Proc. of 2019 6th Int. Conf. on Information Tech., Computer, and Electrical Engineering (ICITACEE), Sep 26-27, 2019, Semarang, Indonesia

MFCC Feature Extraction and KNN Classification in ECG Signals

Siti Agrippina Alodia Yusuf
Dept. of Electrical Engineering and Information Technology, Faculty of Engineering, Gadjah Mada University
Yogyakarta, Indonesia
sit.agrippina.a@mail.ugm.ac.id

Risanuri Hidayat
Dept. of Electrical Engineering and Information Technology, Faculty of Engineering, Gadjah Mada University
Yogyakarta, Indonesia
risanuri@jteti.gadjahmada.edu

***Abstract*—Feature extraction of electrocardiogram (ECG) signal is one of the essential steps to diagnose various cardiovascular disease (CVD). The signal is generated by the hearts electrical activity and able to reveal the abnormal activity of the heart. An accurate feature extraction method is important to produce better identification of ECG signal. ECG feature extraction using Mel Frequency Cepstrum Coefficient (MFCC), Discrete Wavelet transformation and KNN using euclidean distance as the classifier is proposed in this study. The model and testing of the proposed system were performed on the two types of data, normal and myocardial infarction (MI) labeled as abnormal, obtained from PTB-DB database. Total data used were 100 data, with 50 data for each condition. K-fold cross validation also applied to achieve a generalized result. According to the experimental, 13 features that obtained from MFCC shows good result. The accuracy, sensitivity and specificity were achieved 84%, 85% and 84% respectively.**

***Keywords*—** ECG signal, feature extraction, MFCC, Wavelet, k-fold cross validation, KNN

I. Introduction

Cardiovascular disease is one of the leading cause of non-communicable diseases (NCDs) deaths according to the data of Global Health Observatory (GHO). Of 56.9 million global deaths in 2016, 17.9 million deaths or 44% are caused by cardiovascular [1]. Cardiovascular is a general term that includes several specific heart conditions, such as Myocardial Ischemia, Dysrhythmia, Cardiomyopathy, Congenital Heart Disease, Valvular, Peripheral Arterial Disease. Electrocardiogram (ECG) is a noninvasive technique, it represents the contractile of heart activity and can be recorded using surface electrodes on the limbs or the chest of the patient [2]. An ECG waveform holds important information of diagnostic segments, in one cardiac cycle of normal ECG signal consist of P, QRS, and T waves. The blood pressure from the atrium to the ventricle causes atrial depolarization, at that time P wave occurs. The Q, R and S wave that forms QRS complex occurs when ventricular depolarization. The T wave occurs when the repolarization period of the ventricles as a preparation for the next heartbeat [3].

The ECG signal has own characteristic because of the electrical activity of the heart. If the heart rhythm or damage is abnormal, the heart muscle can change the electrical activity of the heart, that is how the ECG signal gets changed [4]. Manual interpretation of ECG signal requires experience and it is hard to determine the small changes that occur in ECG signal waveform. Furthermore, due to the different waveform, variation, the small amplitude (mV) and duration (sec) make the automatic interpretation of the ECG signal is challenging [5][6]. Pattern recognition is an application to interpret the ECG signal. The purpose of pattern recognition is to classify automatically a system into different classes. For that reason, in the last two decades, a researcher in the biomedical engineering community pays special attention to classifying ECG signal [7].

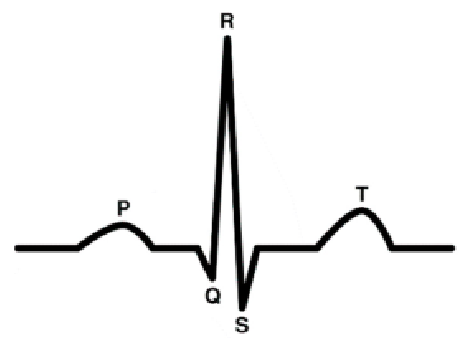

Fig. 1. ECG signal sample showing P, QRS and T waves [2]

