Paper

Application of Feature Extraction Scheme to the Discrimination of Electrocardiogram (ECG)

Member Sadao Fujimura (The University of Tokyo)
Non-member Senya Kiyasu (The University of Tokyo)

Electrocardiograms (ECGs) used for the diagnosis of heart diseases generates large amounts of time-series data. They are regarded as hyperdimensional data. The number of dimension is that of sample points. To automatically recognize any abnormality in the ECG it is essential to extract significant features from the hyperdimensional ECG data. We have already developed a method for purpose-oriented feature extraction and successfully applied it to hyperspectral data which have several hundreds of dimensions. Here we apply the basic idea of this method to the analysis of 12-lead ECGs for the discrimination of abnormal waveforms. ECG data were acquired from normal subjects and from patients who seemed to suffer from one of three classes of abnormalities (anterior myocardial infarction, inferior myocardial infarction, and flattening of the T wave). A small number of features appropriate for discriminating significant patterns of the ECG were extracted. Our method can efficiently process the data and give results relevant to the purpose of diagnosis.

Keywords: time-series analysis, waveform, abnormality, diagnosis, purpose-oriented

1. Introduction

Time series signals are often used for the detection of defects or malfunctions, the amount of data becomes very large for frequent sampling, long periods of observation, or a large number of data channels. In the analysis of electrocardiograms (ECGs), waveforms are simultaneously observed through 12 leads [(1)]. When, for example, the waveforms are observed at 200 sampling points for one cycle of each lead (Fig.1), the total number of dimensions is 2400. Since the time-series data are mostly redundant in the time domain as the data from the 12 leads are correlated with each other, essential information can be expressed by a small number of features. We must derive significant features to extract useful information from according to the purpose of diagnosis.

Many methods have already been developed for diagnosis using ECGs. However, the percentage of ECGs that are correctly classified by computer programs is often lower than that classified by cardiologists [(2)~(4)]. Although most programs intend to directly implement algorithms used by cardiologists and attempt to recognize all of the abnormalities, it is sometimes much more important to recognize specific abnormalities than to discriminate all the classes of them.

Conventional methods for extracting features are not always satisfactory for this type of objective in terms of optimality and computation time. We have already developed a method of purpose-oriented feature extraction for supervised classification and applied it to hyperspectral data [(5)~(7)]. Here, we apply this approach to feature extraction from time-series data. We applied the scheme of feature extraction to the ECG data and discriminated the serious abnormalities which were determined by the waveforms.

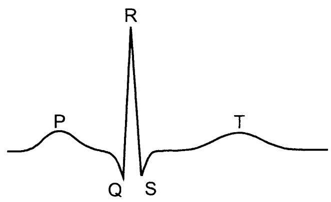


Fig. 1. Waveform of single cycle of ECG

In the application of our method to time-series data, the wavelength in spectral analysis corresponds to sampling time in the time domain. Analogously to the spectral data which are synchronized in wavelength, we have to synchronize the time-series data by the significant peaks in the waveforms.

2. Purpose-oriented feature extraction

2.1 Basic idea In the analysis of data we have certain objectives or intentions. The key idea of our method is to introduce subjective significance explicitly into feature extraction. By this we can extract a set of features which discriminates particular classes of waveforms from others and, at the same time, separates each of the particular classes [(5)].

The features are extracted using a set of training data. After all the data have been orthogonalized and reduced by principal component analysis, a set of appropriate features for prescribed purpose is extracted as linear combinations of the reduced components. The weighting factors to differentiate the normal and the abnormal waveforms were determined as the results of feature extraction. Each dimension is weighted and fused according to the weighting factors.

