Multi-Distance Dispersion Entropy for ECG Signal Classification

https://doi.org/10.3991/ijoe.v18i07.30055

Sugondo Hadiyoso[1](✉), Suci Aulia[1], Indrarini Dyah Irawati[1], Mohamad Ramdhani[2]
[1] School of Applied Science, Telkom University, Bandung, Indonesia
[2] School of Electrical Engineering, Telkom University, Bandung, Indonesia
sugondo@telkomuniversity.ac.id

Abstract—Automatic detection of heartbeat is critical for early cardiovascular disease prevention and diagnosis. Traditional feature methodologies based on expert knowledge cannot abstract and represent multidimensional and multi-view information. Hence traditional research on heartbeat detection pattern recognition cannot produce adequate results. The proposed method in this research used Dispersion Entropy (DisEn) on Multidistance Signal Level Difference (MSLD) for feature extraction and Support Vector Machine (SVM) method for classifying the ECG signals. The ECG datasets used in this research were obtained from the MIT-BIH Arrhythmia database. The experiments result using 5-fold cross-validation revealed that at distance D= 1-15 had the highest accuracy of 91% to classify the ECG data into Normal Sinus Rhythm (NSR), Left Bundle Branch Block (LBBB), and Atrial Fibrillation (AFIB) from the MIT-BIH Arrhythmias database.

Keywords—electrocardiogram, dispersion entropy, SVM, multidistance signal level difference

1 Introduction

The electrocardiogram (ECG) is one of the most frequently used techniques for diagnosing heart disease [1], [2]. The ECG records the heart's electrical activity over time by placing electrodes on the skin[3]. This signal is utilized to diagnose heart health and can be accessible by grasping the essentials of its distinctive waves: T, U, P, and QRS complex [4]-[6]. Our previous work was implemented an ECG classification based on information theory to decompose the ECG signals and employed the SVM algorithms as classifiers [7]. The results conduct the highest accuracy is 81.1%, with sensitivity and specificity of 89.8% and 79.4%. However, the proposed method still needs to be developed to assist clinical diagnosis. Therefore in this work, we proposed optimization of the performance from the previous experimental[7] to classify the ECG signals into three classes: Normal Sinus Rhythm (NSR), Left Bundle Branch Block (LBBB), and Atrial Fibrillation (AFIB).

Some research has been done in the past to classify cardiac disease based on the ECG signal. Some of them were processed to time-domain feature extraction [8]-[13].

Whereas DisEn was just released and performed excellently, suggesting that it should be examined further, particularly in biosignal cases[14]-[16].On the other hand, the prior studies for classifying ECG signals widely used the Support Vector Machine (SVM) approach [17]. Our previous research [18], had been simulated the sample entropy on Multidistance Signal Level Difference (MSLD) for feature extraction and the SVM method for epileptic EEG signal classification. The result was conducted with the highest accuracy of 97.7%. By calculating the entropy at different signal levels, it is thought that it will increase the classification accuracy.

According to the previous related works above, feature extraction plays a critical role in pattern identification, particularly for ECG data. In this research, we used the DisEn on Multidistance Signal Level Difference (MSLD) for feature extraction to describe the dynamics of biological signals and SVM as a classifier. The feature extraction step generates 20 values as feature vectors for each MSLD signal with a distance D of 1 to 20. Finally, the performance evaluation of the proposed method uses 5-cross validation to divide training data and test data. The test was carried out on three different ECG data classes: normal sinus rhythm (NSR) (283 datasets), AFIB (135 datasets), and LBBB (103 datasets) from an accessible MIT-BIH Arrhythmia database.

2 Material and methods

2.1 ECG dataset

The datasets used in this research was obtained from an accessible the MIT-BIH Arrhythmia database of ECG signals on http://www.physionet.org PhysioNet [19]-[21]. The ECG signals of forty-five patients were captured using 200 [adu/mV] amplification and a sampling frequency of 360 Hz. One lead is used to record signals. This database has 17 different ECG classes: normal, cardiac rhythm, and 15 abnormal ECG classifications. In this work, the ECG signals used consist of 283 datasets of normal sinus rhythm (NSR) class, 135 datasets of Atrial Fibrillation (AFIB) class, and 103 datasets of Left Bundle Branch Block (LBBB) class. The visualisation of three classes the ECG signals are shown in Figure 1.