Typically, there are four phases of signal processing for classification: preprocessing, segmentation, feature extraction and classification [8]. The preprocessing aim to remove the unwanted component, enhance and normalize to achieve a reliable ECG signal [5][8]. Then segmentation is the next step after preprocessing, the aim of the segmentation is to divides signal into smaller segments to achieve better electrical activity of the heart [8]. Feature extraction is one of the most important steps for the diagnostic system, the features of ECG signal can be extracted in various domain, such as extracting in time and/or frequency domain, morphological and/or statistical domain [5][8][9]. Variety of algorithm also able to extract feature from ECG signal, such as Pan-Tompkins [5], Hilbert transform [10], and Wavelet packet decomposition [11]. As for classification, theoretically, all classifier can be used for ECG classification [8], such as Neuro fuzzy [9], Support Vector Machine [5], and K-Nearest Neighbour [11].

This paper presents the use of Mel frequency cepstral coefficient (MFCC), discrete wavelet transformation (DWT) as the feature extractor and KNN as the classifier to identify and distinguish heart condition, which is normal and abnormal.

The next section of this paper is organized as follows: the methods used in the study are summarized in Section 2 and the results and discussion are presented in Section 3. Lastly, the conclusion is presented in Section 4

II. Materials and Methods

Methodologies that consist of three steps were proposed in this study, such as preprocessing, feature extraction and the classification

978-1-7281-2689-0/19/$31.00 ©2019 IEEE

of ECG signals.The following passages are the explanation of the methodology

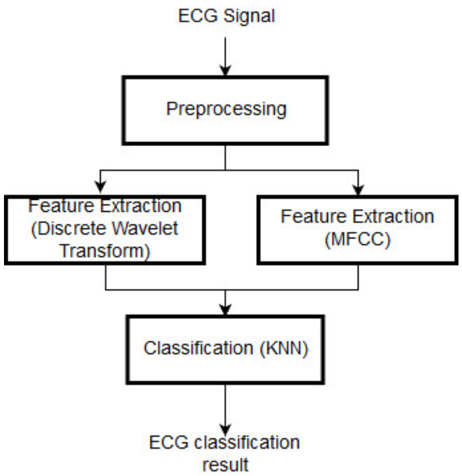

Fig. 2. Identification of ECG signal

A. Dataset

The ECG signals were obtained from the Physikalisch Technische Bundesanstalt Diagnostic ECG Database (PTBDB) available at Physionet that recorded by the Department of Cardiology at Benjamin Franklin University in Berlin, Germany, with the help of healthy volunteers and patients with different cardiac diseases. The database contains 549 records from 290 subjects (aged 17 to 87, mean 57.2; 209 men, mean age 55.5, and 81 women, mean age 61.6). Each subject was represented by one to five records. The classification of the recording contains healthy signal and different cardiac diseases such as myocardial infarction, cardiomyopathy/heart failure, bundle branch block, arrhythmia, myocardial hypertrophy, etc. Each record includes 15 simultaneously measured signals: the conventional 12 leads (i, ii, iii, avr, avl, avf, v1, v2, v3, v4, v5, v6) together with the 3 Frank lead ECGs (vx, vy, vz). This paper performed analyzed ECG signal based on the following criteria: 50 Myocardial Infarction (MI) recordings were labeled as abnormal and 50 healthy control recordings were labeled as normal using lead ii.

B. Preprocessing

ECG signal contains different types of noises, such as frequency interference, baseline drift, electrode contact noise, polarization noise, muscle noise, the internal amplifier noise, and motor artifacts [12] that can lead to corruption of ECG signal. To overcome that condition, it is necessary to eliminate noise from the signal using a suitable filter. The median filter was applied to soften the signal without losing the important information, this act can fix the baseline wander problem.

C. Mel Frequency Cepstral Coefficient

One of the powerful algorithms that is widely used in the signal processing domain especially in the field of speech recognition is the mel-frequency cepstral coefficient (MFCC). It is a linear representation of the cosine transform of a short duration of the logarithmic power spectrum of the speech signal on a nonlinear scale Mel frequency [13]. The main idea in the MFCC is based on Mels criteria and the nature of human hearing perceptions and speech intelligibility [14]. MFCC consists of framing, windowing, FFT, Mel Filter Bank, and DCT. Figure 3 illustrated the MFCC feature extraction process.

