Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control
Journal homepage: http://kinetik.umm.ac.id
ISSN: 2503-2267
Vol. 5, No. 1, February 2020, Pp. 71-78

Multi-scale entropy and multiclass fisher's linear discriminant for emotion recognition based on multimodal signal

Lutfi Hakim*[1], Sepyan Purnama Kristanto[2], Alfi Zuhriya Khoirunnisaa[3], Adhi Dharma Wibawa[4]
Politeknik Negeri Banyuwangi/Department of Informatics Engineering, Indonesia[1,2]
Institut Teknologi Sepuluh Nopember/Department of Electrical Engineering[3]
Institut Teknologi Sepuluh Nopember/Department of Computer Engineering[4]

Article Info

Keywords:
SpO_2 Signal, Pulse Rate Signal, Support Vector Machine, Linear Discriminant Analysis, Emotion Recognition on Physiological Signals

Article history:
Received 26 July 2019
Revised 20 November 2019
Accepted 02 December 2019
Published 12 Ferbuary 2020

Cite:
Hakim, L., Kristanto, S., Khoirunnisaa, A., & Wibawa, A. (2020). Multi-scale Entropy and Multiclass Fisher's Linear Discriminant for Emotion Recognition based Multimodal Signal. Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, 5(1). doi:https://doi.org/10.22219/kinetik.v5i1.896

*Corresponding author.
Lutfi Hakim
E-mail address:
lutfi@poliwangi.ac.id

Abstract

Emotion recognition using physiological signals has been a special topic frequently discussed by researchers and practitioners in the past decade. However, the use of SpO_2 and Pulse rate signals for emotion recognition is very limited and the results still showed low accuracy. It is due to the low complexity of SpO_2 and Pulse rate signals characteristics. Therefore, this study proposes a Multiscale Entropy and Multiclass Fisher's Linear Discriminant Analysis for feature extraction and dimensional reduction of these physiological signals for improving emotion recognition accuracy in elders. In this study, the dimensional reduction process was grouped into three experimental schemes, namely a dimensional reduction using only SpO_2 signals, pulse rate signals, and multimodal signals (a combination feature vectors of SpO_2 and Pulse rate signals). The three schemes were then classified into three emotion classes (happy, sad, and angry emotions) using Support Vector Machine and Linear Discriminant Analysis Methods. The results showed that Support Vector Machine with the third scheme achieved optimal performance with an accuracy score of 95.24%. This result showed a significant increase of more than 22% from the previous works.

1. Introduction

The emotional condition has been of the particular attention in the health world recently. Many studies reported that emotion has a vital relationship with the health condition, both in positive and negative aspects. A study reported that positive emotions (happy, optimistic, etc.) had a significant impact on immune system improvement and decreased potential mortality [1]. On the contrary, negative emotional factors (such as anger, anxiety, sadness, depression, stress, etc.) were reported to cause a decrease in the immune system. In addition, extreme negative emotion conditions in an individual who has a history of chronic diseases can increase the risk of potential mortality more quickly. It is because of the negative emotion associated with a more significant autonomous system response than the positive emotion [1][2]. Therefore, to reduce the negative impact caused by the excessive negative emotion, a quick and accurate human emotion recognition system is needed. This system will later be used as the basis for developing an e-health-based emotional monitoring system. It can be used by doctors to treat patients and individuals to control the emotional condition that arises suddenly and spontaneously.