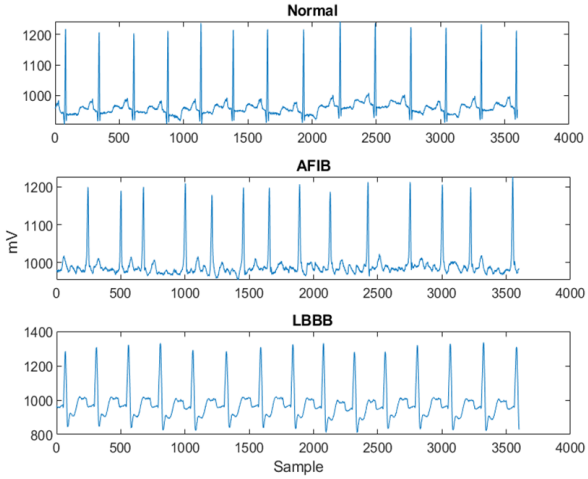

Fig. 1. The visualisation of Normal sinus rhythm, AFB, and LBBB signal

2.2 Multi-distance signal level difference

Multi-distance signal level difference (MSLD) is a method for generating a new signal by calculating the delta of two sample points of the signal according to a specified distance. MSLD is a modification of the grey-level difference (GLD) method [18]. The difference with GLD, MSLD is calculated as a one-dimensional (1D) vector. With MSLD, it can be calculated the pattern of occurrence of one sample point to another sample point at a specific distance. Equation 1 shows the MSLD formula used in this study.

$$S_d(i) = |x(i) - x(i+d)| \tag{1}$$

Where i is sample point (1, 2, …, N- d), d is specific distance (in this study d = 1-20), and S_d is the new signal with distance d. The MSLD computation with distance d is illustrated in Figure 2.

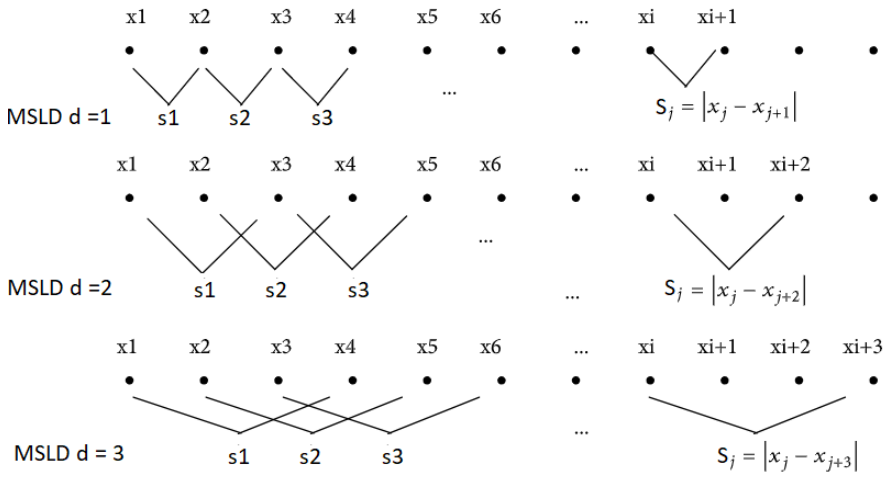

Fig. 2. The MSLD computation

2.3 Dispersion entropy

Entropy in a signal can be interpreted as a measure of the uncertainty or dynamics of the signal. Entropy is thought to represent the complex nature of biological signals. If there is an abnormality in a system, it can cause changes in signal complexity. The basis for calculating entropy is Shannon entropy which was later developed into several methods including approximate entropy, Renyi entropy, Tsallis entropy, sample entropy, and permutation entropy. Recently, dispersion entropy (DisEn) has received great attention, where DisEn has been shown to outperform sample entropy and permutation entropy.