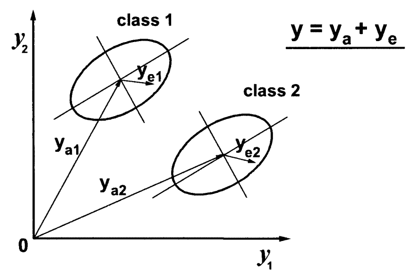


Fig. 2. Description of data

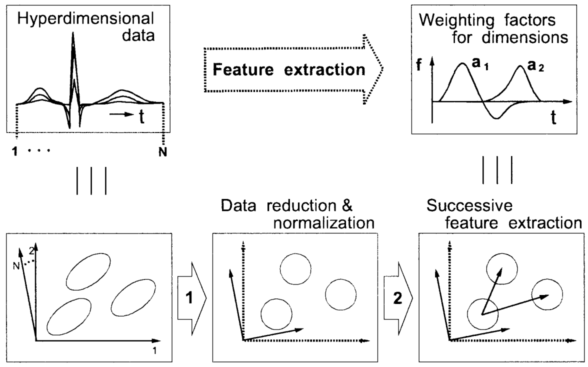


Fig. 3. Procedure of feature extraction

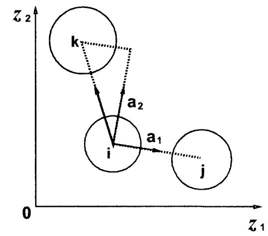


Fig. 4. Feature vectors discriminating between two classes

2.2 Procedure of feature extraction [5]

We assume that we have obtained training data for almost all the classes and can estimate the mean and covariance of these classes. We denote hyperdimensional data (N dimension) by a vector $\boldsymbol{y} = (y_1, \cdots, y_N)^t$ (^{t}: transpose), and assume that they are classified into one of n classes. Then, $\boldsymbol{y}$ can be decomposed into class mean $\boldsymbol{y}_a$ and within-class dispersion $\boldsymbol{y}_e$: that is, $\boldsymbol{y}$ is written as $\boldsymbol{y}_{ij} = \boldsymbol{y}_{a_i} + \boldsymbol{y}_{e_{ij}}$ $(i = 1 \sim n,\ \ j = 1 \sim m_i)$, where $\boldsymbol{y}_{ij}$ is the j-th datum of class i (Fig.2). We write the variance and covariance matrices of $\boldsymbol{y}$, $\boldsymbol{y}_a$ and $\boldsymbol{y}_e$ as $\boldsymbol{C}_{yy}$, $\boldsymbol{C}_a$ and $\boldsymbol{C}_e$ respectively. We call $\boldsymbol{C}_a$ and $\boldsymbol{C}_e$ between-class and within-class covariance matrices, respectively. Here, we assume that the covariance matrix of each class is identical. This assumption is rather reasonable from the viewpoint of the generality of training data. We use a measure of separability between the two classes and extract features which maximize the separability.

Our method consists of two steps: preprocessing and feature extraction (Fig.3). In preprocessing, hyperdimensional data $\boldsymbol{y} = (y_1, \cdots, y_N)^t$ are reduced and normalized to m ($m \ll N$) components $\boldsymbol{z} = (z_1, \cdots, z_m)^t$ by a linear transformation $\boldsymbol{z} = \boldsymbol{A}^t\boldsymbol{y}$. (See 'Appendix' for details of process.) Based on the assumption of $\boldsymbol{C}_e$ that the within-class dispersion is the same for all classes, they are normalized into m dimensional spheres after transformation. This renders the space uniform; this means that the distance measured in terms of variance does not have directionality in space.

In the second step, features are successively extracted until no class remains at a distance from the particular classes less than the minimum distance obtained so far. Feature extraction is performed by determining subspace in the feature space; that is, by making a linear combination of $\boldsymbol{z}$ as $\boldsymbol{a}^t\boldsymbol{z}$, where $\boldsymbol{a}$ is an m dimensional weight vector which we call here a feature vector. Thus, feature extraction consists of the determination of a feature vector. Since the space is now uniform, the direction of an optimal feature vector which discriminates between two classes can be obtained by connecting the centers of these classes. The feature vectors obtained are orthogonalized to make them independent.

The procedure for determining successive feature vectors is as follows.

(1) First, we set an optimal feature vector $\boldsymbol{a}_1$ between the two nearest classes among the prescribed classes.

(2) Next, we evaluate the separability of $\boldsymbol{a}_1$ for all combinations of the prescribed classes.