Recently, the study on emotion recognition has become a particular topic among researchers and practitioners. In the field of computation and human-computer interaction, emotion recognition utilized some physiological signal modalities, such as Electrocardiograph (ECG), Electroencephalograph (EEG), Pulse rate (PR), Galvanic Skin Response (GSR), Oxygen Saturation (SpO_2), blood volume pressure (BVP), Finger Temperature (FT), Respiration Rate (RR), Heart Rate (HR), Carbon Dioxide (CO_2) and so on. In this context, the use of SpO_2 and PR signals are still limited. The previous study utilized physiological signals with both other modality (known as multimodal signals) and a single modality. Some researchers used SpO_2 signal for emotion recognition combined by other physiological signals, such as what is done by Begum et al. [3], Huang et al. [4], and Wen et al. [5]. Begum et al. combined SpO_2 with FT, RR, HR, and CO_2 signals [3], Wen et al. combined with SpO_2 with GSR and HR signals [5], and Huang et al. combined it with GSR and ECG signals [4]. In another case, Marzuki et. al. only used PR signal to propose emotion recognition [6]. Also, our previous work used SpO_2 and PR signals for emotion recognition in elders [7]. However, the results were still not maximum.

SpO_2 and PR signals are two physiological types that represent oxygen saturation and heart rate activities from the human finger [8]. The characteristic of these signals is less complexity of the pattern. If we look directly based on time series data, it will be challenging to analyze the difference in emotional patterns based on these physiological signals. Therefore, in this work, we used Multiscale Entropy (MSE) for the feature extraction method of SpO_2 and PR signals to analyze the feature vectors of three human emotions. This method is an extension of sample entropy and has been used by some researchers to describe complexity in a time series [9][10][11][12][13]. The purpose of complexity in this context can be defined as "a meaningful structural richness" which incorporates correlation over several temporal scales [9]. It is achieved mathematically by creating multiple sub-time series from the main series and calculating the sample entropy of each scaled series [9]. With calculation using this method, many feature vectors of sample entropy values are generated from both SpO_2 and PR signals.

Some researchers in the field of pattern recognition have found that to find optimal feature vectors, they widely used dimensional reduction methods such as Principal Component Analysis (PCA) and Multiclass Fisher's Linear Discriminant (MFLDA) methods. Pane et al. [14] compared both methods to find emotional patterns based on the EEG signal and the results showed that MFLDA could improve an accuracy score significantly. Therefore, we used the MFLDA method to reduce the dimension of feature vectors to get performance in emotion recognition based on SpO_2 and PR signals. This method was implemented after calculating the MSE method with 20-temporal scales. We hypothesize that these methods for the feature extraction can significantly improve the accuracy of three human emotion recognitions based on these physiological signals. Finally, Support Vector Machine and Linear Discriminant Analysis methods were used to classify the data obtained from the feature extraction process.

2. Research Method

The overall methodology includes feature extraction, dimensional reduction, and emotion classification, as shown in the following Figure 1.

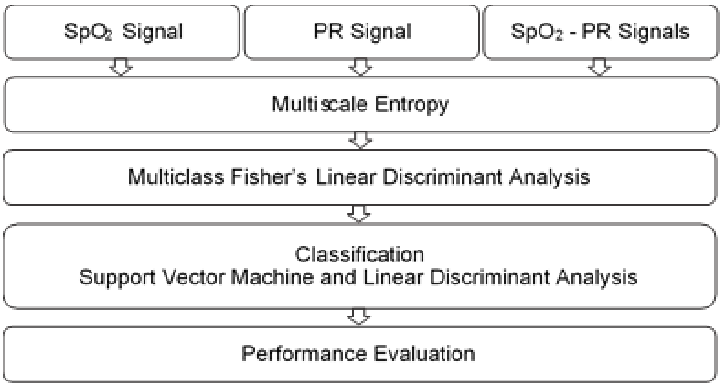


Figure 1. Proposed Method

2.1 Dataset

In this study, we used the dataset of our previous study [7]. It consists of 31 SpO_2 and 31 PR data. All data were the result of emotional induction using video. However, the existing dataset must be validated using questionnaire filling by the participants. The validation results show that there were 21 validated data from three types of emotions (happy, sad, and angry emotions). Furthermore, the validated data were used as input to the data processing.

