British Journal of Medicine & Medical Research
12(12): 1-8, 2016, Article no.BJMMR.22890
ISSN: 2231-0614, NLM ID: 101570965

SCIENCEDOMAIN ***international***
www.sciencedomain.org

Support Vector Machine-based Multi-scale Entropy of Curves Recognition for Electrocardiogram Data

Chien-Chih Wang[1*] and Cheng-Deng Chang[2]

[1]Department of Industrial Engineering and Management, Ming Chi University of Technology, Taiwan.
[2]Unimicron Technology Corporation, Taiwan.

Authors' contributions

This work was carried out in collaboration between both authors. Authors CCW and CDC designed the study, wrote the protocol and wrote the first draft of the manuscript. Author CDC collected the data and searched for the literatures. Authors CCW and CDC analyzed the results. Both authors approved the final manuscript.

Article Information

DOI: 10.9734/BJMMR/2016/22890
Editor(s):
(1) Crispim Cerutti Junior, Department of Social Medicine, Federal University of Espirito Santo, Brazil.
Reviewers:
(1) Anonymous, University Hospital Zurich, Zurich, Switzerland.
(2) Maddury Jyotsna, Deemed University, India.
Complete Peer review History: http://sciencedomain.org/review-history/12711

Original Research Article

Received 3rd November 2015
Accepted 9th December 2015
Published 17th December 2015

ABSTRACT

Objective: Multiscale entropy (MSE) analysis has been widely used to analyze the physiological signals in the frequency domain. Higher complexities of MSE curve present in the physiological system have the better ability to adapt under environmental change. Most people use the subjective experience to distinguish different complexity groups of MSE curves. When the difference between curves is hard to distinguish, the results are often misinterpreted.
Methodology: In this study, four features were designed for the purpose to use the support vector machine technique to develop an automatic recognition procedure for the MSE curve.
Results: A dataset of the electrocardiogram was used to illustrate the proposed analytical process. The results show that AUC is not the only MSE curve feature that should be employed, and new design features may increase recognition ability of MSE curves for electrocardiogram data.
Conclusion: The study results imply that the proposed process can facilitate MSE recognition among nonprofessionals.

**Corresponding author: E-mail: ieccwang@mail.mcut.edu.tw;*

Keywords: Complexity; feature selection; normalization; pattern recognition; classification; R–R interval.

1. INTRODUCTION

Physiological signals provide valuable information to analyze and monitor human health status. Environmental changes or physiological conditions may cause the physiological systems to be up- or down-regulated by the interacting mechanisms that operate across multiple spatial and temporal scales. Some examples are the output signals from physiological monitoring systems such as electromyography (EMG), electroencephalography (EEG), and electrocardiogram (ECG). Costa et al. [1] stated that these signals frequently exhibit complex fluctuations that contain information regarding the underlying dynamics. Multiscale entropy (MSE) analysis, as proposed by Costa et al. [2], is a common method for quantifying the complexity, irregularity, or randomness of physiological time series signals. Physiological signals with higher complexity typically indicate a healthier physiological system [2]. Norris et al. [3] suggested that complexity might be a new clinical biomarker for outcomes. The MSE curves of a patient's heart rate within hours of hospital admission can be used to predict his or her mortality.

Traditional MSE methods that define complexity involve calculating the area under the MSE curve (AUC) or comparing the sample entropy (SampEn) values on the same scale. Trunkvalterova et al. [4] used MSE to detect autonomic dysregulation in young patients with type 1 diabetes mellitus (DM). They found that the MSE of a young patient with DM was significantly reduced on scales 2 and 3. Conversely, SampEn values of SBP and DBP on scale 3 were significantly lower in patients with DM than in healthy participants. Park et al. [5] used MSE to analyze EEG signals from patients with various pathological conditions of Alzheimer's disease (AD) to measure the complexity of the signal. They found that the MSE curves of patients with severe AD showed lower levels of entropy than those of healthy participants and patients with Mild Cognitive Impairment (MCI). Hung and Jiang [6] used MSE to investigate the effect of fatigue on cardiac dynamics during long-term web browsing. They found that the cardiac dynamics of participants who were browsing the web were less complex than those of healthy young participants under free-running conditions. MSE analysis can be used to analyze the signals of physiological time series as well as to investigate the balance problems. Jiang et al. [7] used three cases to introduce MSE analysis applied to the center of pressure (COP) signal and to compare the difference of the COP signal in young and elderly participants. They found higher levels of complexity in young participants compared with elderly participants.