The first step is framing, in this step, the signal is trimmed into an equal section called frames, the usual duration is 10-30 ms. The

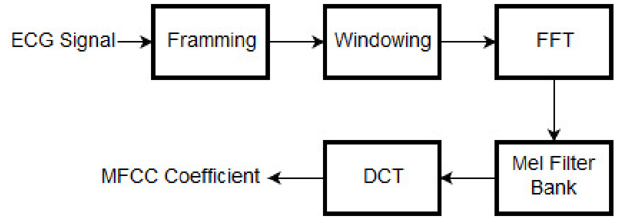

Fig. 3. MFCC Feature Extraction

framing process can cause the discontinuity of the signal, to maintain its continuity, windowing then applied [15]. Hamming window was used in this paper. The following equation represents the Hamming, While equation (2) is the output of each frame after the filtering process.

$$W[n] = 0.54 - 0.4cos[\frac{2\pi n}{N-1}] \tag{1}$$

$$Y[n] = X[n] \times W[n] \tag{2}$$

Where *N* is the number of samples per frame, *Y[n]* is the output signal and *W[n]* is the n^{th} coefficient of the Hamming window [15]. Then, each frame of the sampled signal converted from the time domain to the frequency domain using FFT [13]. The next step is Mel filterbank; it is the bandpass filter that overlaps with each other. Based on the Mel scale, it is linear under the frequency 1kHz and logarithmic over 1kHz [15]. The following equation is the mathematical form of Mel scalling (3).

$$mel = 2595log_{10}(1 + \frac{f}{100}) \tag{3}$$

Where *mel* is the output of the filterbank and f is the input of the filterbank. 2595 and 700 are the fixed values that already been used in many research [15]. The following figure ilustrated the filterbank.

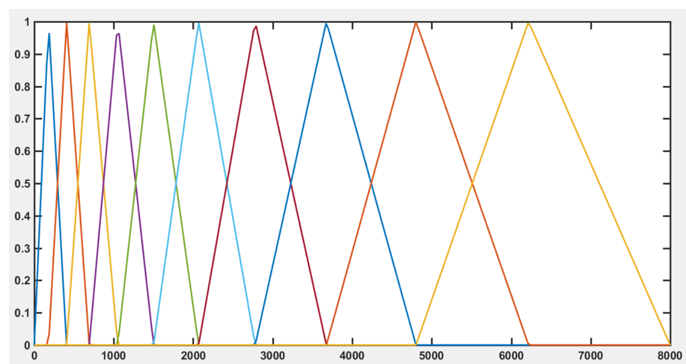

Fig. 4. Mel Spaced Filterbank

The last step is calculating the N characteristics with Discrete Cosine Transform (DCT) to generate the MFCC. The following equation (4) is the mathematical form of DCT.

$$\sum_{k=1}^{N} log(Y(i)) \times cos[mx(k - 0.5)x\pi \div N] \tag{4}$$

D. Discrete Wavelet Transform

Wavelet transform is a mathematical tool to decompose a signal into a set of orthogonal waveform localized both in time and frequency domains. The output of the decomposition is coefficient, which are functions of the scale (of the wavelet function) and position (shift across the signal) [12]. In practical usage of signal processing, a signal filtered into low pass filter (LPF) and high pass filter (HPF). The output coefficient of LPF is called approximation and the output of HPF is called detail. The approximation signal then can be sent again to pass LPF and HPF for the next level decomposition, that way signal has various components at a different level. The following figure is the structure of the DWT filterbank:

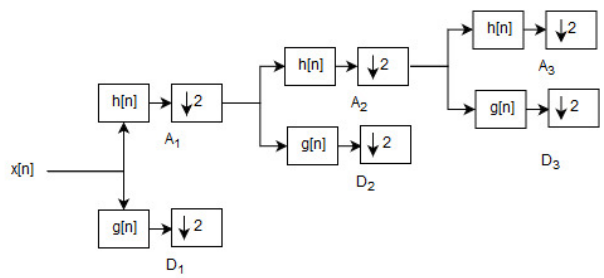

Fig. 5. 3 level DWT Decomposition

LPF removes the frequency component that is above half of the highest frequency but leaves the scale unchanged, and the resolution become half after the operation [5]. The mathematical form of the above process can be defined as:

$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[2n-k] \tag{5}$$

The output of DWT is an approximation and several detail components that provide information about the signal by decomposing the signal into subbands at different frequencies along with different resolutions. The following mathematical form represents one level of decomposition:

$$y_{low}[k] = \sum_{n} x[n]h[2k-n] \rightarrow y_{low} = (x*h)\downarrow 2 \tag{6}$$

$$y_{high}[k] = \sum_{n} x[n]g[2k-k] \rightarrow y_{high} = (x*g)\downarrow 2 \tag{7}$$

Where $y_{high}[k]$ and $y_{low}[k]$ yielded by high-pass low-pass filter respectively after subsampling by 2 [5].

Deciding the type of mother wavelet and the level of decomposition is crucial for signal analysis. Haar and Symlet order 7 function were chosen to extract the features of the ECG signal. Haar wa schosen because of the simplicity while Symlet order 7 was chosen because of its similarity to the ECG signal. Thus, this paper applied Haar and Symlet 7 at level 6 decomposition to extract the information of the subbands.

E. K-Nearest Neighbour

One of the statistical classification algorithm used for classifying object based on the closest training example is KNN. The function is approximated locally and all the computations are postponed until classification, hence it is called lazy learning algorithm. There are two phases in the KNN algorithm, training, and classification phase. In the training phase, the vector is the training examples in a multidimensional featured space. This phase stored the feature vectors and class labels of training samples. While in the classification phase, user define the constant K, a query or test point is classified by assigning a label, which is the most recurrent among the K training samples nearest to that query point [16]. KNN requires only a few parameters to tune to achieve high classification accuracy, which is K and the distance metric. Usually, K is an odd number so it can avoid the tied votes [16]. The distance method used in this study was euclidean distance. The following equation (5) is the mathematical form of euclidean distance.

$$d_{st} = \sqrt{\sum_{j=1}^{n}(x_{sj}-y_{tj})^2} \tag{8}$$

Where *x* and *y* are the two data that the distance must be search and *n* is the number of each data dimension.

F. K-Fold Cross Validation

K-fold cross-validation is applied so the variation of recognition result low. This technique allowed data divided into K number of the dataset. The 1st dataset used as testing data while the 2nd, 3rd...K dataset used as the training data. Then, the process repeated where the 2nd dataset used as the testing data, while the 1st, 3rd...K used as the training data. This process repeated until K times. Total accuracy then divided by K [11]. Since the total data in this research is 100, thus the number of K used was 5.

G. Performance Metric

Evaluating the system performance is necessary, that is why to evaluate it, a well-known performance metrics such as accuracy (Acc), sensitivity (Se), specificity (Sp) were used. Those metrics obtained from the confusion matrix that forms four indices which are true positive (TP), true negative (TN), false positive (FP) and false-negative (FN). TP indicates the number of abnormal samples correctly detected, TN indicates the number of normal samples correctly detected, FP indicates the number of abnormal samples incorrectly detected and FN indicates the number of normal samples detected incorrectly. Table I expressed the mathematical forms of those features.

TABLE I
PERFORMANCE METRICS

Formula	Description
$Acc = \frac{TP+TN}{TP+FP+TN+FN}$	Acc points the overall performance of the model
$Se = \frac{TP}{TP+FN}$	Se refers to the model performance to correctly detect abnormal samples
$Sp = \frac{TN}{TN+FP}$	Sp refers to the model performance to correctly detect normal samples

