JURNAL INFOTEL
Informatics - Telecommunication - Electronics
Website Jurnal : http://ejournal.st3telkom.ac.id/index.php/infotel
ISSN : 2085-3688; e-ISSN : 2460-0997

Classification of ECG signal-based cardiac abnormalities using fluctuation-based dispersion entropy and first-order statistics

Naufal Juhaidi Jafal[1], Yuli Sun Hariyani[2*], and Sugondo Hadiyoso[3]
[1,2,3]School of Applied Science, Telkom University, Indonesia
[1,2,3]Bandung 40257, Indonesia
*Corresponding email: yulisun@telkomuniversity.ac.id

Received 11 March 2022, Revised 12 May 2022, Accepted 16 May 2022

Abstract — The heart is one of the most important organs in the human body. The presence of abnormalities in the heart can be fatal for a person. One way to detect heart abnormalities is an electrocardiogram (ECG) signal examination. To facilitate the detection of ECG signal abnormalities, an automatic classification method is needed. Therefore, in this study, a method for classifying ECG signals using FdispEn (Fluctuation-based dispersion Entropy) and first-order statistics is proposed. FdispEn measures the uncertainty in the signal and is expected to be able to distinguish the physiological state of the ECG signal time series. In this study FdispEn and statistical computing were used as feature extraction of the ECG signal and combined with the Support Vector Machine (SVM) for the classification simulation of Normal, AFIB (Atrial Fibrilation), and CHF (Congestive Heart Failure). The number of simulated signals are 283 Normal, 135 AFIB, and 30 CHF, respectively. In this study, the training and testing phase uses 5-cross validation. The proposed method in this study generates highest accuracy of 91.5% for normal, AFIB, and CHF classification. The system proposed in this study is expected to assist in the clinical diagnosis of abnormalities in the heart.

Keywords – heart abnormalities, electrocardiogram, Fluctuation-based dispersion Entropy, statistics

I. INTRODUCTION

Heart abnormalities or heart disease generally refers to conditions that involve narrowing or blocking of blood vessels [1]. This condition can cause a heart attack, chest pain or stroke [2]. Other heart conditions that affect the heart's valves or rhythm, are also considered forms of heart disease [3]. To prevent heart disease, an early examination is carried out by looking at the results of the recorded electrocardiogram (ECG) signal. ECG is a signal that describes the electrical activity carried out by the heart which is useful for diagnosing heart conditions and diseases [4]. Medics are generally able to identify a disease or heart attack based on the ECG recording. However, many clinicians have difficulty in identifying ECG abnormalities because some patterns of heart abnormalities have similarities [5].

Currently, digital signal processing in the medical world is important to assist doctors in diagnosing diseases. With digital signal processing, it is possible to develop an automatic classification system [6]. One of the most important stages or protocols in the classification of ECG signals is feature extraction [7]. Various studies have proposed an ECG signal feature extraction method to obtain high performance in classification. Feature extraction in the time [8], frequency [9], and time-frequency domains [10] has been reported. However, the ECG classification is still an interesting issue in the exploration of other feature extraction methods to find the best classification performance. Recently, feature extraction methods using a complexity approach to bio-signals have received a lot of attention. Bio-signal is an accumulation of complex processes at the cellular, tissue, and organ level. If there is a functional change in the body, there is a change in complexity [11]. This is what underlies that there are differences in signal complexity when there are abnormal functions in the body. One popular method for measuring the complexity of a time series signal is entropy.

Therefore, in this study proposed a mechanism for the classification of abnormalities in the heart based on ECG signals. This study uses first-order entropy and statistical approaches to calculate the information content of the ECG signal. The entropy method used in this study is fluctuation dispersion entropy, which is a new refinement in measuring signal complexity.

In this study, three types of ECG signals were classified including normal, atrial fibrillation (AFIB), and congestive heart failure (CHF). As a preliminary study, a simulation was carried out on ECG signals sourced from open datasets. The ECG signal data recorded from 45 patients 19 women (aged 23-89) and 26 men (aged 32-89). The CHF data contains Long-term ECG recordings of 15 men, aged 22 - 71 years, and 4 women aged 54 - 63. Feature extraction and classification stage simulations were performed in Matlab.

