# Feature Reduction Method Using Self Organizing Maps

Yakup Kutlu[1] and Damla Kuntalp[1]

[1]Dokuz Eylül University , İzmir, Turkey
yakup.kutlu@deu.edu.tr, mehmet.kuntalp@deu.edu.tr, damla.kuntalp@deu.edu.tr

### Abstract

**In this work, five main groups of arrhythmias in electrocardiograph (ECG) signals are tried to be classified using the features obtained from the output of a Self Organizing Map (SOM) network. The raw ECG signal consists of 81 sample points (60 point before and 20 point after the R peak point of the ECG). Consecutive sample values of a moving window (20 points of width) are used as the input vector of the SOM network. The output of the SOM network is used as the input vector to a classifier. K-nearest neighbor (k-NN) algorithm is chosen as the classifier. The performance of the classifier is evaluated by the average values of sensitivity, specificity, selectivity and overall accuracy. As a result, 96%, 91%, 99%, and 97% sensitivity, selectivity, specificity, and overall accuracy values are obtained.**

Key Words: ECG, Arrhythmia, Self Organizing Maps, K-NN, feature reduction

## 1. Introduction

The detection of the arrhythmias from an electrocardiographic (ECG) record has been a very important subject. The accurate recognition and classification of the various types of arrhythmias is essential for the correct treatment of the patient. In literature, many studies have been applied for arrhythmia detection. Physionet MIT-BIH arrhythmia database is used in most of them. Various algorithms for the automatic detection of ECG beats have been developed by different investigators for this purpose. These researchers used different features and classification methods. In these studies, raw ECG with different window sizes between 144 ms and 400 ms according to R points has been used[1-12]. Different models are used as classifiers such as an MLP [2, 4, 7–10, 12], Fuzzy Logic [2, 7], Support Vector Machine (SVM) [2, 4, 6], linear discriminant function[5], Self Organizing Maps (SOM) [9, 11], etc.

In this study, the window length of 222 ms (81 consecutive points) was constituted from the raw ECG signal according to the location of the R point (52ms right and 166ms left side of R). These raw ECG data are windowed by a 20 point length moving window which is lagged in time 5 point at each step. 20 point length data vector which was created by this method was used as the input of the SOM network. Output of the SOM network was used as the input features for the k-NN classifier

In the following section, the ECG data acquisition, preprocessing, moving window method, SOM structure, k-NN classifier and performance measures are presented. The result and discussion are given in section three.

## 2. Materials and Methods

In this section the ECG data acquisition, pre-processing and feature extraction methods are described in detail. The general block diagram of the constructed system is shown in Fig. 1.

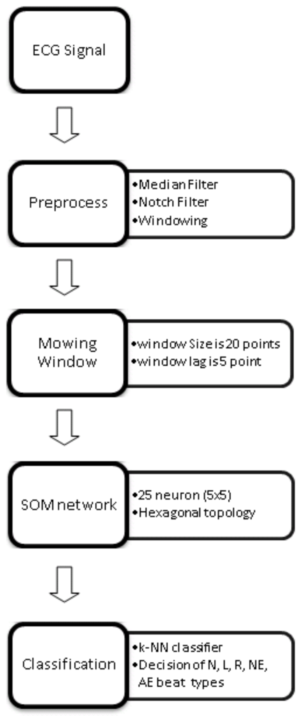


**Fig. 1.** The general block diagram of the study

### 2.1. Data acquisition

ECG signals were obtained from MIT-BIH Arrhythmia Database [13] that was created by the Beth Israel Hospital Arrhythmia Laboratory. The subjects were 25 men, aged 32 to 89 years, and 22 women, aged 23 to 89 years. Each of the records is 30 minutes long and the sampling frequency of the records is 360 Hz.

The ECG signal was processed to reduce the problems of baseline wander and power line interference. All ECG signals were first filtered with two median filters to remove the baseline wander. After the signal is processed with two median filters which have 200ms and 600ms widths respectively, the obtained signal was the baseline of the ECG signal. The result was subtracted from the original signal to remove the baseline wander effect of ECG signal. Then power-line frequency was removed from the median filtered ECG with a notch filter.

The annotation label was used to identify R points. After determining QRS peak in the database, feature sets were constituted by windowing the raw ECG around each R point. The window length was taken as 81 raw ECG points (60 previous points and 20 next points) as shown Fig.2.