2.2 Feature Extraction Using Multiscale Entropy

The use of Multiscale Entropy (MSE) as a feature extraction method in various cases of pattern recognition has been proposed. However, its application was still limited to EEG signals, such as analysis of the quality of sleep conditions [15], complexity analysis in Alzheimer's disease [16], arterial wave contour analysis in healthy and diabetes subjects [12]. Only Michalopoulus et al. [17] had conducted the study on emotion recognition. In this study, the MSE method was applied to feature extraction of SpO_2 and Pulse rate signals for emotion recognition. In the previous work [7], eight statistical methods in the time domain were used as the main features in the recognition of emotion, including mean, maximum, minimum, variance, mode, standard deviation, ratio, and mean absolute deviation calculations.

MSE is a method developed by Madalena Costa et al. [10][11], which is used to measure the complexity of a data series on several temporal scales. The idea of this method combines two approaches, namely:

1) Given a sequence of SpO_2 and Pulse rate data $\{x_i\} = \{x_1, x_2, \ldots, x_n\}$ with the set values denoted by $\Theta_1, \Theta_2, \ldots, \Theta_n$ sequentially. The coarse-grained data series is constructed by calculating the average number of data points, which increases continuously in a non-overlapping window to create a different scale and signal resolution. If illustrated, scale 1 is the original time series data. Scale 2 is formed from the average of two consecutive data points, defined by $y_1 = (x_1 + x_2) / 2$; $y_2 = (x_3 + x_4) / 2$ until the last point data of all time series data. Then, scale 3 is formed as the mean of three consecutive data points, $y_1 = (x_1 + x_2 + x_3) / 3$; $y_2 = (x_4 + x_5 + x_6) / 3$ and so on. This calculation is

repeated sequentially according to the set value. Each element of the coarse-grained data has a timeline, ${y_j}^{\tau}$, calculated by the following Equation 1.

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

where τ represents the temporal scale (SF) factor of the data points. Illustration of the coarse-grained method is explained by Madalena Costa et. al. [9].

2) The next step is Sample Entropy (SampEn) approach used to calculate a time-series data, ${y_j}^{\tau}$, with several temporal scales. Thus, the MSE profile consists of determining SampEn at each scale value. To get the SampEn value, the following formula is used Equation 2, Equation 3, Equation 4, Equation 5, and Equation 6.

$$ SampEn\,(m, r, N) = \log_e[U^{m+1}(r)/U^{m}(r)]\,, \tag{2} $$

Where,

$$ U^{m+1}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} U_i^{m+1} \tag{3} $$

$$ U_i^{m+1} = \frac{[\#from\; x_{m+1} \mid d[x_{m+1}(i), x_{m+1}(j)] \le r]}{N-m-1} \tag{4} $$

and

$$ U^{m}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} U_i^{m} \tag{5} $$

$$ U_i^{m} = \frac{[\#from\; x_{m} \mid d[x_{m}(i), x_{m}(j)] \le r]}{N-m-1} \tag{6} $$

x_m is a vector of *m* vector data series of (N-m) length, and $d[x_m(i), x_m(j)]$ denotes the distance formula (usually using Euclidean distance). See Richman and Moorman for details of the SampEn algorithm [18].

2.3 Multiclass Fisher's Linear Discriminant Analysis (M-FLDA)

M-FLDA is one of the transformation methods developed by R.A. Fisher, whose works begin with the linear Discriminant method [19][20]. This method is used in statistical cases, data mining, pattern recognition, and machine learning to reduce the dimensions of data points. It looks for linear combinations of features that separate more than two classes from data objects [21]. M-FLDA is known as a linear equation operation that is used to find the optional projection of the W^Tmatrix to a lower dimension. The principle of this method is to look for linear projection vectors that maximize the inter-class distribution ratio in the class. It aims to reduce the variation of data in the same class and increase the distance of data distribution between classes [20].