As a guide, this paper is organized as follows. Section II contains the methods used in the simulation of this study. A brief explanation of the dataset used in this study and system design are presented in Section III. Section IV contains an explanation of the results achieved along with a discussion. Section VV is the conclusion of the study, limitations of the current study, and future research work.

II. Research Method

A. *Fluctuation-based dispersion Entropy*

Fluctuation-based dispersion Entropy (FdispEn) is a new approach for estimating the dynamics of signal fluctuation variability. FdispEn is based on Entropy Shannon and fluctuation based dispersion patterns. FdispEn is a nonlinear dynamic analysis method that characterizes the complexity and irregularity of time series. In addition, FdispEn can distinguish various physiological states of biomedical time series, and is commonly used in biomedical midwives [12].

The basic calculation of FdispEn is based on dispersion entropy [13]. Meanwhile, in the FdispEn method consider the difference between adjacent dispersion pattern elements which is called fluctuation-based dispersion entropy or fluctuation-based dispersion pattern. This method can have a vector of length m – 1 where each element changes -c + 1 to c – 1. So $(2c-1)^{m-1}$ is a potential for fluctuation-based dispersion patterns.

$$FDispEn(x,m,c,d) = -\sum_{x=1}^{(2c-1)^{m-1}} p(\pi_{v_0,v_2,\dots v_{m-1}}) \ln\left(p(\pi_{v_0,v_2,\dots v_{m-1}})\right) \quad (1)$$

It can be concluded that these two approaches have different potential patterns. In this study the FdispEn value was normalized as $\frac{FdispEn}{\ln((2c-1)^{m-1})}$.

B. *First-Order Statistic*

Statistical analysis is a mathematical approach to feature extraction using statistical methods. In this study, the statistical features calculated are first order statistics including mean, skewness, kurtosis and variance.

a. *Mean* ($\bar{x}$)

According to Mishra et. al, the mean of a variable is the average of a numerical sample series. The average value can be calculated using the following equation [14].

$$Mean\ (\bar{x}) = \frac{\sum Xi}{n} \quad (2)$$

where, X_i = data-i
n = the nuber of data

b. *Skewness*

Skewness is used to measure the symmetry or flatness of a curve.

$$Sk = \frac{3\ (\bar{x} - Md)}{s} \quad (3)$$

where:
s = standar deviation
$\bar{x}$ = *mean*
Md = median

c. *Kurtosis*

Kurtosis is the peak level of a distribution which is usually taken relative to a normal distribution .

$$Kurtosis = \frac{\frac{1}{n}\sum(x-\bar{x})^4}{s^4} \quad (4)$$

where:
$\bar{x}$ = *mean*
s = standar deviation
x = data

d. *variance*

Variance is the sum of the squares of all deviations from individual values to the group mean. The root of the variance is called the standard deviation.

$$Variance = \frac{\sum(x_i - \bar{x})^2}{n} \quad (5)$$

where:
x_i = data-i
$\bar{x}$ = *mean*
n = number of data

C. Empirical Mode Decomposition (EMD)

EMD is an adaptive method for analyzing non-linear and non-stationary multi-scale real signals. The first objective of this algorithm is to decipher the univariate signal. EMD parses data into a finite number of simple orthogonal oscillatory modes called IMF (Intinsic Mode Function). The EMD algorithm decomposes the original signal into IMF and residuals [15]. An iterative process called filtering process is used [16]. In this study, EMD is used as part of preprocessing.

D. Support Vector Machine

Support Vector Machine (SVM) is a technique for making predictions, both in the case of classification and regression. The SVM concept can be explained simply as an attempt to find the best hyperplane that functions as a separator of two classes in the input space. The process of finding a support vector to obtain the best hyperplane is a learning process in SVM [17]. So that only the support vector has an effect while the other data does not affect the hyperplane.