Although the MSE analysis was often sufficient for distinguishing physiological conditions, Park et al. [5] showed that different physiological conditions may have similar AUCs. In such situations, using only AUC cannot distinguish between the MSE curves during varying physiological conditions; thus, other features of the MSE curves must be considered. This study examined four features from the MSE curves and used a support vector machine (SVM) to classify the various patterns of the MSE curves according to combinations of the four features. The analytical results indicate that the new design features may increase the ability to recognize data from the MSE curves.

2. METHODOLOGY

To evaluate a one-dimensional discrete time-series signal $(x_1,\cdots,x_i,\cdots,x_N)$, we constructed the coarse-grained time series (y_j^τ) by averaging the increasing number of data points in nonoverlapping windows (Fig. 1). Each element of the coarse-grained time series (y_j^τ) was calculated as follows:

$$y_j^\tau = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i \tag{1}$$

here τ represents the scale factor and $1 \le j \le (N/\tau)$. The length of each coarse-grained time series is equal to the length of the original time series (N) divided by τ .

After the construction of the coarse-grained time series of a physiological signal, SampEn values for each coarse-grained time series were calculated and plotted as a function of the scale factor. This procedure is known as MSE analysis. For scale 1, the coarse-grained time series is simply the original time series. SampEn proposed by Richman and Moorman [8] is a

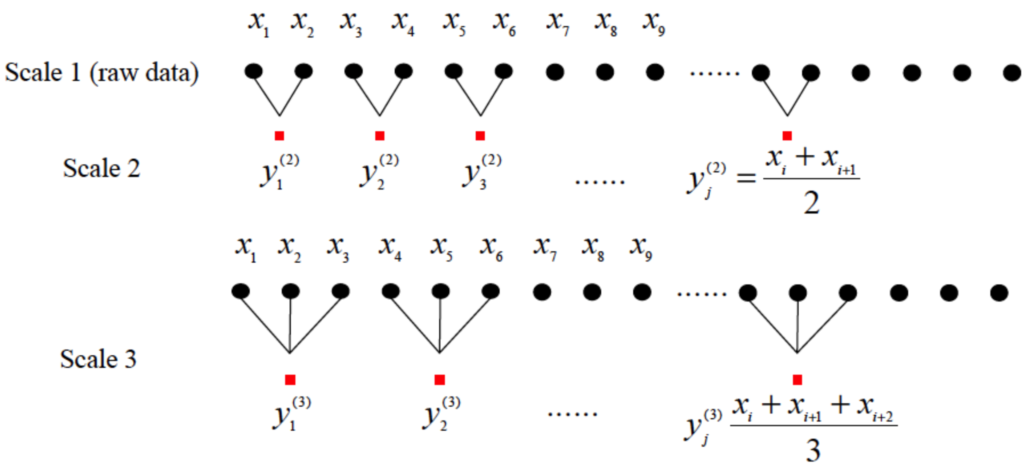

x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9
Scale 1 (raw data)
Scale 2
$y_1^{(2)}$ $y_2^{(2)}$ $y_3^{(2)}$ …… $y_j^{(2)} = \frac{x_i + x_{i+1}}{2}$
Scale 3
$y_1^{(3)}$ $y_2^{(3)}$ …… $y_j^{(3)} \frac{x_i + x_{i+1} + x_{i+2}}{3}$