It is assumed that the sample data is defined as $x_i = \{x^1, x^1, \dots, x^{(m)}\}$. It finds $[y_1, y_2, \dots, y_{C-1}]$ projection using several classes of C - 1 through the projection of matrix-vector $W_i = [\theta_1|\theta_2|\dots|\theta_{C-1}]$, where $y_i = \theta_i^T x_i \Rightarrow y = W_i^T x_i$. Because the projection is no longer scalar (has a C - 1 dimension), a distribution matrix is used to obtain the scalar objective function through the equation: $J(W) = \frac{W^T S_B W}{W^T S_W W}$, where S_B is an inter-class distribution matrix and S_W is a distribution matrix in the class. S_B and S_W are used to measure the distribution in the x_i feature vector and defined by: $S_B = \sum_C (\mu_C - \bar{x})(\mu_C - \bar{x})^T$, and $S_W = \sum_C \sum_{i \in C} (x_i - \mu_C)(x_i - \mu_C)^T$, with $\bar{x}$ as overall average of the data, μ_C as an average of each class C. Furthermore, a projection matrix of W is used to maximize the distribution ratio by solving the eigenvalue for each column in the weight matrix W. The optimal eigenvalue of W represents the column in the W matrix based on the largest eigenvalue λ. The following equation is used to calculate the W weight matrix that is $S_B W = \lambda\, S_W\, W$, where $W = S_W^{-1/2}\, U$ and $U = S_W^{-1/2}\, S_B\, S_W^{-1/2}$ obtained from eigenvectors.

2.4 Classification

Support Vector Machine and Linear Discriminant Analysis algorithms are often used in cases of emotion recognition of physiological signals. Therefore, in this paper, these two algorithms are applied for the classification of emotions into three classes, namely happy, sad, and angry conditions.

1) **Support Vector Machine (SVM)** is a learning method for supervised learning, which was first introduced by Corina Cortes and Vladimir Vapnik [22]. When first introduced, SVM could only be used to classification cases of two classes. However, further study was conducted so that this algorithm can be used for multiclass cases using several approaches, such as one-against-all, one-against-one, directed acrylic graph SVM, Error-correcting output code SVM, and All-at-once SVM approaches. In classification, this algorithm can work in cases of data that can be separated linearly and non-linearly. In a general case, SVM uses non-linear classification. The basic principle of SVM is to determine an optimal hyperplane to separate data [22]. The hyperplane is optimal if it is located in the middle of the training data and has the largest distance d. To obtain this condition is done by minimizing the value of the Lagrange formula of a primal problem ($\min_{w,b,\xi} L_P(w,b,\xi)$) and maximizing margin (d) through the Lagrange formula of dual problem $\max_{\alpha} L_D(\alpha)$, where Equation 7 and Equation 8.

$$L_p(w,b,\alpha,\mu,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\{y_i(x_i^T w + b) - 1 + \xi_i\} - \sum_{i=1}^{n}\mu_i\xi_i \tag{7}$$

$$L_D(\alpha) = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j \tag{8}$$

with C as a parameter which determines a penalty due to errors in classification of data, ξ = slack variable, $\alpha_i \geq 0$ and $\mu_i \geq 0$: Lagrange multiplier. The training data which is closest to the separator field is called a support vector. This study utilized the Library Support Vector Machine (LibSVM) developed by Chang and Lin [23] using a one-against-one approach for multiclass cases. With this approach, n (n-1)/2 classification models are built, where n is the number of classes. Consider the decision function for training data from *i* and *j* classes, with maximum margin to be: $d_{ij}(x) = w_{ij}^T\, g(x) + b_{ij}$, where w_{ij}^T is dimensional vector l, $g(x) = sign(x^T w + b)$ is a function of mapping which maps to the dimensional feature space l, and b_{ij} is bias term. In addition, for non-linear cases, the transformation of data space into the feature space is applied using "Kernel Trick". Some commonly used kernels are linear, polynomials, Radial Basis Function (RBF), and sigmoid. This study used RBF kernel as recommended by Hsu et. al. [24] which is defined by: $K(x_i x_j) = \exp\left(-\gamma\, \|x_i - x_j\|^2\right), \gamma > 0$.