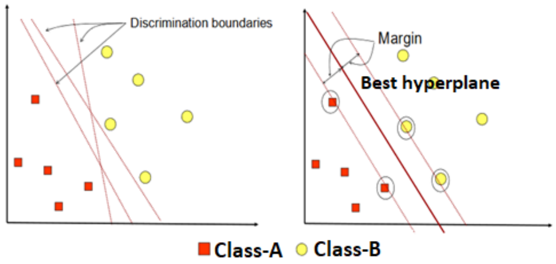

Fig.1. Finding the best Hyperplane on SVM [17]

E. Proposed Method

The following Figure 2 is a system flowchart that presents the simulation stages in this study.

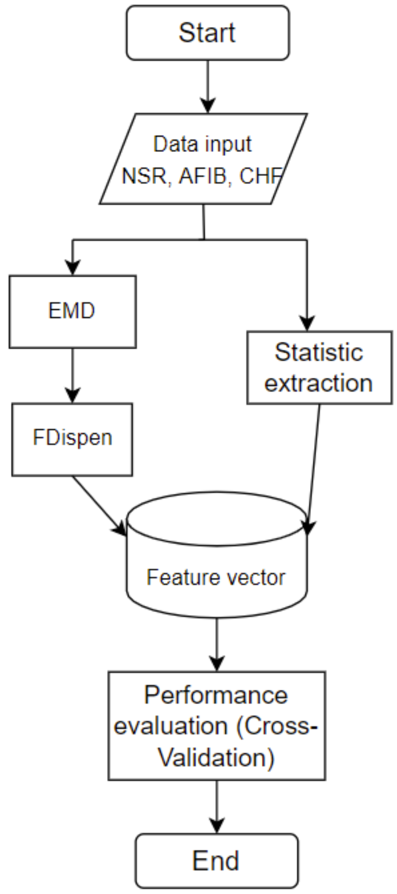

Fig.2. System flowchart

Based on Figure 2, the system designed in this study consists of three types of ECG datasets that will be processed and divided into 3 features including FdispEn features, statistical features, and combined features consisting of FdispEn and statistical features. FdispEn features through 2 processes, namely through preprocessing and not using preprocessing, statistical characteristics using predetermined statistical calculations. After the characteristics are collected, a performance evaluation is carried out, namely the classification process by conducting training and testing data. The results of the classification use 2 parameters, namely the confusion matrix and the ROC graph.

F. Dataset

The ECG signals used in this study were sourced from physionet.org. This dateset consists of NSR (Normal Synus Rhythm), AFIB (Atrial Fibrilation), and CHF (Congestive Heart Failure). The NSR data contains 283 signals, AFIB contains 135 signals and CHF contains 30 signals. ECG was recorded in 45 patients consisting of 19 women (age 23-89) and 26 men (age (32-89). The signal was recorded at a frequency of 360 Hz and gain of 200 adu/mV, duration 10 seconds. And the CHF data contained recordings. Long-term EKGs of 15 subjects, 11 males (aged 22-71) and 4 females (aged 54-63). The recordings were each sampled at 250 samples per second with 12-bit resolution over a range of approximately 10 millivolts.

G. Pre-processing the signal using EMD

Pre-processing is done by using EMD for signal decomposition. The process is by decomposing one signal into several signals. In this research, the signal is decomposed into 5 signals. Statistical feature extraction and FdispEn were then applied to the decomposed signal.

H. Feature Extraction

The next step is feature extraction, including feature extraction using the FdispEn method without preprocessing which aims to collect FdispEn feature extraction. So the collection of FdispEn feature there are 6 value including five features that have done preprocessing and one value that has not done preprocessing. Then added with statistical cues in the form of mean, skewness, kurtosis, and variance. The following Figure 3 is a flowchart of the FdispEn feature extraction process and Figure 4 is a flowchart of the statistics feature.