Fig. 1. The coarse-graining procedure of scales 2 and 3

modification of the approximate entropy (ApEn). The greatest difference between these two entropies is that SampEn does not count self-matches, whereas ApEn does. SampEn has the advantage of being less dependent on the time series length and has greater consistency over a broad range of possible *r*, *m*, and *N* values. For finite length *N*, SampEn is calculated as follows:

$$S_E(m,r,N) = \ln\left(\frac{\sum_{i=1}^{N-m} n_i^{'m}}{\sum_{i=1}^{N-m} n_i^{'m+1}}\right) \qquad (2)$$

where $n_i^{'m}$ differs from n_i^{m} to the extent that for SampEn self-matches are not counted $(i \neq j)$ and $1 \leq i \leq (N-m)$.

A two-stage process was proposed to analyze the MSE curve. The first stage emphasizes describing the features of the MSE curves, and the second stage classifies the MSE curves using SVM.

2.1 Features of the MSE Curves Calculated

The MSE curves are used to compare the relative complexity of normalized time series (same variance as scale 1) by the following guidelines: (1) If the entropy values for the majority of the scales are highest for the past series, the unique series is considered the most complex. (2) A monotonic decrease in the entropy values indicates that the original signal contains information only on the smallest scale. Because higher entropy values connect and form a curve with a larger area, guideline (1) can be interpreted to mean that a greater area under an MSE curve represents a more complex physiological signal.

This study designed four features of the MSE curves to form a coordinated matrix and used this matrix to describe the curve. The features are defined as follows.

2.1.1 Feature 1: AUC

AUC is the most commonly used feature to describe the complexity of an MSE curve. A larger AUC indicates a physiological signal with higher complexity. We used a trapezoidal area to determine the AUC approximately:

$$area(i) = \sum_{j=1}^{\tau-1}\left(\frac{SE_{i,j} - SE_{i,j+1}}{2}\right) \qquad (3)$$

where the *area*(*i*) is the *i*th AUC between the curve *i* and the *X* axis, $j = 1, 2, \cdots, \tau - 1$ represents the scale number, and τ is the maximum scale number. In this study, τ is set to 20. $SE_{i,j}$ represents the *i*th curve's SampEn value in scale *j*. Fig. 2a shows the gray area under the MSE curve between scales 1 and 20.

2.1.2 Feature 2: The slope of maximum difference of small-scale entropies

The MSE curves for physiological signals typically show a relatively substantial increase or decrease on small scales and gradually stabilize on large scales. Therefore, we selected the slope of the maximum difference of small-scale entropies as a feature. According to the MSE curve trend, in this study, the first seven scales (one-third of all scales) were defined as small scales. Fig. 2b shows the slope of maximum difference in the first seven scales. Feature 2 is calculated as follows:

$$slope\left(i_7\right)=\frac{SE_{i,j}-SE_{i,k}}{j-k} \tag{4}$$

where $slope(i_7)$ means the i^{th} curve's maximum slope of the SampEn of the first seven scales, $SE_{i,j}$ is the maximum SampEn value of the first seven scales, and $SE_{i,k}$ represents the minimum SampEn value of the first seven scales. The j and k variables are the numbers on the scales that have the maximum and minimum SampEn values, respectively.

2.1.3 Feature 3: Average entropy value on large scales

Although the values of entropy increase in smaller scales for healthy participants, in larger scales the entropy values become stable. Moreover, the values of entropy obtained from elderly participants are notably lower than those from younger participants. Therefore, different physiological conditions, as well as aging, may be defined in average entropy values on large scales. (4) According to the MSE curve trend, this study used the average entropies of the last of the five scales (one-fourth of all scales) as the third feature, as shown in Fig. 2c. Feature 3 is calculated as follows:

$$ael\left(i_5\right)=\sum_{j=16}^{20}\frac{SE_{i,j}}{5} \tag{5}$$

The $ael(i_5)$ variable is the average SampEn value of the i^{th} MSE curve of the last five scales.