Dispersion entropy was first introduced by Rostaghi & Azami in 2016 [16]. Dispersion entropy transforms data into new signals with several predetermined patterns and then the probability of the occurrence of these patterns is calculated. The DisEn calculation method is based on a new signal pattern mapping function with parameters: length m template; the number of classes c represents the number of patterns, and the delay time d. In this study, m=2, c=6, and d=1 are used. For each c^m of potential dispersion pattern, the frequency is calculated using the equation:

$$p(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{Number\{i|i \leq N-(m-1)d, z_i^{m,c} has\ type\ \pi_{v_0 v_1 \dots v_{m-1}}\}}{N-(m-1)d} \quad (2)$$

$p(\pi_{v_0 v_1 \dots v_{m-1}})$ represents the number of dispersion patterns $\pi_{v_0 v_1 \dots v_{m-1}}$ that are assigned to $z_i^{m,c}$. Based on Shannon entropy, then DisEn is calculated by:

$$DisEn(signal, m, c, d) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 v_1 \dots v_{m-1}}).ln(p(\pi_{v_0 v_1 \dots v_{m-1}})) \quad (3)$$

2.4 Performance evaluation

Figure 3 shows the proposed ECG signal classification system. Dispersion entropy is calculated for each new MSLD signal with a distance D of 1 to 20. This process generates 20 values as feature vectors. Finally, the performance evaluation of the proposed method uses 5-cross validation to divide training data and test data and SVM as a classifier. Figure 4 shows an illustration of the distribution of test data and training data with five iterations where each iteration is then evaluated for its performance. At the end of the step, the average performance is calculated. Performance parameters include accuracy, sensitivity, and specificity.

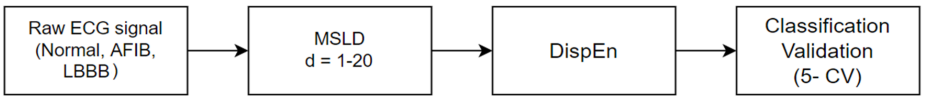

Fig. 3. Proposed system

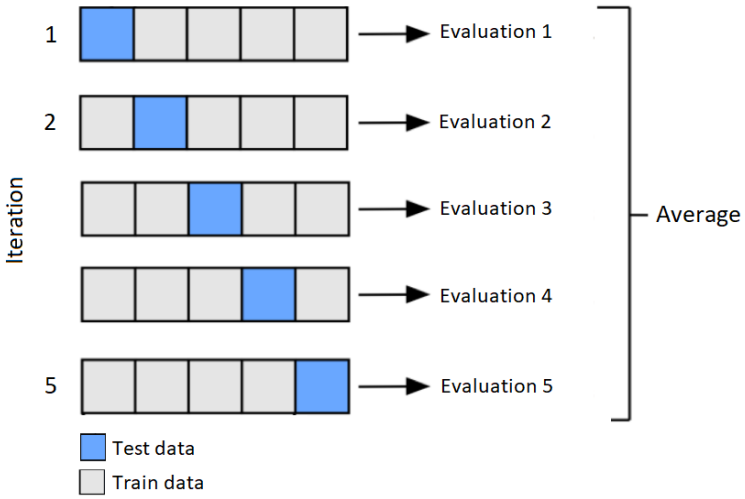

Fig. 4. The illustration of 5-cross validation

3 Results and discussion

Figure 5 depicts an example of the MSLD results at a distance of 1 to 5 for the AFIB signal. MSLD will generate a new signal for each distance. This new signal is the difference of two signal samples with a specified distance D. This new signal has a basic shape that is similar to the original signal, but the amplitude value is lower.

