ICIC Express Letters
Volume **16**, Number **12**, December **2022**

ICIC International ©2022 ISSN 1881-803X
pp. **1359–1366**

BIOMETRIC SIMULATION BASED ON SINGLE LEAD ELECTROCARDIOGRAM SIGNAL USING DISPERSION ENTROPY AND LINEAR DISCRIMINANT ANALYSIS

Suci Aulia[1], Sugondo Hadiyoso[1], Inung Wijayanto[2]
and Indrarini Dyah Irawati[1]

[1]School of Applied Science
[2]Faculty of Electrical Engineering
Telkom University
Jl. Telekomunikasi Terusan, Buah Batu, Bandung 40257, West Java, Indonesia
{ suciaulia; sugondo; iwijayanto; indrarini }@telkomuniversity.ac.id

Received December 2021; accepted March 2022

Abstract. *Biometrics based on Electrocardiograms (ECGs) has recently gained popularity. However, a person's ECG signal might vary emotionally or physically, making identification difficult. In this research, we offer an ECG-based authentication approach as biometrics to simulate a personal identification system. The dispersion entropy approach is provided as a feature extraction method of signal complexity. The validation of the performance system is carried out using Linear Discriminant Analysis (LDA) combined with 5-fold and 10-fold Cross-Validation (CV). A feature selection scenario is also applied to determining the best performance. The raw ECG signal is denoised using a High Pass Filter (HPF) with a cutoff frequency of 0.5 Hz and a Low Pass Filter (LPF) with a cutoff frequency of 50 Hz, then normalized to obtain an amplitude range of 0-1. The signal is segmented into ten segments with 6 seconds long and 50% overlap, then decomposed into five levels using Variational Mode Decomposition (VMD). DisEn is calculated for each decomposition signal. This result value is a feature vector in the validation process. The test results show the highest accuracy of 95%, with all feature variables used as predictors.*
Keywords: Biometrics, ECG, Dispersion entropy, Linear discriminant analysis, Variational mode decomposition

1. **Introduction.** Physiological and behavioral biometrics are the two types of biometrics [1,2]. Handwriting, face pattern, iris, fingerprints, DNA, hand geometry, and fragrance are examples of physiological biometrics. On the other hand, these biometric features are easily falsified and can be obtained forcefully or physically damaged. As a result, an innovative biometric system with a distinct feature that is challenging to falsify is needed [3]. Bio-potential or bio-signal is the model predictive or a biometric modality that meets the referred requirements because it has the potential to become a future biometric that is difficult to falsify. However, when compared to EEG signals, ECG has some advantages, such as a tendency to be linear, a regular rhythm or continuous signal, low complexity, and a relatively easy signal acquisition process [4,5]. Based on this interpretation, the ECG signal was chosen as the biometric modality in this study. The advantage of biometrics based on ECG signals is that the electrical activity of the human heart is almost impossible to replicate [6]. Furthermore, compared to other conventional biometric systems, the natural characteristics of biometrics have allowed for increased protection [7].

For the past few decades, ECG-based biometric authentication has risen in popularity as a next-generation potential tool [8,9] that has been applied with various approaches to optimize authentication efficiency [10-13]. Sample Entropy (SampEn) is widely used [14] for its effectiveness, and the commonly popular SampEn is based on ConEn and

DOI: 10.24507/icicel.16.12.1359

Permutation Entropy (PE) [15], based on the order relationships among signal values. PE is a simplified and computationally quick process, but it ignores amplitude mean values and variations between intervals [16]. The entropy metric of Dispersion Entropy (DisEn) was recently introduced to measure the uncertainty of time series [17]. It is fast, performed admirably in terms of time series characterization. An interval-based Linear Discriminant Analysis (LDA) algorithm is used to extract specific individual feature vectors that reflect distance and angle characteristics on short Autocorrelation Profile (ACP) segments for individual verification using ECG (Electrocardiogram) [18]. The Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) ECG tests yielded EERs (Equal Error Rates) of 0.143%, demonstrating that the proposed algorithm is both realistic and stable when confirming an individual's identity. In other research [19], the likelihood threshold was introduced into the LDA recognition system, effectively lowering the False Accept Rate (FAR) and False Rejection Rate (FRR) while increasing the Average Recognition Rate (ARR). LDA has been proposed in another study [20]. This research entails a thorough examination of the heartbeat. The results proved effective in resolving global accuracy up to 99.78% in the arrhythmia classification and 100% in the Atrial Fibrillation (AF) classification.