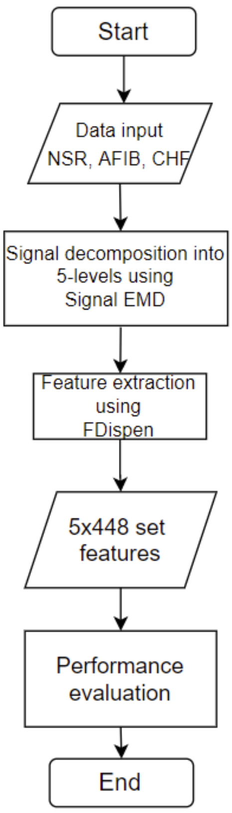

Fig.3. Flowchart of FdispEn feature extraction

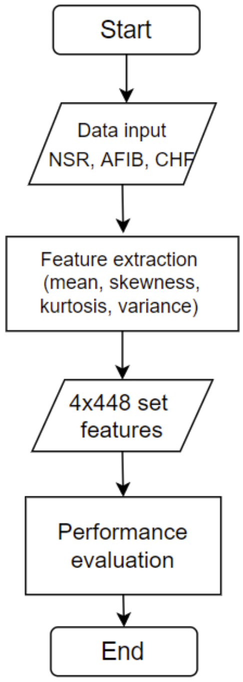

Fig.4. Flowchart of statistical feature extraction

I. Classification Stage

After the features are calculated, the next step is to classify the ECG signal using SVM. The parameters used in this classification are the accuracy of each type of SVM, ROC graph, and confusion matrix. The classification process is carried out by training data.

In this study, in conducting data training using 5 cross validation, which means doing the experiment 5 times. In this process, NSR, AFIB, and CHF ECG data are taken randomly to evaluate the performance of the system model or algorithm used. Figure 5 below is a cross validation mechanism

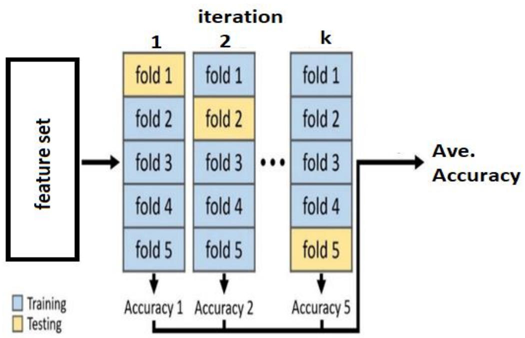

Fig.5. The ilustration of 5 *cross validation* [18]

J. Performance Evaluation

Performance evaluation aims to measure the performance of the proposed method. The performance parameter used in this study is accuracy. This parameter is calculated by the values of TP (True Positive), TN (True Negative), FP (False Positive), and FN (False Negative) [19]. Meanwhile, TP, TN, FP, and

FN are calculated based on the matrix presented in Figure 6.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} x100$$

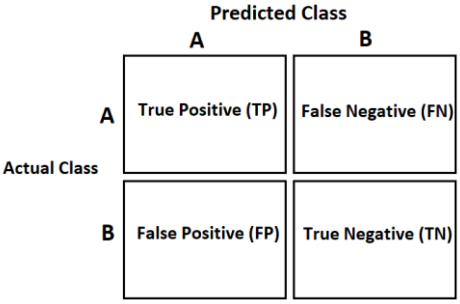

Fig.6. TP, TN, FP, and FN calculation

III. RESULTS

Table 1 shows the classification accuracy of the simulation results.

Table 1. Classification result

Scenario	Accuracy (%)			
	Gaussian	Quadratic	Cubic	Linear
1 (statistic)	90.8	90	92.4	84.6
2 (FdispEn)	72.8	71	69	69
3 (Combined)	90.2	91.5	88.6	85.9

From Table 1, it can be seen the results of several types of SVM classification. The FdispEn feature generates lower accuracy but is quite good because from several types of classifications an accuracy rate of more than 70% is obtained. The combined features generate the highest accuracy compared to other scenarios. It can be concluded that statistical features are very influential in increasing detection accuracy.