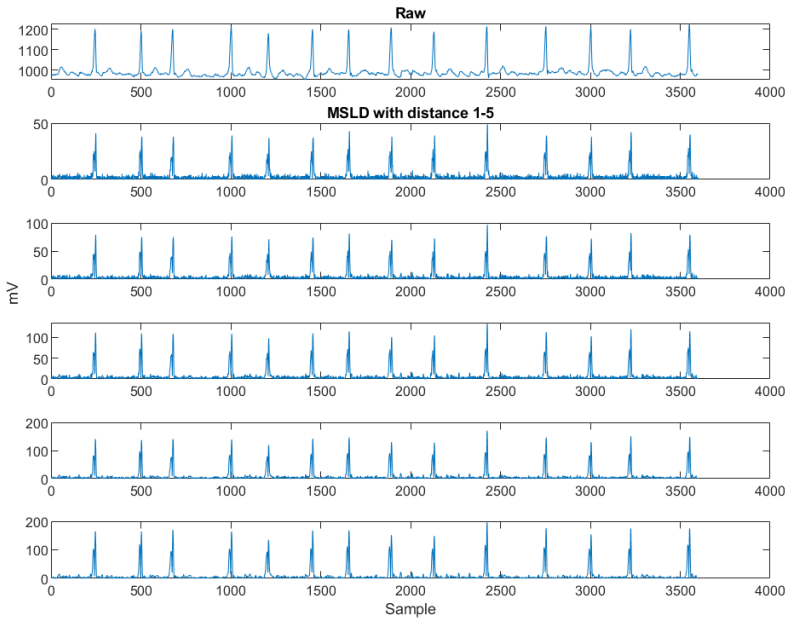

Fig. 5. MSLD results of AFIB

The average DisEn calculation results for each distance D for normal, AFIB, and LBBB are presented in Figure . Based on Figure 6, it can be seen that the normal ECG, AFIB, and LBBB generate different DisEn values as discriminant features. ECG LBBB has the greatest DisEn value compared to others. Normal ECG has a DisEn value between AFIB and LBBB. The significance test was not carried out in this study because the evaluation of the performance of the proposed method directly uses a classification simulation.

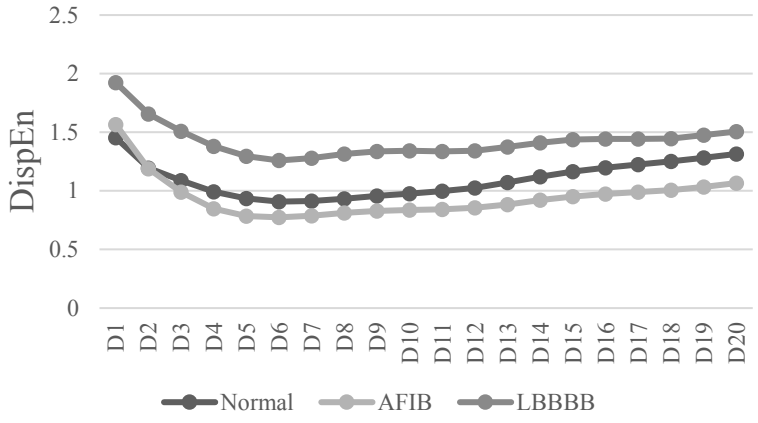

Fig. 6. Dispersion entropy for each signal and distance D

In the classification stage there is no special treatment in feature selection, only determined based on distance D=1-5, D=1-10, D=1-15, and D=1-20 for evaluation and finding the highest classification accuracy. Table 1 shows the accuracy for each scenario. The highest accuracy achieved is 91% with D=1-15 using Gaussian SVM. At D=1-5 and D=1-10 it produces the lowest accuracy because the DisEn features with these distances may not be significantly different. As with D=1-20, the feature D=16-20 has similar characteristics so that it produces an accuracy value that is not higher than D=1-15. So it can be concluded, MSLD-dispersion entropy with D=1-15 is the most best scenario that can be applied.

Table 1. Accuracy of each scenario

	D = 1-5	**D = 1 -10**	**D = 1-15**	**D = 1 - 20**
Linear SVM	63.5	68.9	80	80.8
Quadratic SVM	72.7	82.3	85.2	85.2
Cubic SVM	72.2	87.3	89.4	88.5
Gaussian SVM	81.4	87.7	91	90.2

Based on the confusion matrix shown in Table 2, it is known that AFIB produces the highest error. Some AFIB ECGs are classified as normal ECGs. Based on the characteristics shown in Figure 4, at D=1-10, AFIB and normal have a similar DisEn value so that it can reduce accuracy. The proposed method yielded a sensitivity and specificity of 90% and 92%, respectively. All performance parameters outperformed previous studies as reported in [7]. Another implication is that the measure of signal dynamics represented by dispersion entropy can be considered in the analysis and classification of ECG signal abnormalities.