(3) If there is any pair of prescribed classes which does not have sufficient separability, we set an additional feature vector $\boldsymbol{a}_2$ between them. We orthonormalize the new vector $\boldsymbol{a}_2$ with $\boldsymbol{a}_1$, as shown in Fig.4, so that this feature vector is independent of the first one.

(4) Features are successively extracted in the same way until all the distances among the prescribed classes are larger than the minimum distance obtained so far.

(5) Then, we reapply steps (2)∼(4) to the distances between the prescribed and the other classes.

When only one class is prescribed, the procedure begins by setting a feature vector between that class and its nearest class in the feature space.

The feature $\boldsymbol{a}_i{}^t\boldsymbol{z}$ is equivalent to the $(\boldsymbol{A}\,\boldsymbol{a}_i)^t\,\boldsymbol{y}$ expression using original data $\boldsymbol{y}$, because $\boldsymbol{z} = \boldsymbol{A}^t\boldsymbol{y}$, where $\boldsymbol{A}\,\boldsymbol{a}_i$ means the weighting factor for waveforms.

3. Application to abnormality detection in ECGs

The ECG is one of the most commonly used tools for the diagnosis of heart diseases. Automatic recognition of abnormalities in ECGs will be of great help in medical examinations or the monitoring of general health. In diagnosis using ECG, it is sometimes much more important to be able to find a specific abnormality than to discriminate all classes. We applied the method of purpose-oriented feature extraction to discriminating the waveform of significant abnormalities with higher

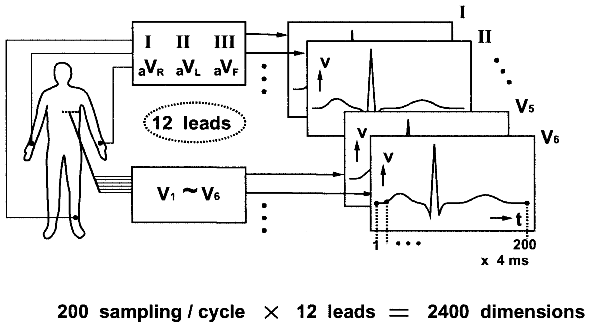


Fig. 5. Detection of 12-lead ECG

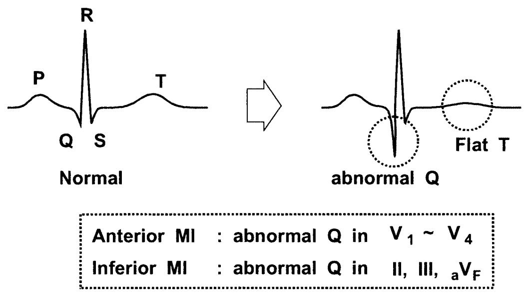


Fig. 6. Abnormalities appearing in ECG corresponding to ischemic heart diseases

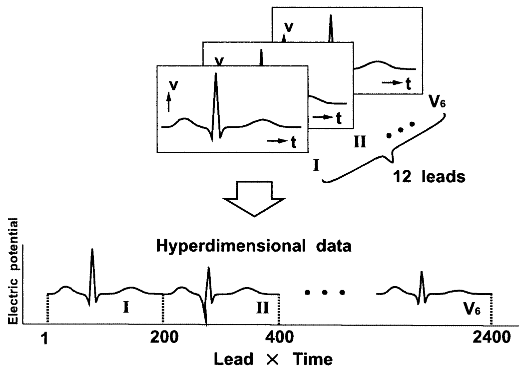


Fig. 7. Hyperdimensional data from 12-lead ECG

priority.

The most commonly used ECG detects the waveform of electric potential through 12 leads, as shown in Fig.5. Therefore, when the waveforms are observed at 200 sampling points for one cycle of each lead , the total number of dimensions is 2400. Since the data from each lead are highly correlated with each other, we can derive essential information from a small number of features according to the purpose of the diagnosis.