2.1.4 Feature 4: The variation in the absolute difference between the two scales that comprise the first half of the scale

In this study, although the slope of the maximum difference in the first seven scales is a feature, the MSE curves may have differing patterns but similar slopes. Therefore, feature 4 was selected to overcome this situation. Feature 4 is the variation of two entropy values in the two scales that comprise the first half of the scale. The left side of Fig. 2d is an MSE curve plot. Each dot on the right side of Fig. 2d is the absolute value of the difference calculated from every two SampEn values. Feature 4 is calculated as follows:

$$vhs(i)=sd\left\lfloor\left|SE_{i,j+1}-SE_{i,j}\right|\right\rfloor \quad \text{for}(j=1,2,\cdots,9) \tag{6}$$

where the suffix j is the number of the scale, and $vhs(i)$ means the i^{th} curve's standard deviation of the absolute difference between the two scales that comprise the first half of the scale. The variable $sd\left[\left|SE_{i,j+1}-SE_{i,j}\right|\right]$ represents the standard deviation of all $\left|SE_{i,j+1}-SE_{i,j}\right|$ from $\left|SE_{i,2}-x_{i,1}\right|$ to $\left|SE_{i,10}-SE_{i,9}\right|$ in the i^{th} curve. Because of limited data length, signals are used only to calculate the entropies of the first 20 scales; thus, feature 4 focuses on the variations of the first ten scales.

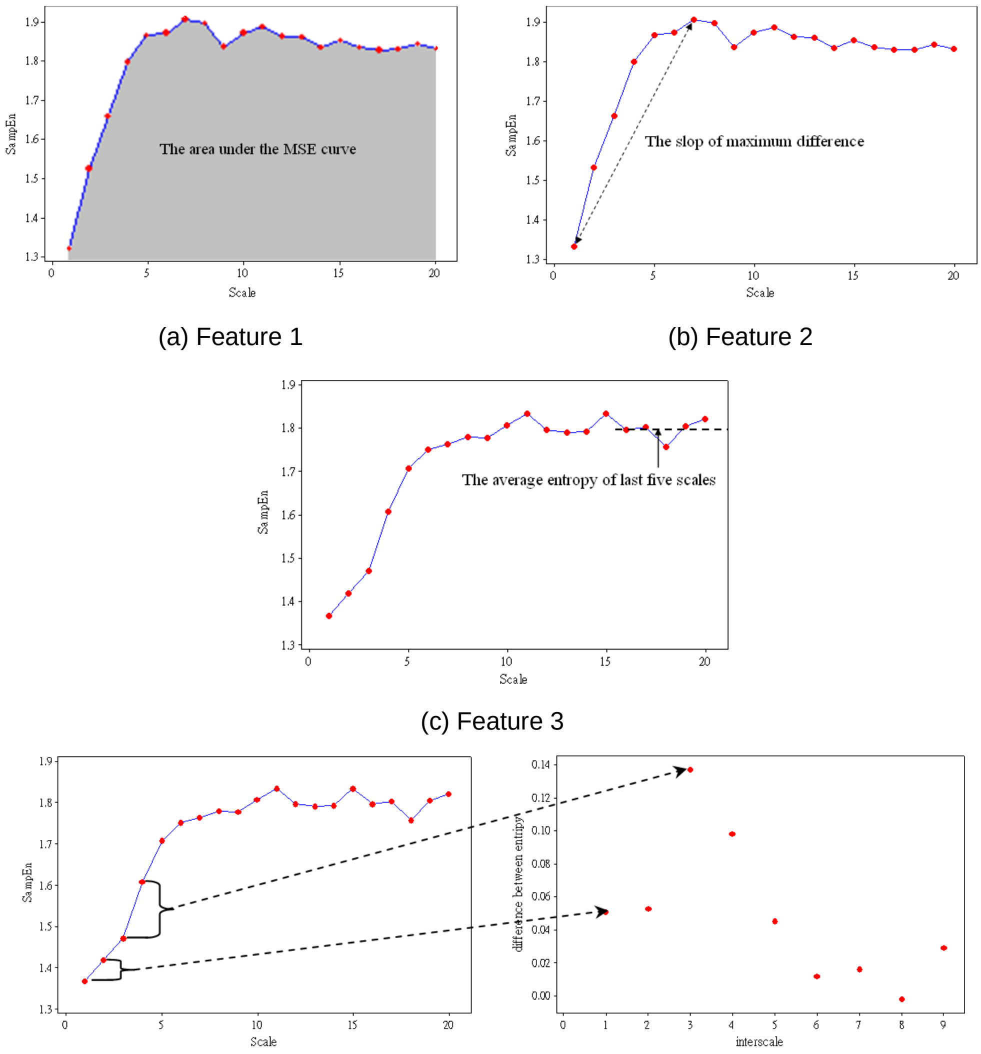
After calculating the four features of the MSE curves in an MSE plot, we normalized each feature vector to avoid differing scale measurements for each feature. The normalization formula is calculated as follows:

$$z_{m,i}=\frac{x_{m,i}-\mu_{X_m}}{\sigma_{X_m}} \tag{7}$$

where $z_{m,i}$ and $x_{m,i}$ are the standardized value and real value of the i^{th} curve, respectively. For example,

$$norm\_area\left(i\right)=\frac{area\left(i\right)-mean\left(area\right)}{stdev\left(area\right)} \tag{8}$$

Here *norm_area*(*i*) is the normalization value of the area of the i^{th} curve. The *mean*(*area*) variable refers to the average of all the curve areas, and *stdev*(*area*) is the standard deviation of all the curve areas. After normalization, the analysis produces four normal vectors of the features; each curve can be drawn in spatial coordinates according to the four features.



(a) Feature 1

(b) Feature 2

(c) Feature 3

(d) Feature 4

Fig. 2. Features of the MSE curves

2.2 Using SVM to Classify the MSE Curves

SVM is a supervised learning technology for classification [9,10]; Chen et al. [11] have stated that SVM performs appropriately for problems with low training sets and nonlinear and multidimensional data. Thus, we used SVM to test the classification accuracy of various physiological states using different feature combinations of the MSE curves. Assuming a set of points as follows:

$$D=\{(\mathbf{x}_i,c_i)\,|\,\mathbf{x}_i\in R^p, c_i\in\{\text{healthymiddle-aged,healthyelderly,CHF}\}\}_{i=1}^{j} \qquad (9)$$

where c_i indicates the class of participant $\mathbf{x}_i$, and each $\mathbf{x}_i$ is a p-dimensional singular values vector. The term $i=1,2,\ldots,j$ is the participant's number. The aim of the SVM is to identify a maximum-margin hyperplane that divides the points into different physiological states (healthy middle-aged, healthy elderly, or patients with congestive heart failure - CHF). This hyperplane can be written as the set of points $\mathbf{x}$ and is expressed as:

$$\mathbf{w}\cdot\mathbf{x}-b=0 \qquad (10)$$

where $\mathbf{w}$ is a normal vector perpendicular to the hyperplane, and $b/\|\mathbf{w}\|$ is the offset of the hyperplane from the origin along the normal vector $\mathbf{w}$. Geometrically, the width of the margin is $2/\|\mathbf{w}\|$ under the minimum $\|\mathbf{w}\|$.