Table 2. Confusion matrix for highest accuracy

		Predicted class			**Acc. (%)**	**Sens. (%)**	**Spec. (%)**
		AFIB	***LBBB***	***Normal***			
True class	AFIB	120	5	10	88.89	90	92
	LBBB	3	94	6	91.26		
	Normal	8	15	260	91.89		

4 Conclusion

We have succeeded in developing the ECG Signal Classification using a combination of MSLD and dispersion entropy methods. MSLD generates a new signal at each distance, while dispersion entropy describes the dynamics of biological signals so that analysis and classification of ECG signal abnormalities can be carried out. The system can classify into 3 classes including normal, AFIB and LBBB. Performance results are better than previous studies for maximum accuracy parameters of 91%, sensitivity of 90% and specificity of 92%. The Gaussian SVM method is superior to

other methods. In further research, this classification method can be developed into the introduction of ECG and determine the follow-up treatment based on the data obtained.

5 References

[1] R. J. Martis, U. R. Acharya, and L. C. Min, "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform," Biomed. Signal Process. Control, vol. 8, no. 5, pp. 437–448, Sep. 2013. https://doi.org/10.1016/j.bspc.2013.01.005

[2] M. Kaur and A. S. Arora, "Classification of ECG signals using LDA with factor analysis method as feature reduction technique," J. Med. Eng. Technol., vol. 36, no. 8, pp. 411–420, Nov. 2012. https://doi.org/10.3109/03091902.2012.702851

[3] P. Huang, "Study on a Low Complexity ECG Compression Scheme with Multiple Sensors," 2017, [Online]. Available: http://arxiv.org/abs/1704.01612

[4] M. C. T. Manullang, J. Simanjuntak, and A. L. Ramdani, "Implementation of AD8232 ECG Signal Classification Using Peak Detection Method For Determining RST Point," Indones. J. Artif. Intell. Data Min., vol. 2, no. 2, p. 61, 2019. https://doi.org/10.24014/ijaidm.v2i2.7593

[5] J. Pestana, D. Belo, and H. Gamboa, "Detection of Abnormalities in Electrocardiogram (ECG) using Deep Learning," in Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies, 2020, pp. 236–243. https://doi.org/10.5220/0008967302360243

[6] S. Sahoo, M. Mohanty, S. Behera, and S. K. Sabut, "ECG beat classification using empirical mode decomposition and mixture of features," J. Med. Eng. Technol., vol. 41, no. 8, pp. 652–661, Nov. 2017. https://doi.org/10.1080/03091902.2017.1394386

[7] S. Aulia and S. Hadiyoso, "Abnormal ECG Classification using Empirical Mode Decomposition and Entropy," J. Rekayasa Elektr., vol. 17, no. 3, pp. 191–198, 2021. https://doi.org/10.17529/jre.v17i3.22070

[8] W. Wei-quan, L. U. Pan, L. I. N. Jia-lun, and Z. Jin, "ECG Identification Based on Wavelet Transform," in Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC), 2016, no. Jimec, pp. 497–501. https://doi.org/10.2991/jimec-16.2016.89

[9] H. U. Si-yang, D. Kui, Q. Wang, Z. Chao-hao, Y. I. Yong-jie, and Y. Chang-rong, "A Detection and Analysis System of ECG Signals Based on Wavelet Entropy," IJSSST, pp. 1–6, doi: 10.5013.

[10] B. Natwong, P. Sooraksa, C. Pintavirooj, S. Bunluechokchai, and W. Ussawawongaraya, "Wavelet Entropy Analysis of the High Resolution ECG," in 2006 1ST IEEE Conference on Industrial Electronics and Applications, May 2006, pp. 1–4. https://doi.org/10.1109/ICIEA.2006.257276

[11] H. He, Y. Tan, and Y. Wang, "Optimal Base Wavelet Selection for ECG Noise Reduction Using a Comprehensive Entropy Criterion," Entropy, vol. 17, no. 12, pp. 6093–6109, Sep. 2015. https://doi.org/10.3390/e17096093

[12] H. Candra, M. Yuwono, R. Chai, H. T. Nguyen, and S. Su, "EEG emotion recognition using reduced channel wavelet entropy and average wavelet coefficient features with normal Mutual Information method," in 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2017, pp. 463–466. https://doi.org/10.1109/EMBC.2017.8036862

[13] M. C. Devi and S. Audithan, "Diagnosis of Breast Cancer using Wavelet Entropy Features," Int. J. Electr. Comput. Eng., vol. 10, no. 1, pp. 45–51, 2017.