2) **Linear Discriminant Analysis (LDA).** In this study, the LDA method was used to classify three emotions as a comparison. This method is another way to determine linear transformation from data samples by reducing the dimensional number needed to represent them. LDA can be used for dimensional reduction and classification. This method is included in supervised learning. Like the SVM method, LDA uses the hyperplane principle to separate data samples between classes. However, on LDA, this hyperplane is defined as the distance between the cluster center and data, which is used to determine the sample data into the right target class. For multiclass classification, it needs more than one hyperplane. The principle of LDA is that the data is modeled with a Multivariate Gaussian distribution in each class and applies Bayes rules for classification purposes. Each class is assumed to have a different normal distribution with the same covariance matrices for several different classes (so the resulting classifier will be linear). For each class c, the data sample is determined based on a Multivariate Gaussian Distribution, with a specific covariance (Σ) and average (μ_c) matrix. It is assumed that the covariance matrix for each class is similar, so the distribution has a linear shape. Then, the linear discriminant function is determined by: $y_c = x^T\Sigma^{-1}\mu_c - \frac{1}{2}\mu_c^T\,\Sigma^{-1}\mu_c + \log\left(\frac{n_c}{n}\right)$, where n is a sum of all data and n_c is the amount of data from each class c. Sample data is classified based on the selection of the largest y_c value.

2.5 Performance Evaluation

Cross-Validation is used to get an optimal classification result and avoid overfitting conditions. There are several methods included in the cross-validation type, but this study used k-fold cross-validation. It is a validation technique model that is often used for performance evaluation of classification cases. It is useful for generalizing data into an independent subset. In this method, the sample data is divided into random folds. One fold is used as a test data and the other (K-1) is used as training data [25]. The classification result is arranged in a confusion matrix table. Accuracy parameters are used to measure the performance of the proposed method whose values are obtained from the Equation 9.

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{9}$$

where TN = amount of predicted data correctly in a positive class, TP = amount of mispredicted data in a positive class, TN = amount of predicted data correctly in a negative class, and FP = amount of mispredicted data in a negative class.

3. Results and Discussion

The overall classification results, from three schemes which are described in the previous section, are illustrated in Figure 4 in the first experimental scheme, as shown in Figure 4(a), it is evident that the fluctuations in the classification accuracy score using SVM show inconsistency (up and down score) along with differences in the value of m and r compared to the LDA method. In LDA, the classification accuracy on the set of parameter values m = 2 shows consistently superior to m = 1. This situation occurs in the set of parameter values r between 0.15 and 0.23. The Optimum accuracy in this classification scheme obtained the same score (69.84%), both on SVM and LDA methods. In addition, the average accuracy on the set value m = 1 using SVM is better than m = 2. The results are the opposite when using the LDA method where the accuracy score in the set m = 2 is better than m = 1.

Scheme 2 has similar steps starting from the feature extraction stage to the classification. The tuning parameters of m and r are also applied to 21 validated datasets. This second experimental scheme also shows a less significant difference in accuracy scores. It is shown in Figure 4(b) where the average accuracy of both classification methods and the set of parameter values m = 1 and m = 2 shows a score between 61.5% - 63.5%. In this scheme, average accuracy of m = 2 (63.49% in both SVM and LDA methods) is better than m = 1 (SVM = 61.51% and LDA = 63.0%). In contrast, an optimal score is obtained in the set values m = 1 and r = 0.21 using the SVM method. This maximum score is 71.43%. It increases by 1.59% compared to the optimal accuracy in the previous experimental scheme. However, a minimum accuracy in this scenario shows a lower score than before, which is down to 4% in the set value of m = 1 and r = 0.10 using the SVM method.