4. Experiments and results

Twelve-lead ECG data were acquired from normal subjects and from patients who seem to suffer from one of three classes of abnormalities (anterior myocardial infarction, inferior myocardial infarction, and flattening of the T wave). These abnormalities are known to be typical symptoms of ischemic heart diseases which are the most common and serious heart diseases, especially among middle-aged persons. Figure 6 illustrates the positions of abnormalities in a single cycle of ECG. It is thought that myocardial infarctions (MI) are characterized by abnormal Q waves, and the anterior MI and the inferior MI can be distinguished by the positions of leads where the abnormal Q wave are observed. It is said that anterior MI can be recognized as the presence of abnormal Q in $V_1 \sim V_4$ but not in leads II, III, aV_F. Inferior MI can be recognized as the presence of abnormal Q only in II, III, and aV_F (1). There are several other variations of leads in which abnormal Q waves can be observed. However, since the position of abnormal Q corresponds to the damaged part of the heart, this method can be directly extended to the recognition of other types of myocardial infarctions when appropriate training data are available. A flattened or negative T wave is considered to be a generally observed abnormality in ischemic heart diseases. Although a specific disease cannot be diagnosed solely on the bases of this abnormality, it will be useful in discriminating between this type of waveform and a normal one.

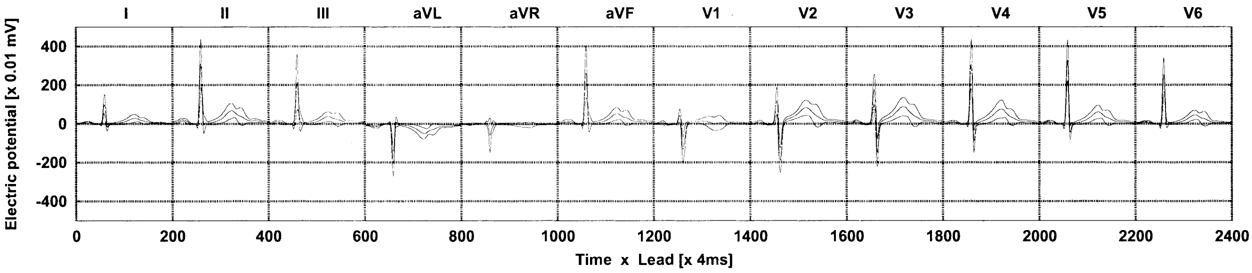
We acquired 15 sets of data for each class and used them in the experiments. The waveforms were measured at 200 sampling points with the sampling period of 4ms for each of the 12 leads (I, II, III, aV_L, aV_R, aV_F, $V_1 \sim V_6$). The waveforms were automatically aligned so that the R waves come to 50th points. We connected the data from 12 leads, as shown in Fig.7, and treated the data as 2400-dimensional long waveforms. Mean values and the regions of $\pm\sigma$ for four classes are shown in Fig.8.

We assumed a case in which one class is to be discriminated from the others. Since the number of classes is only four in this experiment, the number of available features is less than three.

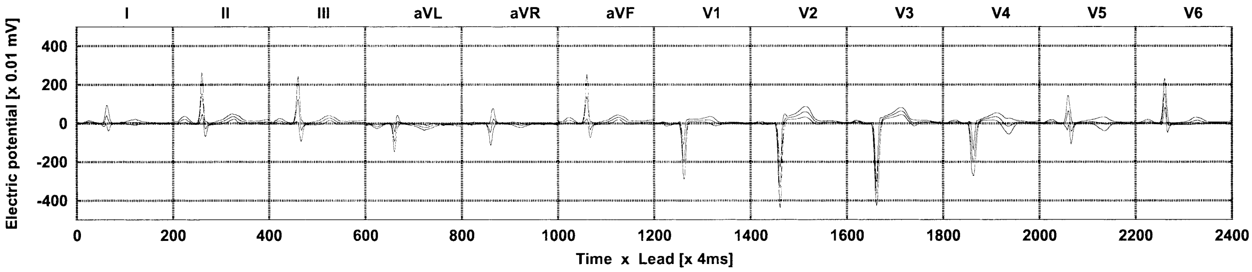
Figure 9 shows the derived weighting factors (first feature) for discriminating the anterior MI from the other classes. It can be seen that the weighting factors are relatively high in the region which corresponds to the abnormal Q waves observed in $V_1 \sim V_4$. This means that this region is the most significant in discriminating the abnormal waves of anterior MI.