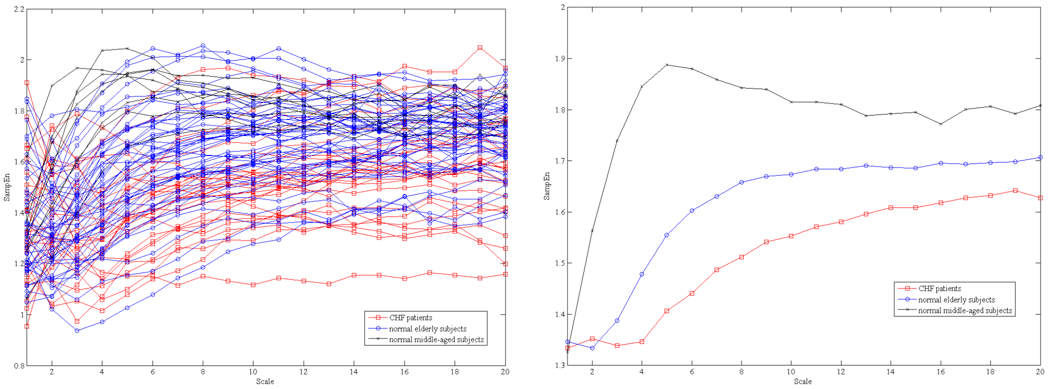
3. ANALYSIS AND DISCUSSION

In this study, a dataset from PhysioNet [12] was used to demonstrate the performance of the proposed method. The dataset included two subsets of participants. The first subset was labeled as the regular sinus rhythm R–R interval database. The data comprised the heart rate R–R interval of 54 healthy participants. The 46 elderly patients were aged 65.87±3.97 years, ranging between 58 and 76 years and eight middle-aged male patients were aged 35.44±4.52 years, ranging between 28.5 and 40 years. The other subset was labeled the CHF R–R interval database and contained 29 patients (aged 55.28±11.60 years) diagnosed with CHF. All data were recorded using an ECG Holter monitor (sampled at 128 Hz) for six-hour when the participants were awake. This study used the R–R interval of healthy participants and patients with CHF and calculated the MSE values of each participant by setting the parameters of MSE to m = 2 and r = 0.15. The MSE curves of the R–R interval signals for all participants are presented in Fig. 3.

Fig. 3 shows the MSE curves of the heart rates from participants in three physiological states. As illustrated in Fig. 3a, most of the MSE curves of patients with CHF were lower than those of the other two groups. For healthy middle-aged participants, the MSE curves were mostly higher than the other curves. However, some of the patients with CHF and healthy elderly participants had AUCs similar to those of the healthy middle-aged participants. The average MSE curve for the healthy middle-aged participants was higher than that for the other two groups. The patients with CHF had the lowest average MSE curve, as shown in Fig. 3b. The healthy middle-aged participants had R–R interval signals with greater complexity than those of the patients with CHF. The complexities of the R–R signals of healthy elderly participants were at an intermediate level.

After extracting the four features of the MSE curves (discussed above), we used the SVM method to compare the classification accuracy of different feature combinations. Using random sampling, we selected 70% of the samples as the training data; the remaining 30% of the data set were used to test the classification effect. Each feature combination was tested ten times, and we used the average accuracy rate of the SVM classification results to compare the effectiveness of all feature combinations. The classification results of all the feature combinations are shown in Table 1.

As shown in Table 1, the average accuracy rate for ten replications of the SVM classification using Fig. 1 (AUC) was 60.1%. Using Feature 1, 2 and 4 can achieve a 68.8% average accuracy rate. Moreover, the fewer features used higher was the accuracy rate. When all the features were used to classify the MSE curves



(a) MSE curves of all participants

(b) Average of MSE curves in different physiological states

Fig. 3. MSE curves of various physiological situations

Table 1. The average accuracy of the SVM classification for feature combinations

Feature combination[1]	Average accuracy rate[2]	Max. accuracy rate
Feature 1	60.1 (2.6)	63.8
Feature 2	56.9 (2.6)	62.6
Feature 3	59.3 (2.4)	61.4
Feature 4	56.1 (1.6)	60.2
Feature 1 & 2	62.8 (2.2)	67.4
Feature 1 & 3	59.7 (1.9)	62.6
Feature 1 & 4	65.3 (4.7)	71.1
Feature 2 & 3	59.7 (2.6)	63.8
Feature 2 & 4	59.1 (3.1)	65.1
Feature 3 & 4	61.5 (3.3)	65.1
Feature 1, 2 & 3	61.6 (2.3)	63.8
Feature 1, 2 & 4	68.8 (2.8)	72.3
Feature 1, 3 & 4	62.4 (3.5)	66.2
Feature 2, 3 & 4	63.1 (2.4)	66.2
All features	65.5 (3.3)	69.8