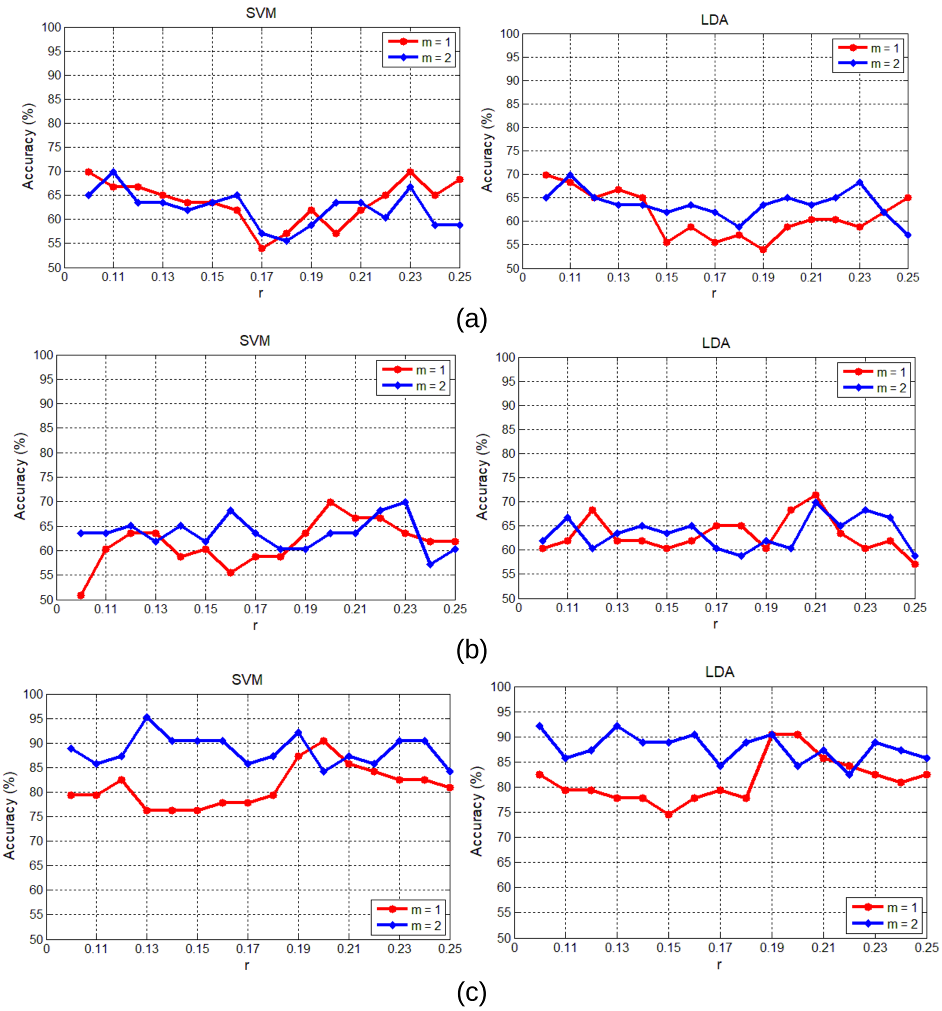
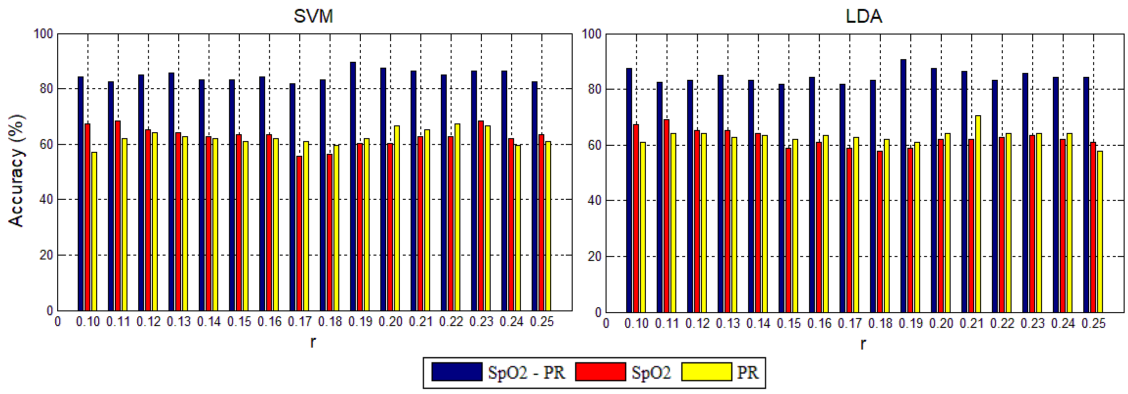
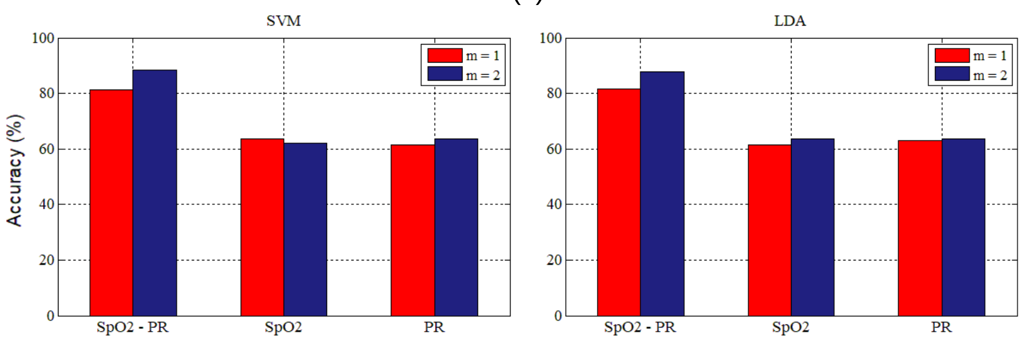