Since the number of data were restricted to 15 sets for each class, we applied the "leaving one out" method to estimate the accuracy of classification. Each time we set aside one datum for testing and extracted the feature using the rest of the data which used at the same time as training data. When the weighting factors were derived as a result of feature extraction, the feature values can be calculated as inner products with the observed waveforms. After converted to feature values, test data were classified using the classifier determined by the training data. We checked whether the "leaved out" datum was correctly classified. We shifted the data one by one until all the data were examined to be classified.

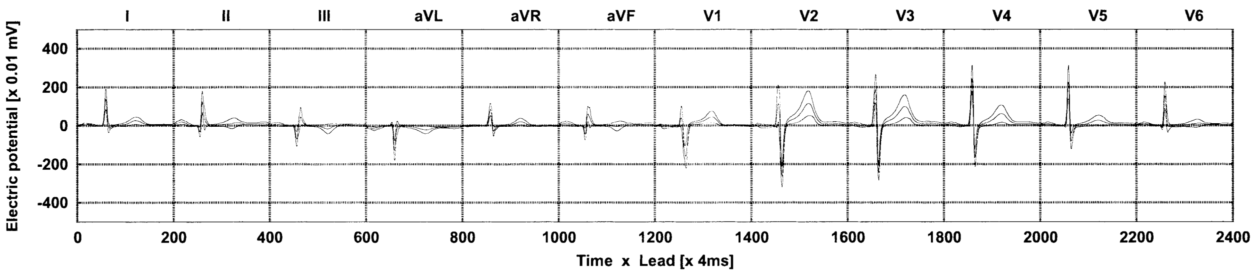
The confusion matrix of classification (15 data sets for each class) when one feature is used is listed in Table 1. Fourteen out of 15 data sets from the anterior MI were classified correctly (93%) by one feature. We neglected



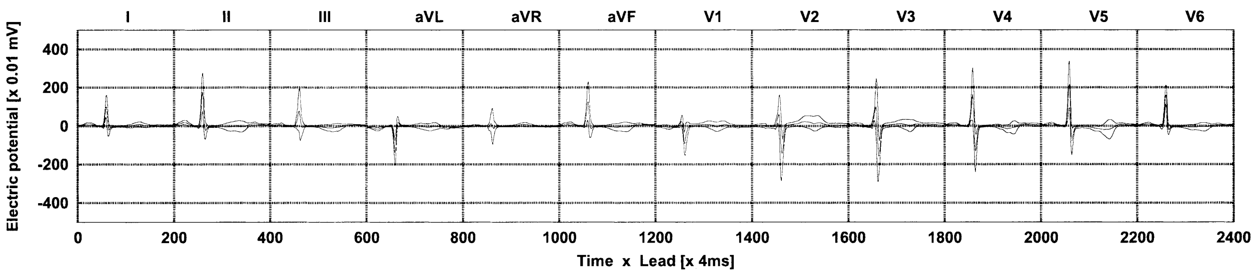
(a) Normal



(b) Anterior myocardial infarction (Abnormal Q are observed in V_1-V_4)



(c) Inferior myocardial infarction (Abnormal Q are observed in II, III, aV_F)



(d) Flattening of the T wave

Fig. 8. Twelve-lead ECG (mean$\pm\sigma$)

the classification of other classes in feature extraction.

Figure 10 illustrates the weighting factors (first feature) extracted for the discrimination of the inferior MI. The weighting factors are relatively high in the region of abnormal Q waves observed in II, III and aV_F. The confusion matrix is listed in Table 2. Twelve out of 15 data sets from the inferior MI were classified correctly (80%).

The difference between the results in Tables 1 and 2 suggests that the performance of our purpose-oriented feature extraction is very high. Table 5 lists the confusion matrix when all (three) features were used, which means the maximum available accuracy of classification. We can see that the extracted features give the same accuracy as the values in Table 5 for the prescribed class.

