh, C. W. Chiou, and H. J. Lin, “Analyzing ECG for cardiac arrhythmia using cluster analysis,” Expert Syst. Appl., vol. 39, no. 1, pp. 1000–1010, 2012.

[4] W. Li and J. Li, “Local Deep Field for Electrocardiogram Beat Classification,” IEEE Sens. J., vol. 18, no. 4, pp. 1656–1664, 2017.

[5] E. D. Ü beyli, “ECG beats classification using multiclass support vector machines with error correcting output codes,” Digit. Signal Process. A Rev. J., vol. 17, no. 3, pp. 675–684, 2007.

[6] H. El-Saadawy, M. Tantawi, H. A. Shedeed, and M. F. Tolba, “Hybrid hierarchical method for electrocardiogram heartbeat classification,” IET Signal Process., vol. 12, no. 4, pp. 506–513, Jun. 2018.

[7] K. Alfarhan, M. Y. Mashor, and M. Al-Qaisee, “Effects of the Window Size and Feature Extraction Approach for Arrhythmia Classification,” Biomimetics, Biomater. Biomed. Eng., vol. 30, pp. 1–11, 2017.

[8] E. J. Da Silva Luz, G. J. P. Moreira, L. S. Oliveira, W. R. Schwartz, and D. Menotti, “Learning Deep Off-the-Person Heart Biometrics Representations,” IEEE Trans. Inf. Forensics Secur., vol. 13, no. 5, pp. 1258–1270, 2018.

[9] G. Bortolan, I. I. Christov, and W. Pedrycz, “Hyperbox classifiers for ECG beat analysis,” Comput. Cardiol., vol. 34, pp. 145–148, 2007.

[10] S. H. Rampengan, Buku Praktis Kardiologi. 2014.

[11] M. S. Spach and J. M. Kootsey, “The nature of electrical propagation in cardiac muscle,” Am. J. Physiol., vol. 244, no. 1, pp. H3–H22, 1983.

[12] L. Breiman, “Random Forest,” pp. 1–33, 2001.

[13] N. Emanet, “ECG beat classification by using discrete wavelet transform and Random Forest algorithm,” 2009 Fifth Int. Conf. Soft Comput. Comput. with Words Perceptions Syst. Anal. Decis. Control, pp. 1–4, 2009.

[14] P. Kathirvel, M. S. Manikandan, S. R. M. Prasanna, and K. P. Soman, “An Efficient R-peak Detection Based on New Nonlinear Transformation and First-Order Gaussian Differentiator,” Cardiovasc. Eng. Technol., vol. 2, no. 4, pp. 408–425, 2011.

[15] M. F. Amri, M. I. Rizuyawan, and A. Turnip, “ECG signal processing using offline-wavelet transform method based on ECG-IOT device,” 2016 3rd Int. Conf. Inf. Technol. Comput. Electr. Eng., pp. 1–6, 2016.

Acknowledgement

This research is the development, who was conducted by Technical Implementation Unit for Instrumentation Development, Indonesian Institute of Sciences, Indonesia.