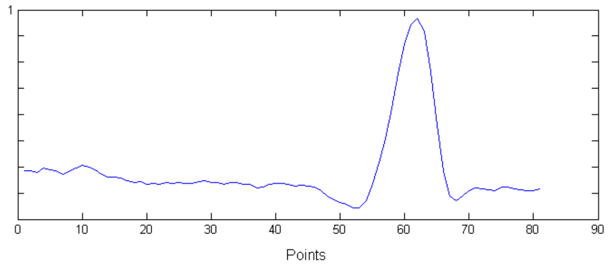


**Fig. 2.** Windowed normal ECG signal

After 81 point length beats were created from raw ECG signal, the input vectors of the SOM network were prepared by a moving window in time. The moving window size is 20 points and this window is lagged 5 points in time as shown in Fig. 3. For each step, the windowed points were taken as the input of the SOM.

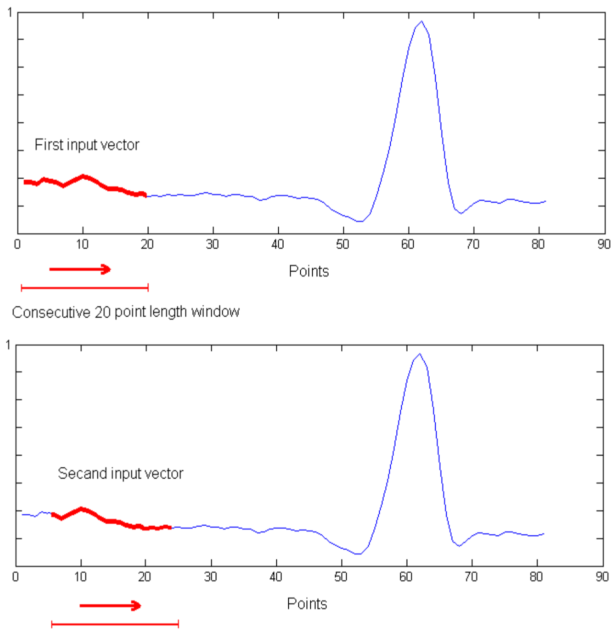


**Fig. 3.** Windowed normal ECG signal

The vectors composed of the moving window outputs were entered into the SOM. The activated neuron is called the Best Match Unit (BMU). The labels of the successively activated neurons were used to construct a specific chain code for the windowed raw ECG signal. These codes were used as the inputs of the k-NN classifier.

Total of 1895 heartbeats were used (1173 beats for training and 722 beats for testing) as shown in Table 1. The system was used to classify Normal Beat (N), Left Bundle Branch Block Beat (L), Right Bundle Branch Block Beat (R), Nodal Escape Beat (NE), and Atrial Escape Beat (AE).

**Table 1.** Number of ECG records of each arrhythmia type in the database

| Class | Train | Test |
|---|---|---|
| Normal Beat (N) | 350 | 200 |
| Left Bundle Branch Block Beat (L) | 350 | 200 |
| Right Bundle Branch Block Beat(R) | 350 | 200 |
| Nodal Escape Beat (NE) | 115 | 114 |
| Atrial Escape Beat (AE) | 8 | 8 |

## 2.2. Self-Organizing Map

The SOM is an unsupervised type neural network architecture used to visualize and interpret high-dimensional data sets [14]. The map usually consists of a two-dimensional regular (rectangular or hexagonal) grid of nodes called neurons. Each sample of high dimensional input data is associated with a unit which is the winner. Not only the winning neuron but also its neighbors on the lattice are allowed to learn and adapt their weights towards the input. This way, the representations will become ordered on the map.

An N-dimensional input is presented to each neurons of a SOM network as shown in Fig. 4. Then the winner unit (indicated by the index *c*), i.e. best match unit, is identified by the condition shown below for each sample,

$$\| x(t) - w_c(t) \| = \min_{c \in i} \| x_i(t) - w_i(t) \| \qquad (1)$$

where $x_i$ is the input vector with N dimension, $w_i$ is the *i*th weight, and *c* indicates the winning neuron.