In the same way, Fig.11 shows the weighting factors for the discrimination of the flat T wave, and Table 3 lists the confusion matrix for one feature. Figure 12 shows the result for the case where both the anterior and inferior MIs were designated to be discriminated from the other classes. The weighting factors show higher weights in the regions of abnormal Q waves of both anterior MI and inferior MI. The confusion matrix in Table 4 shows that the classification accuracy of the two MI

Table 1. Confusion matrix (number of data sets) (object class: Anterior MI, number of features: 1)

	Normal	Ant MI	Inf MI	Flat T
Normal	8	0	3	4
Anterior MI	0	14	0	1
Inferior MI	5	0	9	1
Flat T	5	2	1	7

Table 2. Confusion matrix (number of data sets) (object class: Inferior MI, number of features: 1)

	Normal	Ant MI	Inf MI	Flat T
Normal	2	4	0	9
Anterior MI	4	10	0	1
Inferior MI	0	1	12	2
Flat T	5	0	1	9

Table 3. Confusion matrix (number of data sets) (object class: Flat T wave, number of features: 1)

	Normal	Ant MI	Inf MI	Flat T
Normal	9	3	3	0
Anterior MI	2	8	3	2
Inferior MI	8	5	0	2
Flat T	0	2	0	13

Table 4. Confusion matrix (number of data sets) (object class: Anterior & Inferior MI, number of features: 1)

	Normal	Ant MI	Inf MI	Flat T
Normal	6	1	3	5
Anterior MI	0	13	0	2
Inferior MI	3	0	11	1
Flat T	6	1	0	8

Table 5. Confusion matrix (number of data sets) (number of features: 3 (maximum))

	Normal	Ant MI	Inf MI	Flat T
Normal	13	0	1	1
Anterior MI	0	14	0	1
Inferior MI	2	0	12	1
Flat T	0	1	0	14

classes were weighted at the same time.

This means that our feature extraction method yields the position to be used for diagnosis. These position agree well with those used by medical doctors.

5. Conclusions

We have applied the approach of purpose-oriented feature extraction to the discrimination of electrocardiograms (ECGs). The method was applied to the discrimination of four classes of waveforms (normal, inferior myocardial infarction, anterior myocardial infarction, and flattening of the T wave).

Although the number of data sets used in the experiments was limited (15 data sets for each class), the prescribed classes of myocardial infarction were classified successfully using only the first extracted feature. Our method can efficiently process the time-series hyperdimensional data and give results matched to the purpose of diagnosis. The application of our method to the cases where more classes of ECG are included and the confirmation of its validity for medical applications are subjects for future study.

6. Acknowledgments

Thanks are due to Professor H. Tomoike and Dr. M. Yamaki of Yamagata University, Japan, for providing the ECG data.

(Manuscript received February 4, 2000, revised December 12, 2000)

References

(1) A. R. Houghton and D. Gray: "Making Sense of the ECG: A Hands-On Guide", Arnold Pubns (1997)
(2) Y. Nose: "Performance of automatic interpretative program of ECG", Jpn. J. Med. Electron. Bio. Eng., vol.1, no.1, pp.37–39 (1987)
(3) N. Okamoto, M. Yokoi, N. Ohsawa and T. Iwatsuka: "Evaluation of accuracy on automatic ECG analysis", Jpn. J. Med. Electron. Bio. Eng., vol.1, no.1, pp.40–43 (1987)
(4) J. L. Willems, et al.: "The Diagnostic Performance of Computer Programs for the Interpretation of Electrocardiograms", New England Journal of Medicine, vol.325, pp.1767–1773 (1991)
(5) S. Fujimura and S. Kiyasu: "Significance-Weighted Feature Extraction from Hyper-Dimensional Data", Proc. SPIE, vol.2318, pp.63–68 (1994)
(6) S. Fujimura and S. Kiyasu: "A Method for Object-Oriented Feature Extraction from Hyperspectral Data —Generation of New Channels by Fusion of Data—", Proc. IGARSS'97, pp.975–977 (1997)
(7) S. Fujimura and S. Kiyasu: "Application of an Object-Oriented Feature Extraction Method to Quantitative Estimation from Hyper-Spectral Data", Proc. IGARSS'98, pp.1061–1063 (1998)