[14] E. Kafantaris, I. Piper, T.-Y. M. Lo, and J. Escudero, "Application of Dispersion Entropy to Healthy and Pathological Heartbeat ECG Segments," in 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jul. 2019, pp. 2269–2272. https://doi.org/10.1109/EMBC.2019.8856554

[15] H. Azami and J. Escudero, "Amplitude- and fluctuation-based dispersion entropy," Entropy, vol. 20, no. 3, pp. 1–21, 2018. https://doi.org/10.3390/e20030210

[16] M. Rostaghi and H. Azami, "Dispersion Entropy: A Measure for Time-Series Analysis," IEEE Signal Process. Lett., vol. 23, no. 5, pp. 610–614, 2016. https://doi.org/10.1109/LSP.2016.2542881

[17] S. Hadiyoso, S. Aulia, and A. Rizal, "One-Lead Electrocardiogram for Biometric Authentication using Time Series Analysis and Support Vector Machine," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 2, pp. 276–283, 2019. https://doi.org/10.14569/IJACSA.2019.0100237

[18] A. Rizal and S. Hadiyoso, "Sample Entropy on Multidistance Signal Level Difference for Epileptic EEG Classification," Sci. World J., vol. 2018, pp. 1–6, Sep. 2018. https://doi.org/10.1155/2018/8463256

[19] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," IEEE Eng. Med. Biol. Mag., vol. 20, no. 3, pp. 45–50, 2001. https://doi.org/10.1109/51.932724

[20] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet," Circulation, vol. 101, no. 23, Jun. 2000. https://doi.org/10.1161/01.CIR.101.23.e215

[21] P. Pławiak, "Novel methodology of cardiac health recognition based on ECG signals and evolutionary-neural system," Expert Syst. Appl., vol. 92, pp. 334–349, Feb. 2018. https://doi.org/10.1016/j.eswa.2017.09.022

6 Authors

Sugondo Hadiyoso received the Master.in Electrical-Telecommunication Engineering from Telkom University, Bandung, Indonesia in March 2012. His research interests are wireless sensor network, embedded system, logic design on FPGA and biomedical engineer-ing. In 2018-present, he became a doctoral student in electrical engineering at the Bandung Institute of Technology. The focus of his doctoral research is signal processing and analysis of EEG waves (email: sugondo@telkomuniversity.ac.id).

Suci Aulia received the Master in Electrical-Telecommunication En-gineering from Telkom University, Bandung, Indonesia, in June 2012. Her research interests are image processing, biomedical imaging, and computer vision. In 2020-present, she became a doctoral student in electrical engineering at the Bandung Insti-tute of Technology. Her doctoral research focuses on biomedical imaging and analy-sis of pulmonary tuberculosis diagnosis (email: suciaulia@telkomuniversity.ac.id).

Indrarini Dyah Irawati obtained a doctoral degree in the School of Electrical and Information Engineering, Institute of Technology Bandung. She joined Telkom Applied Science School, Telkom University as an Instructor (2007-2019), Associate Professor (2019-present). Her main research interests are in the areas of compressive sensing, watermarking, signal processing, and computer network. She is currently a member of the Association for Computing Machinery (ACM) and the International Association of Engineers (IAENG) (email: indrarini@telkomuniversity.ac.id).

Mohamad Ramdhani completed his master's program in information technology at School of Electrical and Information Engineering, Institute of Technology Bandung. Since 2002 he has joined Telkom University as a lecturer and researcher, Associate Professor in the field of applied electronics (email: mohama-dramdhani@telkomuniversity.ac.id).

Article submitted 2022-02-08. Resubmitted 2022-04-06. Final acceptance 2022-04-12. Final version published as submitted by the authors.