III. Result and Discussion

The proposed model was conducted on an open database which consists of 50 each of ECG signals, normal and abnormal, thus total data set used was 100. To achieve a generalized result, K-fold cross-validation was applied to the proposed model with the number of K=5. After the preprocessing step using the median filter, the length of the signal was adjusted to a certain number to follow the power of two. Then, each signal was extracted using MFCC. 27 triangular filters were applied to the signals, then 13 coefficients of the 27 were kept as the features. Aside from MFCC, discrete wavelet transformation also applied to extract the features of the signal. The result will be compared to the MFCC. Two mother wavelet were chosen, Haar and Symlet order 7 (Sym7). ECG signal was decomposed into 6 levels. By applying 6 level decomposition, there were six details and an approximation coefficient that obtained, which were D1, D2, D3, D4, D5, D6, and A6. 3 subbands were chosen for further process, which were D4, D5, and D6. The chosen details were based on the consideration that the biggest information on ECG signal was stored at the frequencies below 40 Hz. 5 statistical features were extracted in every chosen subbands. They were mean, variance, standard deviation, kurtosis and skewness. 15 features were obtained from each of mother wavelet.Those features were fed to the KNN classifier. To find the highest accuracy, determining the value of K or the neighborhood in KNN is needed, since it influences the accuracy of the system. KNN classifier was combined with the 5-fold cross-validation and tried a various odd value of K in KNN classifier to avoid ties. Table II shows the obtained results.

TABLE II
THE EFFECT OF K VALUES ON AVERAGE ACCURACY ON FEATURE EXTRACTORS AND KNN CLASSIFIER

	MFCC			Symlet 7			Haar		
K Number	3	5	7	3	5	7	3	5	7
Accuracy (%)	84	82	78	71	67	65	68	65	64
Sensitivity (%)	85	84	83	81	71	72	70	71	69
Spesificity (%)	84	83	73	69	63	64	67	65	62

Table II shows that the smallest value of K gives the highest accuracy and the highest value gives the smallest accuracy. This is because a small value for K is more flexible and have low bias but high variance. On the other hand, the larger value of K gives a lower variance but increased bias. The accuracy obtained using MFCC outperform other feature extractors, with an accuracy of 84%. Furthermore, the following tables show the accuracy, sensitivity and specificity of the classifier on each dataset using each feature extractor.

TABLE III
ACCURACY, SENSITIVITY & SPECIFICITY EACH DATASET USING MFCC FEATURE EXTRACTOR

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dataset 1	95	88	100
Dataset 2	85	85	86
Dataset 3	80	75	88
Dataset 4	80	78	82
Dataset 5	80	100	64
Average (%)	84	85	84

TABLE IV
ACCURACY, SENSITIVITY & SPECIFICITY EACH DATASET USING SYMLET 7 FEATURE EXTRACTOR

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dataset 1	85	100	79
Dataset 2	55	100	44
Dataset 3	75	90	60
Dataset 4	65	50	80
Dataset 5	75	67	82
Average (%)	71	81	69

TABLE V
ACCURACY, SENSITIVITY & SPECIFICITY EACH DATASET USING HAAR FEATURE EXTRACTOR

	Accuracy (%)	Sensitivity (%)	Specificity (%)
Dataset 1	60	75	50
Dataset 2	70	63	75
Dataset 3	70	55	89
Dataset 4	80	92	63
Dataset 5	60	67	59
Average (%)	68	70	67

On Table III, the highest accuracy and specificity obtained by dataset 1 and dataset 5 obtained the highest sensitivity, while on

TABLE VI
AVERAGE ACCURACY, SENSITIVITY & SPESIFICITY

	Accuracy	**Sensitivity**	**Specificity**
Haar	68	70	67
Symlet 7	71	81	69
MFCC	**84**	**85**	**84**

Table IV the highest accuracy and sensitivity obtained by dataset 1, dataset 2 also obtained the highest sensitivity. On Table V, the highest accuracy and sensitivity obtained by dataset 5 and the highest specificity obtained by dataset 3.Table VI summarize the accuracy, sensitivity and specificity of each feature extractor.

MFCC obtained the highest average accuracy, sensitivity, and specificity compared with Symlet 7 and Haar. The highest accuracy obtained by MFCC with 13 features, followed by Symlet 7 and Haar with 15 features of each of them. Accuracy, sensitivity and specificity obtained by MFCC are 84%, 85% and 84%, respectively. Accuracy, sensitivity and specificity obtained by Symlet 7 are 71%, 81% and 69%, respectively. While, Accuracy, sensitivity and specificity obtained by Haar are 68%, 70% and 67%, respectively.