Appendix

The transformation of hyperdimensional data $\boldsymbol{y} = (y_1, \cdots, y_N)^t$ to reduced and normalized data $\boldsymbol{z} = (z_1, \cdots, z_m)^t$ is formulated as follows. First we transform N-dimensional data $\boldsymbol{y}$ to only m ($m \ll N$) principal components $\boldsymbol{u} = (u_1, \cdots, u_m)^t$ by

$$\boldsymbol{u} = \boldsymbol{Q}^t \boldsymbol{y}. \qquad \text{(A1)}$$

The transformation matrix $\boldsymbol{Q}$ can be constructed from eigenvectors of the covariance matrix $\boldsymbol{C}_{yy}$ as $\boldsymbol{Q} = (\boldsymbol{q}_1, \cdots, \boldsymbol{q}_m)$. Each eigenvector $\boldsymbol{q}_1, \cdots, \boldsymbol{q}_m$ corresponds to the m largest eigenvalues $\varphi_1, \cdots, \varphi_m$.

After reducing the data, we normalize the within-class dispersion of each class by a linear transformation so that the dispersion is transformed to an m-dimensional sphere. The within-class covariance matrix of reduced data $\boldsymbol{u}$ can be written as $\tilde{\boldsymbol{C}}_e = \boldsymbol{Q}^t \boldsymbol{C}_e \boldsymbol{Q}$. When the eigenvalues of $\tilde{\boldsymbol{C}}_e$ are expressed by $\lambda_1, \cdots, \lambda_m$ ($\lambda_1 > \cdots > \lambda_m$) and the eigenvectors by $\boldsymbol{p}_1, \cdots, \boldsymbol{p}_m$, the transformation is

$$\boldsymbol{z} = \boldsymbol{P}^t \boldsymbol{u}, \qquad \text{(A2)}$$

where $\boldsymbol{P} = (\, \boldsymbol{p}_1/\sqrt{\lambda_1},\ \boldsymbol{p}_2/\sqrt{\lambda_2},\ \cdots,\ \boldsymbol{p}_m/\sqrt{\lambda_m}\,)$. Then, the transformation from $\boldsymbol{y}$ to $\boldsymbol{z}$ is written as

$$\boldsymbol{z} = \boldsymbol{A}^t \boldsymbol{y}, \qquad \text{(A3)}$$

where $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{P}$.