Figure 2. Classification Results with Tuning Parameters of m = 1-2 and r = 0.10 - 0.25 on MSE Method with Feature Vectors Input: (a) SpO2, (b) PR, and (c) SpO2-PR Signals

Considering the accuracy score in the first and second schemes, we try to combine MSE feature vectors from both physiological signals (known as multimodal signals). 40 MSE features, each consisting of 20 feature vectors of SpO_2 and PR signals, are then reduced in dimensions and classified by the same methods. The dimensional reduction and classification process are also done by tuning parameters of *m* and *r* with the same range values as the previous scheme. With this scheme, the classification accuracy is better than the two previous schemes. It is indicated by an increase in classification accuracy, which is very significant both in the set value m = 1 and m = 2. The set value m =1 and r = 0.19-0.20 shows better accuracy where the highest accuracy is obtained with a score of 90.48% in both SVM and LDA methods. However, the accuracy score decreases on the set values between 0.12 and 0.18 and the lowest value is 74.60% (r=0.15). Meanwhile, the classification results on the set m = 2 show better scores compared to m = 1 and the previous classification scheme. Accuracy scores show the value of more than 80% when r value is set from 0.10 to 0.25. The highest accuracy is 95.24%, which is obtained in the set value m = 2 and r = 0.13 with the SVM method. This score shows higher than 3.18% compared to the highest accuracy on the LDA method. This experiment shows that the average accuracy score on SVM also shows a better score than LDA (mean accuracy of SVM = 88.49% and LDA = 87.08%).

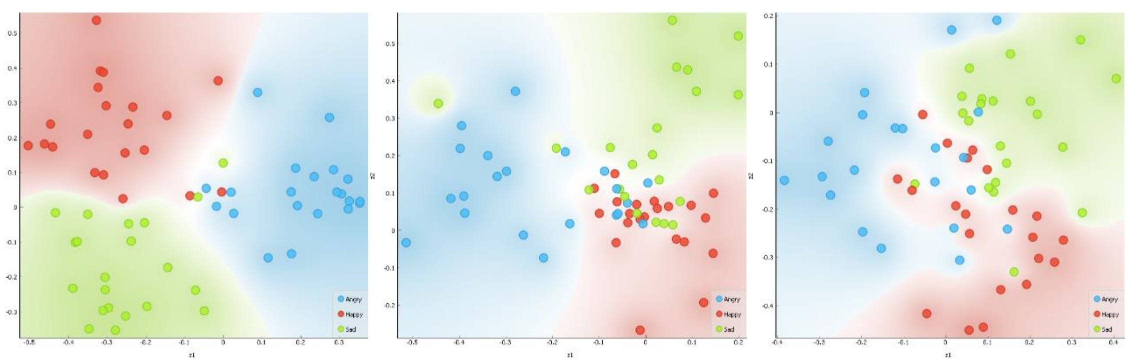


(a)