In this work, we proposed a biometric system based on ECG signals using a signal complexity approach. The degree of complexity of each individual's ECG signal is thought to be different from one another. Since the SampEn and PE have shortcomings, as described above, this study uses the dispersion entropy method to measure signal complexity. DisEn was recently introduced and gave an excellent performance, and it promises to be explored, especially in biosignal cases. The ECG signal was recorded using a single-lead ECG device, then dispersion entropy was used as a feature extraction method, and LDA was used as a classifier.

The contribution of this study is to develop ECG-based biometrics using a signal complexity approach. This study also confirms previous studies that the ECG signal is naturally unique and has a great opportunity as a new biometric system. The proposed ECG biometric system in this study is expected to be implemented in real-time authentication for high security access, as well as the opportunity to develop various methods to improve accuracy and evaluate feasibility by referring to biometric standards.

2. Material and Methods.

2.1. Proposed system.

The main purpose of this work is to extract the person's ECG signal and collect the feature for biometrics authentication to simulate a personal identification system. This research must pass several stages to obtain a high level of accuracy, as seen in the workflow in Figure 1.

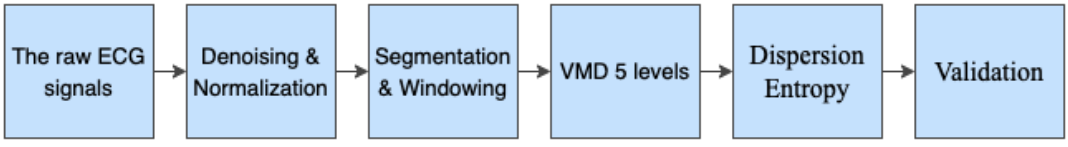

FIGURE 1. Workflow of the proposed method

The raw ECG signal is denoised using a High Pass Filter (HPF) with a f cutoff of 0.5 Hz and a Low Pass Filter (LPF) with f cutoff of 50 Hz, then normalized to obtain an amplitude range of 0-1. The signal is segmented into ten segments with 6 seconds long and 50% overlap, then decomposed into five levels using Variational Mode Decomposition (VMD). DisEn is calculated for each decomposition signal. This result value is a feature vector in the validation process.

2.2. **ECG recording.** The ECG signal was recorded using a single lead ECG device designed in [3]. Signal recordings were performed on 10 adult volunteers, both male and female, with a recording duration of 1 minute. The ECG signal is then cut into 10 segments with a length of 6 seconds and an overlap of 50%. It is intended for the simulation of training data and test data. A total of 100 ECG signal segments were simulated in this study. Figure 2 shows a sample of the ECG signal segments from each subject.