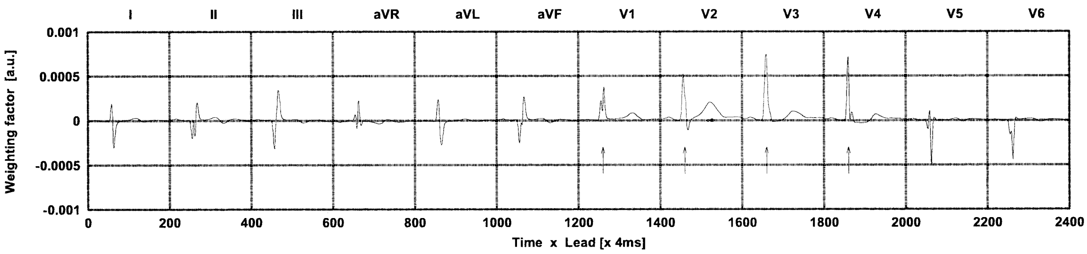


Fig. 9. Extracted first feature for discrimination of anterior MI

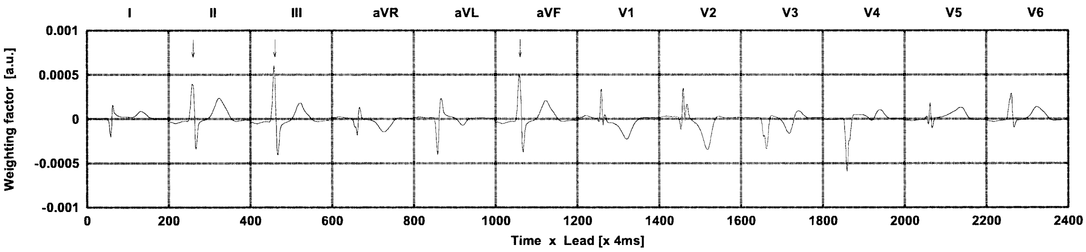


Fig. 10. Extracted first feature for discrimination of inferior MI

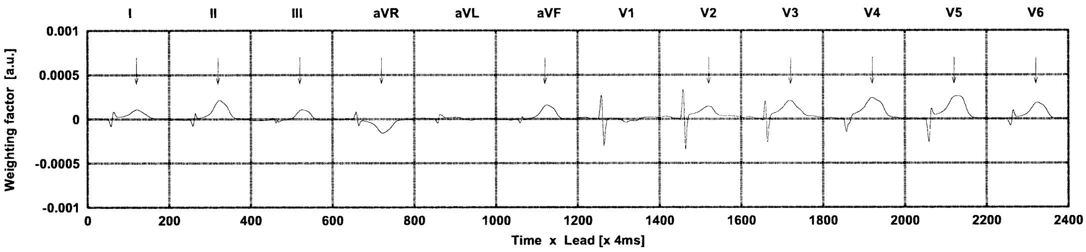


Fig. 11. Extracted first feature for discrimination of the flat T wave

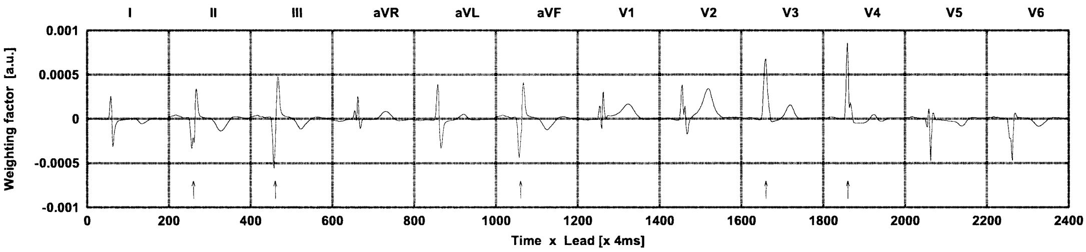
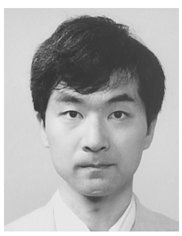


Fig. 12. Extracted first feature for discrimination of both of the anterior and inferior MI

Sadao Fujimura (Member) received the B. Sc., M. Sc., and Ph. D degrees in applied physics from the University of Tokyo, Tokyo, Japan in 1963, 1965 and 1968, respectively. He is currently a Professor at the Department of Information Engineering, Faculty of Informatics, Teikyo Heisei University, Ichihara, Japan. He served as a lecturer in the same department from 1968 to 1973. After that he joined the University of Electro-Communications, Tokyo, as an associate professor, and in 1978 rejoined the Department above of the University of Tokyo, and served as professor there from 1980 to 2000. He joined the present university from 2000. He was a visiting professor at the Department of Electric Engineering, University of Tennessee, Knoxville, Tennessee. His research interests are in image processing, especially multi-dimensional image processing, pattern recognition, infrared thermometry, remote sensing and 3D distribution and shape measurement. He is a fellow of the Society of Instrument and Control Engineers, and a sinior member of IEEE, a member of SPIE, the Japan Society of Applied Physics, the Institute of Electric Engineering and so forth.



Senya Kiyasu (Non-member) received the B.Eng. degree in mathematical engineering and information physics from the University of Tokyo in 1986. From 1986 to 1991, he worked at the Production Engineering Research Laboratory of Hitachi Ltd. After that, he was a research associate at the University of Tokyo from 1991 to 2000, and received the Dr. Eng. degree from the University of Tokyo in 2000. He is now a lecturer at the Department of Information Engineering, Faculty of Informatics, Teikyo Heisei University, Ichihara, Japan. His research interests are in pattern recognition and remote sensing.