IV. CONCLUSION

This paper has presented features extraction method which is MFCC and using KNN as the classifier to identify and distinguish the ECG signal as normal and abnormal. 13 coefficients were used as the features of the signal, with the value of K = 3 on the classification process. 5-fold cross-validation was applied to achieve generalize results. This study also presented the classification result using discrete wavelet transform with mother wavelet Haar and Symlet 7. The highest results were obtained by MFCC with accuracy, sensitivity and spesificity respectively 84%, 85% and 84%. In the future, the improvement can be made to the proposed system, like combining MFCC with the delta and delta-deltas.

REFERENCES

[1] G. H. O. (GHO), "Ncd mortality and morbidity." https://www.who.int/gho/ncd/mortality_morbidity/en/, 2016. [Online; accessed 18-January-2019].

[2] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," *Procedia Computer Science*, vol. 120, pp. 268 – 275, 2017. 9th International Conference on Theory and Application of Soft Computing, Computing with Words and Perception, ICSCCW 2017, 22-23 August 2017, Budapest, Hungary.

[3] A. Diker, D. Avci, E. Avci, and Mehmet, "A new technique for ecg signal classification genetic algorithm wavelet kernel extreme learning machine," *Optik*, vol. 180, pp. 46 – 55, 2019.

[4] S. H. Jambukia, V. K. Dabhi, and H. B. Prajapati, "Classification of ecg signals using machine learning techniques: A survey," in *2015 International Conference on Advances in Computer Engineering and Applications*, pp. 714–721, March 2015.

[5] A. Diker, Z. Cmert, E. Avci, and S. Velappan, "Intelligent system based on genetic algorithm and support vector machine for detection of myocardial infarction from ecg signals," in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, pp. 1–4, May 2018.

[6] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals," *Information Sciences*, vol. 415-416, pp. 190 – 198, 2017.

[7] V. Arjunan, "Ecg signal classification based on statistical features with svm classification," *International Journal of Advances in Signal and Image Science*, vol. 2, 2016.

[8] T. Li and M. Zhou, "Ecg classification using wavelet packet entropy and random forests," *Entropy*, vol. 18, 2016.

[9] E. M Tamil, N. Hafeezah Kamarudin, R. Salleh, M. Yamani, M. Idris, N. N.M, and A. Mohd, "Heartbeat electrocardiogram (ecg) signal feature extraction using discrete wavelet transforms (dwt)," 12 2018.

[10] R. Rodrguez, A. Mexicano, J. Bila, S. Cervantes, and R. Ponce, “Feature extraction of electrocardiogram signals by applying adaptive threshold and principal component analysis,” *Journal of Applied Research and Technology*, vol. 13, no. 2, pp. 261 – 269, 2015.

[11] A. Rizal, “Perbandingan skema dekomposisi paket wavelet untuk pengenalan sinyal ekg,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 4, pp. 80–86, 05 2015.

[12] M. P Nageswari, S. Banu, K. Kumar, S. Maruthi Sujith, and U. Student, “Feature extraction of ecg using daubechies wavelet and classification based on fuzzy c-means clustering technique,” 05 2013.

[13] M. Boussaa, I. Atouf, M. Atibi, and A. Bennis, “Ecg signals classification using mfcc coefficients and ann classifier,” in *2016 International Conference on Electrical and Information Technologies (ICEIT)*, pp. 480–484, May 2016.

[14] B. Farzam and J. Shirazi, “The diagnosis of heart diseases based on pcg signals using mfcc coefficients and svm classifier,” *IJISET - International Journal of Innovative Science, Engineering ‘—&’ Technology*, vol. 1, pp. 454–459, December 2014.

[15] R. Hidayat, A. Bejo, S. Sumaryono, and A. Winursito, “Denoising speech for mfcc feature extraction using wavelet transformation in speech recognition system,” in *2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE)*, pp. 280–284, July 2018.

[16] I. Saini, D. Singh, and A. Khosla”, “Qrs detection using k-nearest neighbor algorithm (knn) and evaluation on standard ecg databases,” *Journal of Advanced Research*, vol. 4, no. 4, pp. 331 – 344, 2013.