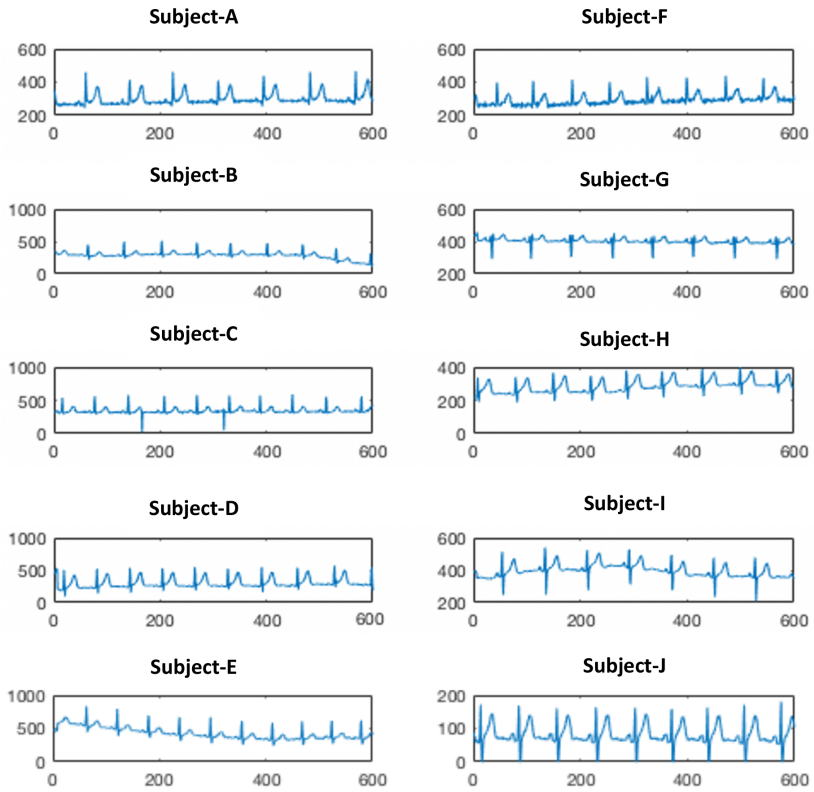

FIGURE 2. Sample ECG signal from each subject

2.3. **Dispersion Entropy (DisEn).** DisEn is a fast and reliable algorithm that uses Shannon entropy and semantic dynamics to evaluate an observed signal sample [15]. The method entails representing the dynamics of a signals sample using a dispersion pattern distribution. After the entire signal segment has been quantized with a defined scaling factor, dispersion characteristics are symbol sequences identified by the relative amplitude of observations of the whole signal sample analyzed. As a result, the signal segment can be represented with a small portion of dispersion patterns. It has a low DisEn value compared to that when all possible dispersion patterns must be considered equally likely. The segment has a maximum DisEn value. The following parameters are set for applying a DisEn algorithm [21]:

a) The number of observations (embedding dimension m) of each dispersion pattern being used reflects the signal segment;
b) The number of classes nc: the maximum number of different values a dispersion pattern observation can have;
c) A mapping method (logarithm sigmoid function) is used to spread the values of the dispersion pattern classes across the amplitude spectrum of the signal under investigation;
d) The time delay is used to add a pause to the samples that the algorithm analyzes.

The parameter model in this case refers to [15], where these parameters are commonly used. Parameters include embedding dimension $m = 2$, number of classes $nc = 6$, Mapping Approach (MA) using 'LOGSIG', and the time delay equal to 1.

2.4. **Linear Discriminant Analysis (LDA).** Linear discriminant analysis is a statistical technique that uses linear combinations of variables to divide a multivariate data set containing m variables into a number (k) of pre-defined classes [22]. LDA has two primary applications: grouping analysis and classification analysis [23]. LDA is used to classify the data in this article. LDA allows us to discriminate between various data types to the greatest extent possible, allowing us to identify the data correctly. In reducing the data dimensionality using LDA, the covariance matrix classes were defined as Equation (1) and Equation (2) [19,24].

$$S_W = \sum_{k=1}^{K} (x_n - m_k)(x_n - m_k)^T, \quad n \in C_k \tag{1}$$

$$m_k = \frac{1}{N_k} \sum x_n \tag{2}$$

m_k represents the global mean of the number of patterns N_k in class C_k, x_n represents the DisEn coefficient vector for the nth pattern, and k represents the number of classes in the results. S_B is the covariance matrix between classes with formula in Equation (3), m is the average mean of the data on k-class with formula in Equation (4), and S_T is the total covariance matrix as seen in Equation (5).

$$S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T, \quad n \in C_k \tag{3}$$

$$m = \frac{1}{N} \sum_{n=1}^{N} x_n = \frac{1}{N} \sum_{k=1}^{K} N_k m_k \tag{4}$$

$$S_T = S_W + S_B \tag{5}$$

Finally, the projection matrix is computed as W in Equation (6), and the LDA coefficients are found in Equation (7).

$$W = \arg_W \max \left\{ \left(W S_W W^T \right)^{-1} \left(W S_B W^T \right) \right\} \tag{6}$$

$$y = W^T x \tag{7}$$

where y is the vector of LDA coefficients and x is the DisEn coefficient vector in a decomposition level for a given sequence.

3. **Results and Discussion.** Figure 3 presents the results of the decomposition of the ECG signal into 5 levels. It can also be seen in Figure 3 that this decomposition is also used to reject baseline noise as well as the ECG signal in Figure 2, which may reduce system performance. The degree of signal complexity is then estimated on the signal series using DisEn. Thus, each analyzed ECG signal will generate five characteristics: DisEn IMF-1, DisEn IMF-2, DisEn IMF-3, DisEn IMF-4, and DisEn IMF-5. Figure 4 shows the average DisEn for each subject.