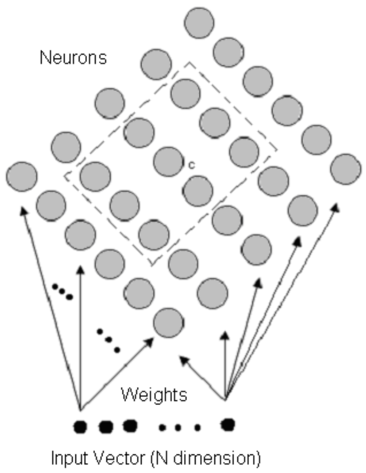


**Fig. 4.** Self-Organizing Map Structure

The update of the weights in the SOM network is limited by a neighborhood function ($\Omega_c(i)$). The neighborhood function plays a main role in the SOM algorithm regardless of the type of the learning algorithm. Three frequently used neighborhood functions are Gaussian, rectangular and cut Gaussian. The weight of the winning unit and its neighbors are updated by the formula

$$\Delta w_i = \eta(x - w_i)\Omega_c(i) \quad i \in NB_c \qquad (2)$$

where $\eta$ is the learning rate in the interval $0 < \eta < 1$, $\Omega_c(i)$ is the neighborhood function and $NBc$ indicates the neighbor neurons centered around node $c$, i.e. the winning neuron.

### 2.3. k- Nearest Neighborhood

k-NN algorithm is one of the most classical methods in pattern recognition because of its effective non-parametric nature [15]. The k-NN algorithm is a method for classifying objects based on closest training examples in the feature space. The k-NN algorithm is among the simplest of all machine learning algorithms. The algorithm is independent from statistical distribution of training examples. Classifying process of objects is realized according to the closest neighbourhood of training examples. Several distance measures are used in this algorithm. Euclidean distance is commonly used as the distance measure. An object is classified by a majority vote of its neighbours, with the object being assigned to the class most common amongst its k nearest neighbours. k is usually chosen small and odd. If $k=1$, then the object is simply assigned to the class of its nearest neighbor.

### 2.4. Performance Measures

The performance of the classification system was measured based on three standard statistical measures for each class: sensitivity (SEN), specificity (SPE), and selectivity (SEL) which are calculated from multi-class classification table [16].

Multi-class classification table for Left Bundle Branch Block Beat (L) is shown in Table 2 as an example.

**Table 2.** Multi-class classification table for beat L

| | | Classifier results | | | | |
|---|---|---|---|---|---|---|
| | | N | L | R | NE | AE |
| Reference label | N | TN | FP | TN | TN | TN |
| | L | FN | TP | FN | FN | FN |
| | R | TN | FP | TN | TN | TN |
| | NE | TN | FP | TN | TN | TN |
| | AE | TN | FP | TN | TN | TN |

where

TP - True Positives: Number of L type heartbeats correctly classified by the system

TN - True Negatives: Number of other heartbeats correctly classified by the system

FP - False Positives: Number of heartbeats incorrectly classified by the system as L type

FN - False Negatives: Number of L type heartbeats incorrectly classified by the system

The measures are defined for any specific heartbeat as:

$$\text{Sensitivity} \quad = \frac{\text{TP}}{\text{TP} + \text{FN}} x100\% \qquad (3)$$

$$\text{Specificity} \quad = \frac{\text{TN}}{\text{TN} + \text{FP}} x100\% \qquad (4)$$

$$\text{Selectivity} \quad = \frac{\text{TP}}{\text{TP} + \text{FP}} x100\% \qquad (5)$$

$$\text{Overall Accuracy} \quad = \frac{\text{TP}_\text{N} + \text{TP}_\text{L} + \text{TP}_\text{R} + \text{TP}_\text{NE} + \text{TP}_\text{AE}}{\text{All Beats}} x100\% \qquad (6)$$

## 3. Results

The classifiers implemented in this study were developed by using Matlab 7 on a Pentium-4 2.4 GHz PC. Classification procedure was performed off-line on data stored on the hard disk. Total of 1895 windowed Raw ECG vectors were used. 1173 vectors were used for training, and 722 vectors were used for testing. The system was used to classify five different arrhythmia types: Normal Beat (N), Left Bundle Branch Block Beat (L), Right Bundle Branch Block Beat(R), Nodal Escape Beat (NE), and Atrial Escape Beat (AE).

SOM network was prepared in a hexagonal topology with 25 neurons (5x5). The output neurons were labeled from 1 to 25. The labels of the successively activated neurons constituted a specific chain code for the windowed raw ECG signal. Each chain code constructed in this manner is a 13x1 vector. These codes were used as the inputs to the k-NN classifier. The average performances of 96.47%, 90.93%, and 99.15%, sensitivity, selectivity, and specificity, respectively, were obtained. Moreover, the overall accuracy of the classifier is 96.7%. The results are shown in Table 3.