A. Confusion Matrix

The results of the confusion matrix using the FdispEn feature

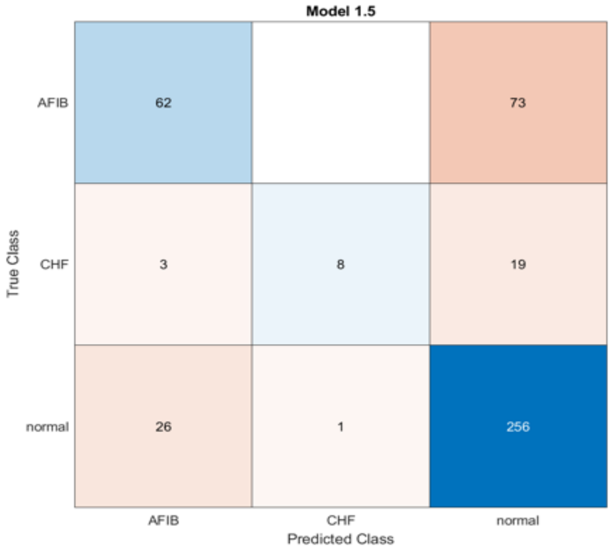

Fig.7 Confusion Matrix using FdispEn feature

From Figure 7. It is known that the system is capable of producing 8 correct CHF predictions, AFIB class 62 detected correctly and NSR class 256 data detected correctly.

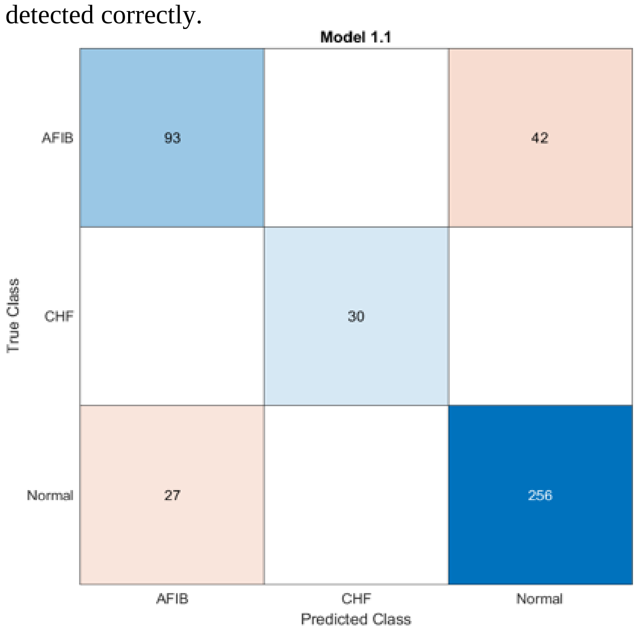

Fig.8. Confusion Matrix using statistic feature

From Figure 8, it is shown that statistical features produce predictions with higher accuracy where there are only fewer prediction errors in each class. The statistical feature method produces better performance compared to the previous scenario.

Fig.9. Confusion Matrix using FdispEn and statistic feature

From Figure 9, it is known that the combined feature consisting of FdispEn characteristics and statistical characteristics produces the highest accuracy compared to other scenarios. The results obtained are very good where the system has only a few errors in predicting each class. The level of accuracy increases after combining the two characteristics, namely the FdispEn feature and the statistical feature.