The average value of the DisEn feature, as presented in Figure 4, shows that there are differences in the characteristics of each ECG signal. Visually, it can be seen that DisEn1 and DisEn2 have significant differences compared to others. These differences could be because the level-3 to level-5 decomposition results tend to be monotonous.

The Kruskal Wallis method was used to test the difference with a 95% confidence level. Table 1 shows the significance test results, and it is known that all variables generate a P-value < 0.05. The lowest P-values were found in DisEn1 and DisEn2, proving that these two features have significant differences compared to the others.

Further validation is carried out using LDA combined with 5-fold and 10-fold Cross-Validation (CV). In this study, two cross-validation scenarios were used to test the robustness of the proposed method. In the validation simulation, feature selection was carried

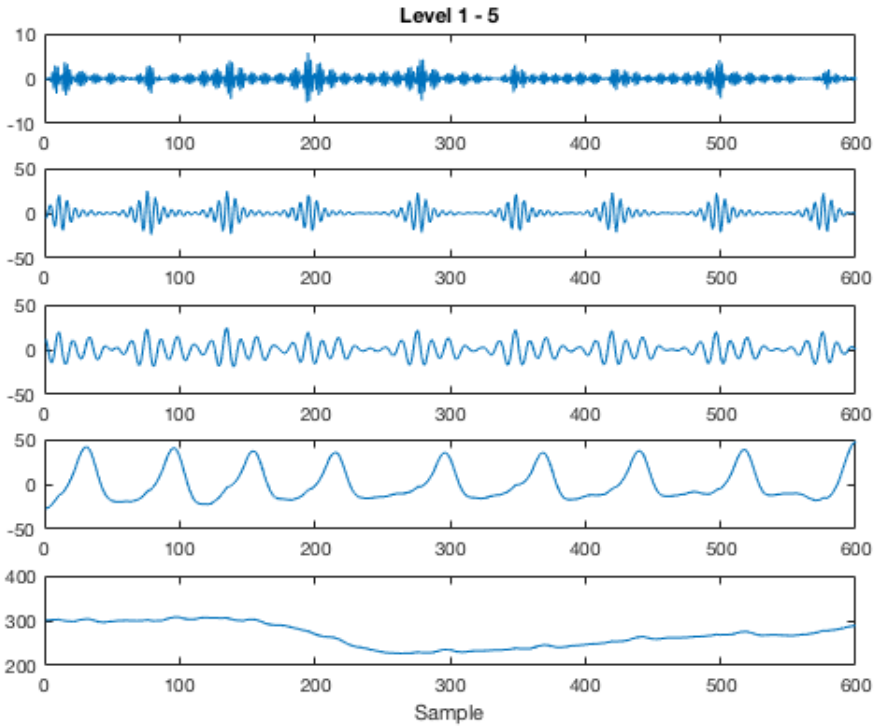

FIGURE 3. Signal decomposition result

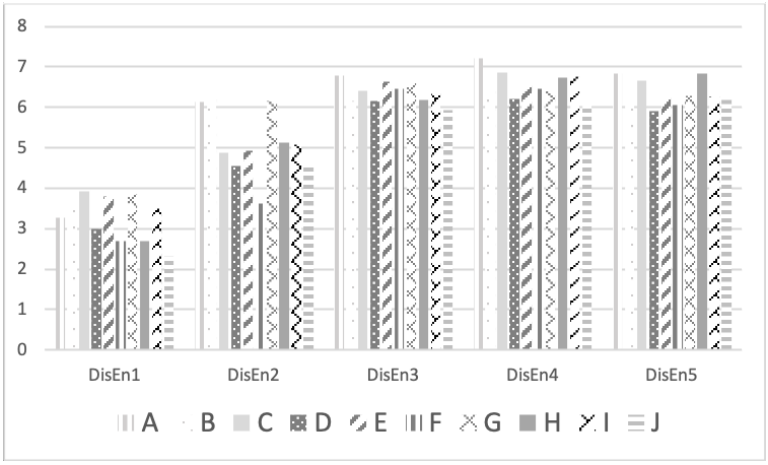

FIGURE 4. Average DisEn (subject-A to subject-J)

TABLE 1. Kruskal Wallis test result

Variable	P-value
DisEn1	6.65E-14
DisEn2	7.30E-16
DisEn3	9.35E-13
DisEn4	8.24E-13
DisEn5	2.67E-12

out with the scenario as presented in Table 2. Validation is performed using the learner toolbox in Matlab. The results of each simulation scenario are shown in Table 3.