**Table 3.** The performance of k-NN based classifiers

| Class Label | SEN | SEL | SPE |
|---|---|---|---|
| **N** | 98.50 | 93.81 | 97.51 |
| **L** | 98.00 | 98.99 | 99.62 |
| **R** | 99.00 | 100.00 | 100.00 |
| **NE** | 86.84 | 95.19 | 99.18 |
| **AE** | 100.00 | 66.67 | 99.44 |
| **Average** | **96.47** | **90.93** | **99.15** |

The comparison of the system constructed in this study with similar systems given in the literature is difficult due to the varieties in arrhythmia types which are classified, the number of arrhythmia types, the classification techniques, data sources, and performance measures. Nevertheless, some results of these studies are presented below to give an indication.

Chazal et al. [5] classified five classes used in this study with a sensitivity of 75.9%. Song et al. [2] reported the results of 99.6%, 95.1%, and 98.9% for sensitivity, specificity and overall accuracy, respectively, to classify six types of arrhythmias. Acır [4] reported the results of %97.6, %93.8, and %95.2, for sensitivity, specificity and accuracy, respectively, for six classes of arrhythmias. Hosseini reported that overall recognition rate was 0.883 for six arrhythmias. [9]. Übeyli [6] reported a total classification accuracy of 98.61% for four types of ECG beats. Osowski et al. [7] reported the mean error of 3.94% for 7-rhythm types.

The performance values of 96.47%, 90.93%, 99.15%, and 96.7% were obtained in this study for sensitivity, selectivity, specificity, and overall accuracy, respectively. Therefore, the system presented in this study seems to outperform most of the similar systems in the literature.

## 5. References


[1] Osowski, S., Hoai, L.T. and Markiewicz, T. “Support Vector Machine-Based Expert System for Reliable Heartbeat Reliable Heartbeat Recognition”, IEEE Trans. on BME, 51(4), 2004.

[2] Song, M.H., Lee, J., Cho, S.P., Lee, K.J. and Yoo, S.K. “Support Vector Machine Based Arrhythmia Classification Using Reduced Features”, International Journal of Control, Automation, and Systems, 3(4), 2005, 571–579.

[3] Palreddy, S., Hu, Y.H. and Tompkins, W.J. "A patient adaptable ECG beat classifier using a mixture of experts approach", IEEE Trans. on BME, 44(9), 1997, 891–900.

[4] Acır, N. “Classification of ECG beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm”, Neural Computing and Applications, 14(4), 2005, 299–309.

[5] P. de Chazal, M. O’Dwyer, and R. Reilly, “Automatic classification of heartbeats using ECG morphology and heartbeat interval features,” IEEE Trans. Biomed. Eng., vol. 51, no. 7, pp. 1196–1206, Jul. 2004.

[6] Übeyli, E.D., ECG beats classification using multiclass support vector machines with error correcting output codes Digital Signal Processing, Volume 17, Issue 3, May 2007, 675-684.

[7] Osowski, S. and Linh, T.R. “ECG Beat Recognation Using Fuzzy Hybrid Neural Network”, IEEE Trans. on BME, 48(11), 2001.

[8] Nadal, J. and Bossan, M. “Classification of cardiac arrhythmia based on principal components analysis and feedforward neural Networks”, Comput Cardiol, 1993, 341–344.

[9] Hosseini, H.G., Reynolds, K.J., and Powers, D., “A Multi-stage Neural Network Classifier for ECG Events”, Proceedings of the 23rd International Conference of the IEEE Engineering in Medicine and Biology Society, 2001, October 25-28.

[10] Minami, K., Nakajima, H. and Toyoshima, T. “Arrhythmia diagnosis with discrimination of rhythm origin and measurement of heart-rate variation”, Comp Cardiol, 1997, 243–246.

[11] Braccini, G., Edenbrandt, L., Lagerholm, M., Peterson, C., Rauer, O., Rittner, R. and Sornmo, L., “Self-organizing maps and Hermite functions for classification of ECG complexes”, Comp Cardiol, 1997, 425–428.

[12] Minami, K., Nakajima, H. and Toyoshima, T. “Real-time discrimination of ventricular tachyarrhythmia with Fourier transform neural network”, IEEE Trans. BME, 46, 1999, 179–185.

[13] Physiobank Archieve Index, MIT-BIH Arrhythmia Database. http://www.physionet.org/physiobank/database

[14] Kohonen, T., (2001). Self-Organizing Maps, 3rd ed., Heidelverg: Springer-Verlag Berlin.

[15] T. M. Cover and P. E. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Inform. Theory, vol. IT-13, pp. 21–27, Jan 1968.

[16] McNeil, B.J., Keler, E., and Adelstein, S.J., Primer on certain elements of medical decision making, J Med (The New England) 293, 211-215 1975