[1] In the “feature combination” column. Feature 1: AUC, Feature 2: The slope of maximum entropy difference in scales 1 to 7, Feature 3: Average entropy value in scales 16 to 20, Feature 4: Variation in the absolute difference in the following two scales for scales 1 to 10

[2] The value in brackets for the “average accuracy rate” column refers to the standard deviation of the accuracy rate in 10 replications. For example, “60.121 (2.691)” means the average accuracy rate of 10 replications of SVM classification using feature 1 is 60.121, and the standard deviation is 2.691

by SVM, the average accuracy rate of 10 replications was 65.5%. However, when we used only features 1, 2, and 4 as a feature combination, we achieve an average accuracy rate of 68.795%.

The highest accuracy rate was 72.3% for all the analysis results using features 1, 2, and 4 as the feature combination. A comparison of the classification results of using features 1, 2, and 4 as a feature combination with that of using only AUC as the feature shows that using features 1, 2, and 4 as a feature combination for SVM can provide a highest accuracy rate.

4. CONCLUSION

MSE analysis is frequently used to measure system complexity. Although MSE was easily employed when comparing various physiological conditions, feature selection for the MSE curves is a critical task. This study identified four features to describe MSE curves and compared the feature combinations applied to classify the MSE curves with SVM. The electrocardiogram data from PhysioNet was used to illustrate the proposed analytical process. The results showed that the feature combination of AUC, the slope of maximum entropy difference in scales 1 to 7, and the variation in the absolute difference of the following two scales for scales 1 to 10 can provide the highest accuracy rate for all the feature combinations. The contributions of this paper can be summarized into two points. First, AUC is not the only feature that can be used for clustering the MSE curves. We suggest the slope of maximum entropy difference for scales 1 to 7 and the variation in the absolute difference between the two scales following scales 1 to 10 as additional features. Second, fewer the features used, higher is the accuracy rate.

After classifying the MSE curves into several groups, creating estimation models with groups is a critical task. The use of estimation models can help to assess the complexity and physiological condition of patients and can provide invaluable information.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Costa M, Pen CK, Goldberger AL, Hausdorff JM. Multiscale entropy analysis of human gait dynamics. Physica A. 2003; 330:53-60.
2. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Physical Review Letter. 2002;89:068102.
3. Norris PR, Anderson SM, Jenkins JM, Williams AE, Morris Jr JA. Heart rate multiscale entropy at three hours predicts hospital mortality in 3,154 trauma patients. Shock. 2008;30:17-22.
4. Trunkvalterova Z, Javorka M, Tonhajzerova I, Javorkova J, Lazarova Z, Javorka K, Baumert M. Reduced short-term complexity of heart rate and blood pressure dynamic in patients with diabetes mellitus type 1: Multiscale entropy analysis. Physiological Measurement. 2008;29:817-828.
5. Park JH, Kim S, Kim CH. Multiscale entropy analysis of EEG from patients under different pathological conditions. Fractals. 2007;15:399-404.
6. Hung CH, Jiang BC. Multi-scale entropy approach to physiological fatigue during long-term web browsing. Human Factors and Ergonomics in Manufacturing. 2009; 19:478-493.
7. Jiang BC, Yang WH, Shieh JS, JFan JSZ, Peng CK. Entropy-based method for COP data analysis. Theoretical Issues in Ergonomics Science. 2013;14:227-246.
8. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology Heart and Circulatory Physiology. 2000;278:2039-2049.
9. Wang T, Huang H, Tian S, Xu J. Feature selection for SVM via optimization of kernel polarization with Gaussian ARD kernels. Expert Systems with Application. 2010;37: 6663-6668.
10. Shieh MD, Yang CC. Multicall SVM-RFE for product from feature selection. Expert Systems with Applications. 2008;35:531-541.
11. Chen Y, Zhang L, Zhang D. Computerized wrist pulse diagnosis using modified auto-regressive models. Journal of Medical Systems. 2011;35:321-328.
12. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):215-220.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/12711