The simulation results show that the highest accuracy is 95% by scenario V where all variables are predictors. In this scenario, both 5-CV and 10-CV produce the same

TABLE 2. Test result based on five scenarios

Scenario	Variable feature
I	DisEn1
II	DisEn1, DisEn2
III	DisEn1, DisEn2, DisEn3
IV	DisEn1, DisEn2, DisEn3, DisEn4
V	DisEn1, DisEn2, DisEn3, DisEn4, DisEn5

TABLE 3. Performance accuracy

Scenario	Acc. (5-CV)	Acc. (10-CV)
I	0.41	0.38
II	0.77	0.77
III	0.93	0.93
IV	0.93	0.91
V	0.95	0.95

TABLE 4. Confusion matrix from scenario V

True \ False	A	B	C	D	E	F	G	H	I	J	Acc.
A	10										100%
B		9					1				90%
C			10								100%
D	1			9							90%
E			1		9						90%
F						10					100%
G							10				100%
H								9		1	90%
I					1				9		90%
J										10	100%
Predicted											95%

accuracy. This result shows the robustness of the proposed method. These results also indicate that all variables affect the resulting accuracy, thus if the best accuracy is desired, all variables must be used as predictors in the classification process.

If we observe the confusion matrix in Table 4, misdetection only occurs in one signal in each subject. The validation scenario in this study is not appropriate for biometric applications. However, feature variables will always be classified to the most similar target. This is still a limitation and weakness of this study. The validation test is needed in the next work by measuring the valid rejection rate and false rejection rate.

DisEn, on the other hand, outperforms our earlier approach of using sample entropy [3] in terms of obtaining signal features. For authentication purposes, some research uses Support Vector Machine (SVM) algorithms to classify the signals. The 10-cross validation method was used in this research for the validation procedure. The Sample Entropy with the SVM Gaussian reached the highest accuracy at 86.2%. Compared to Sample Entropy, DisEn appears to be a promising candidate for implementation, with the highest accuracy value of 95%.

4. **Conclusions.** This study has simulated individual recognition based on ECG signals using the dispersion entropy approach as a feature extraction method from signal complexity. The characterization results show that each individual has a different degree of

signal complexity. The validation of the performance system is carried out using Linear Discriminant Analysis combined with 5-fold and 10-fold cross-validation. A feature selection scenario is also applied to determining the best performance. The test results show the highest accuracy of 95% in scenario V with 10-fold cross-validation, where scenario V uses DisEn1, DisEn2, DisEn3, DisEn4, and DisEn5 variable features. The limitation of this research is that the validation test applies cross-validation, which should be measured using the true rejection rate and false rejection rate. However, the result of the proposed method in this study can strengthen ECG opportunities for biometric applications.

REFERENCES

[1] P. L. Silva, E. Luz, G. Moreira, L. Moraes and D. Menotti, Chimerical dataset creation protocol based on Doddington Zoo: A biometric application with face, eye, and ECG, *Sensors*, vol.19, no.13, doi: 10.3390/s19132968, 2019.

[2] R. Palaniappan and K. Revett, PIN generation using EEG: A stability study, *Int. J. Biom.*, vol.6, no.2, pp.95-105, doi: 10.1504/IJBM.2014.060960, 2014.

[3] 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, doi: 10.14569/IJACSA.2019.0100237, 2019.

[4] W. Wang, P. Lu, J. Lin and J. Zhang, ECG identification based on wavelet transform, *Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC)*, pp.497-501, 2016.

[5] H. J. Kim and J. S. Lim, Study on a biometric authentication model based on ECG using a fuzzy neural network, *IOP Conf. Ser. Mater. Sci. Eng.*, vol.317, no.1, doi: 10.1088/1757-899X/317/1/012030, 2018.