IV. Discussion

A. ROC (*Receiver Operating Characteristic*)

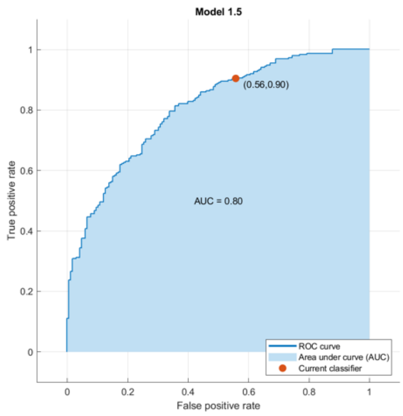

Fig.10. ROC graph of the NSR class using the FdispEn feature

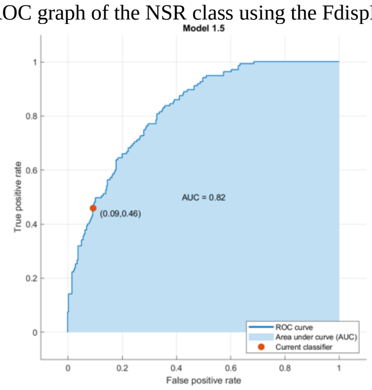

Fig.11. ROC graph of the AFIB class using the FdispEn feature

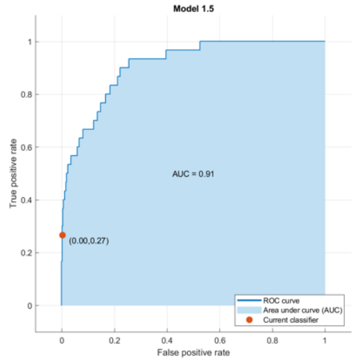

Fig.12. ROC graph of the CHF class using the FdispEn feature

From Figure 10, Figure 11, and Figure 12 is the result of the ROC graph indicated by the ROC area at coordinates (0,1). The results obtained for each classifier point class are not close to 1 where the system performance is still experiencing some errors.

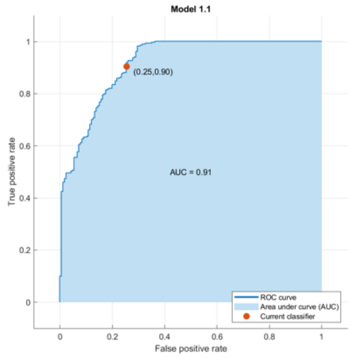

Fig.13. ROC graph on NSR class using statistical features

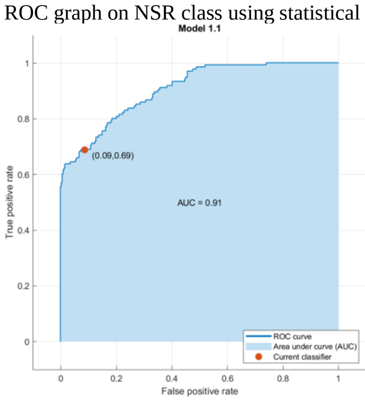

Fig.14. ROC graph on AFIB class using statistical features

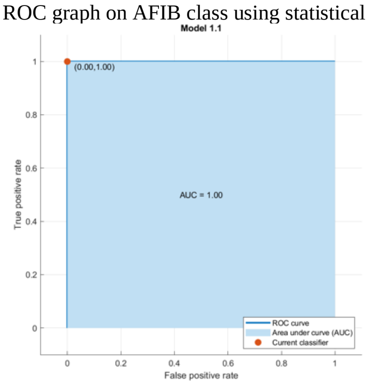

Fig.15. ROC graph on CHF class using statistical feature

From Figure 13, Figure 14, and Figure 15, the results obtained have increased in the classification where the classifier point of each class approaches number 1. For the CHF class it reaches number 1.

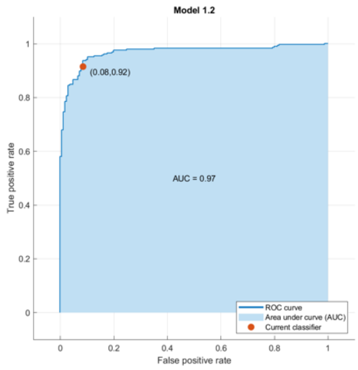

Fig.16. ROC graph on NSR class using combined features

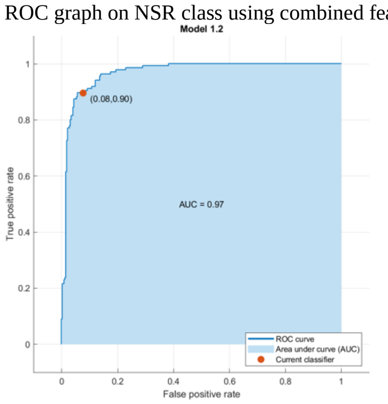

Fig.17. ROC graph on AFIB class using combined features

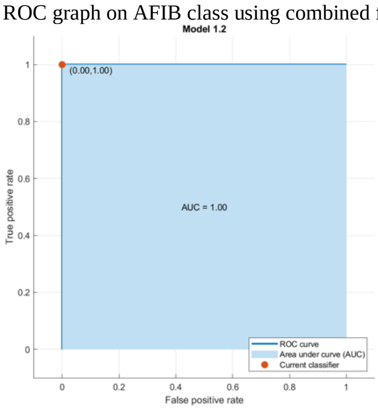

Fig.18. ROC graph on CHF class using combined features

From Figure 16, Figure 17, and Figure 18 the results of the ROC graph obtained using the Combined feature show an increase in predictions where the classifier point of each class is very close to 1, which means that the predictions of each class are almost correct and have few errors. It can be concluded from several characteristics that have been tried adding statistical features can increase the prediction accuracy of the system used.

The classification simulation both confusion matrix and ROC shows that the FdispEn feature is able to discriminate against normal, AFIB, and CHF signals. These results indicate that the ECG signal has a different complexity. The degree of complexity of the ECG signal estimated using FdispEn has the potential to be a discriminatory feature even though the classification accuracy is lower than the statistical feature. FdispEn can be an additional feature in improving classification accuracy. Other entropy methods also need to be explored to find the best performance. From this study also concluded that there is a change in signal complexity in abnormal ECG compared to normal ECG. Analysis of the significance of the difference is also needed to find the features that can produce the highest accuracy.

V. CONCLUSSION

From the results of tests using FdispEn and statistical feature extraction and classification using several types of SVM, the proposed method generate high classification accuracy. FdispEn with Gaussian SVM classification of 72.8%, Quadratic SVM of 71%, Cubic SVM of 69%, and Linear SVM by 69%. Statistical characteristics with Gaussian SVM classification of 90.8%, Quadratic SVM of 90%, Cubic SVM of 92.4%, and Linear SVM of 84.6%. The combined characteristics with the classification of SVM Gaussian is 90%, SVM Quadratic 91.5%, SVM Cubic 88.6% and Linear SVM 85.9%. In this case, the feature extraction test using the FdispEn method needs to add statistical features to increase the level of prediction accuracy. The results of testing of several types of SVM obtained SVM which has the highest accuracy, namely SVM Gaussian which is a non-linear SVM using the kernel method. This study is still limited in the number of samples. In future research, it is hoped that the proposed method can be applied to a larger database and more varied types of ECG signals

ACKNOWLEDGMENT

All authors would to thank "Telkom University" specially to the signal processing laboratory in designing, and testing implementing this study.

REFERENCES

[1] D. Dona, H. Maradona, and M. Masdewi, "Sistem Pakar Diagnosa Penyakit Jantung Dengan Metode Case Based Reasoning (CBR)," *ZONAsi: Jurnal Sistem Informasi*, vol. 3, no. 1, pp. 1–12, Apr. 2021, doi: 10.31849/zn.v3i1.6442.

[2] I. Lazulfa, U. Hasyim Asy'ari, T. Jombang, and R. Augusta, "Analisis Faktor Prediksi Diagnosis Tingkat Keparahan Penyakit Jantung (Heart Disease) Menggunakan Metode Stepwise Binary Logistic Regression," *Inovate (Junal Ilmiah Inovasi Teknologi Informasi)*, vol. 2, no. 1, Mar. 2017, Accessed: Mar. 11, 2022. [Online]. Available: http://ejournal.unhasy.ac.id/index.php/inovate/article/view/211

[3] S. Khurshid *et al.*, "Frequency of Cardiac Rhythm Abnormalities in a Half Million Adults," *Circulation: Arrhythmia and Electrophysiology*, vol. 11, no. 7, Jul. 2018, doi: 10.1161/CIRCEP.118.006273.

[4] M. Rifali and D. Irmawati, "Sistem Cerdas Deteksi Sinyal Elektrokardiogram (EKG) untuk Klasifikasi Jantung Normal dan Abnormal Menggunakan Jaringan Syaraf Tiruan (JST)," *Elinvo (Electronics, Informatics, and Vocational Education)*, vol. 4, no. 1, pp. 49–55, Nov. 2019, doi: 10.21831/elinvo.v4i1.28242.