[6] H. Silva, ECG biometrics: Principles and applications, *Proc. of the International Conference on Bio-Inspired Systems and Signal Processing*, pp.215-220, doi: 10.5220/0004243202150220, 2013.

[7] M. Pelc, Y. Khoma and V. Khoma, ECG signal as robust and reliable biometric marker: Datasets and algorithms comparison, *Sensors*, vol.19, no.10, doi: 10.3390/s19102350, 2019.

[8] J. Kim, G. Yang, J. Kim, S. Lee, K. K. Kim and C. Park, Efficiently updating ECG-based biometric authentication based on incremental learning, *Sensors*, vol.21, no.5, pp.1-17, doi: 10.3390/s21051568, 2021.

[9] A. Condon and G. Willatt, ECG biometrics: The heart of data-driven disruption?, *Biometric Technol. Today*, vol.2018, no.1, pp.7-9, doi: 10.1016/S0969-4765(18)30011-0, 2018.

[10] N. Neehal, D. Z. Karim, S. Banik and T. Anika, Runtime optimization of identification event in ECG based biometric authentication, *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, pp.1-5, doi: 10.1109/ECACE.2019.8679286, 2019.

[11] A. F. Hussein, A. K. AlZubaidi, A. Al-Bayaty and Q. A. Habash, An IoT real-time biometric authentication system based on ECG fiducial extracted features using discrete cosine transform, *Comput. Vis. Pattern Recognit.*, pp.1-6, http://arxiv.org/abs/1708.08189, 2017.

[12] A. Sellami, A. Zouaghi and A. Daamouche, ECG as a biometric for individual's identification, *2017 5th International Conference on Electrical Engineering – Boumerdes (ICEE-B)*, pp.1-6, doi: 10.1109/ICEE-B.2017.8192201, 2017.

[13] O. Al-Hamdani et al., Multimodal biometrics based on identification and verification system, *J. Biom. Biostat.*, vol.4, no.2, doi: 10.4172/2155-6180.1000163, 2013.

[14] A. Holzinger et al., On entropy-based data mining, in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, A. Holzinger and I. Jurisica (eds.), Springer, Berlin, Heidelberg, 2014.

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

[16] M. Zanin, L. Zunino, O. A. Rosso and D. Papo, Permutation entropy and its main biomedical and econophysics applications: A review, *Entropy*, vol.14, no.8, pp.1553-1577, doi: 10.3390/e14081553, 2012.

[17] M. Rostaghi and H. Azami, Dispersion entropy: A measure for time-series analysis, *IEEE Signal Process. Lett.*, vol.23, no.5, pp.610-614, doi: 10.1109/LSP.2016.2542881, 2016.

[18] C. Yang, G. W. Ku, J.-G. Lee and S.-H. Lee, Interval-based LDA algorithm for electrocardiograms for individual verification, *Appl. Sci.*, vol.10, no.17, p.6025, doi: 10.3390/app10176025, 2020.

[19] X. Zhang, H. Li, Q. Fan and G. Geng, LDA method based on probability threshold for ECG beat recognition, *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp.111-115, doi: 10.1109/ICCWAMTIP.2017.8301460, 2017.

[20] J. A. Queiroz, L. M. A. Azoubel and A. K. Barros, Support system for classification of beat-to-beat arrhythmia based on variability and morphology of electrocardiogram, *EURASIP J. Adv. Signal Process.*, vol.2019, no.1, p.16, doi: 10.1186/s13634-019-0613-9, 2019.

[21] E. Kafantaris, I. Piper, T.-Y. M. Lo and J. Escudero, Application of dispersion entropy to healthy and pathological heartbeat ECG segments, *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp.2269-2272, doi: 10.1109/EMBC.2019.8856554, 2019.

[22] 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, doi: 10.3109/0309 1902.2012.702851, 2012.

[23] S. Bharali, M. P. Sarma, K. K. Sarma and N. Mastorakis, Statistical and learning aided classifier for ECG based predictive diagnostic tool, *MATEC Web Conf.*, vol.125, doi: 10.1051/matecconf/2017 12505008, 2017.

[24] 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, doi: 10.1016/j. bspc.2013.01.005, 2013.