[5] D. A. Cook, S. Y. Oh, and M. V. Pusic, "Accuracy of Physicians' Electrocardiogram Interpretations: A Systematic Review and Meta-analysis," JAMA Intern. Med., vol. 180, no.

11, pp. 1461–1471, 2020, doi: 10.1001/jamainternmed.2020.3989.
[6] A. S. Ahuja, "The impact of artificial intelligence in medicine on the future role of the physician," *PeerJ*, vol. 2019, no. 10, 2019, doi: 10.7717/peerj.7702.
[7] S. Kuila, N. Dhanda, and S. Joardar, "Feature extraction of electrocardiogram signal using machine learning classification," Int. J. Electr. Comput. Eng., vol. 10, no. 6, pp. 6598–6605, 2020, doi: 10.11591/IJECE.V10I6.PP6598-6605.
[8] A. Rizal and I. Wijayanto, "Classification of premature ventricular contraction based on ECG signal using multiorder rényi entropy," in Proceeding - 2019 International Conference of Artificial Intelligence and Information Technology, ICAIIT 2019, 2019, pp. 225–229, doi: 10.1109/ICAIIT.2019.8834590.
[9] Hindarto, I. Anshory, and A. Efiyanti, "Feature Extraction of Heart Signals using Fast Fourier Transform," in Proceeding The 1st IBSC: Towards The Extended Use Of Basic Science For Enhancing Health, Environment, Energy And Biotechnology, 2016, vol. 10, no. 2, pp. 165–167.
[10] M. Kropf, D. Hayn, and G. Schreier, "ECG classification based on time and frequency domain features using random forests," Comput. Cardiol. (2010)., vol. 44, pp. 1–4, 2017, doi: 10.22489/CinC.2017.168-168.
[11] H. Azami, A. Fernández, and J. Escudero, "Refined multiscale fuzzy entropy based on standard deviation for biomedical signal analysis," Med. Biol. Eng. Comput., vol. 55, no. 11, pp. 2037–2052, 2017, doi: 10.1007/s11517-017-1647-5.
[12] H. Azami *et al.*, "Multiscale Fluctuation-Based Dispersion Entropy and Its Applications to Neurological Diseases," *IEEE Access*, vol. 7, pp. 68718–68733, 2019, doi: 10.1109/ACCESS.2019.2918560.
[13] H. Azami and J. Escudero, "Amplitude- and Fluctuation-Based Dispersion Entropy," *Entropy*, vol. 20, no. 3, p. 210, Mar. 2018, doi: 10.3390/e20030210.
[14] P. Mishra, C. M. Pandey, U. Singh, A. Gupta, C. Sahu, and A. Keshri, "Descriptive statistics and normality tests for statistical data," Ann. Card. Anaesth., vol. 22, no. 1, pp. 67–72, 2019, doi: 10.4103/aca.ACA_157_18.
[15] 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, doi: 10.1080/03091902.2017.1394386.
[16] A. D. Candra and P. E. Suryani, "Perbandingan Metode EEMD dan EMD Untuk Mereduksi Noise Pada Sinyal Seismik," *Jurnal Ilmiah Teknosains*, vol. 4, no. 1, pp. 47–55, Jun. 2018, doi: 10.26877/jitek.v4i1.1814.
[17] P. Novenando, M. Weking, I. G. Ngurah, W. Partha, and A. I. Weking, "Application of Data Mining with Support Vector Machine (SVM) in Selling Prediction Trend of Spiritual Goods (Case Study : PT . X Bali)," Int. J. Eng. Emerg. Technol., vol. 4, no. 1, pp. 20–24, 2019.
[18] A. Rizal and S. Hadiyoso, "Sample entropy on multidistance signal level difference for epileptic EEG classification," *Sci. World J.*, vol. 2018, pp. 1–7, 2018, doi: 10.1155/2018/8463256.
[19] I. Wijayanto, A. Humairani, A. Rizal, and S. Hadiyoso, "Klasifikasi Sinyal EKG menggunakan Ciri Statistik dan Parameter Hjorth dengan SVM dan k-NN," ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron., vol. 10, no. 1, pp. 132–145, 2022.