(b)

Figure 3. Average Accuracy for All Classification Schemes: (a) for Parameter set m = 1 and m = 2, (b) Set Parameter r in the Range Value 0.10 – 0.25



(a) (b) (c)

Figure 4. Scatter Plot Data of the Optimum Classification From Each Type of Physiological Signals: (a) SpO_2 and PR Signals, (b) SpO_2 Sgnal, c) PR Signal

Three experimental schemes above show that the change in *r*-value of the MSE method shows an inconsistent performance. It occurs in both sets of parameter values (m=1 and m=2), as shown in Figure 2(a). Therefore, the average accuracy score of the set *r*-value = 0.10 – 0.25 is calculated for both m=1 and m=2 and in both classification methods. The barplot in Figure 3 illustrates a calculation of average accuracy. Moreover, the classification in the parameter set m=2 shows a better accuracy score than m = 1. The higher accuracy score range (more than 6%) is obtained in a multimodal signal (SpO_2 and PR), which occurs in both classifier methods.

The classification scheme of combined features from SpO_2 and PR signals (scheme 3) shows the optimal results compared to the first and the second classification schemes. This result is also shown in Figure 4, where scatter plots show illustrations of clearly separated data compared to experimental schemes of single signal modality. These results indicate more feature vectors involved in the process of dimensional reduction. It will affect the performance of the proposed method. MFLDA can work better to get optimum accuracy by involving as many feature vectors as input in the dimensional reduction process. It was also confirmed by the results of a study conducted by Septiana Pane et al. [14] where more involved electrodes of EEG signals can affect the performance of the MFLDA algorithm.

The results of this study are better than the previous work [7]. We know that the optimal accuracy of the previous work only reached 72.86% using the SVM method. The use of the MSE method for feature extraction and dimensional reduction using MFLDA had proven to contribute to the improvement of better accuracy of emotion recognition. The proposed method of this study shows a significant increase by 22.38% compared to previous work.

4. Conclusion

A systematic investigation had been carried out to achieve the main objective of this study. Two proposed methods (MSE and MFLDA), as feature extraction and dimensional reduction, can improve the accuracy of emotion recognition in elders based on SpO_2 and PR signals. Based on the experimental results in each scheme, the classifier methods (SVM and LDA) can work well in emotion recognition on scheme 3. The optimum accuracy is 95.24% obtained by using the SVM method.

These results indicate that SVM is better than LDA in the classification process of emotion. Also, the use of multimodal signals is more recommended in the emotional classification process based on physiological signals. Involving more multimodal signals in feature extraction using MSE and dimensional reduction using the MFLDA method can be considered in the development of a real-time monitoring system of human emotion.

Notation

N or n : the number of data
τ : Temporal Scale.
μ : The average value of data.
x_m : A vector of m vector data point.
S_B : Inter-class distribution matrix
S_W : Distribution matrix.
$\bar{x}$: Overall average of the data.
μ_C : Average of each class
C : Cost or penalty
ξ : Slack variable
w_{ij}^T : Dimensional vector l

Acknowledgement

We would like to convey gratitude to Endownment Fund for Education (LPDP) for research funding with contract number 20160711018121 and LPPM Poliwangi for supporting our research.

References

[1] S. D. Pressman and S. Cohen, “Does positive affect influence health?,” *Psychol. Bull.*, Vol. 131, No. 6, Pp. 925–971, 2005. https://doi.org/10.1037/0033-2909.131.6.925
[2] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” *Biol. Psychol.*, Vol. 84, No. 3, Pp. 394–421, 2010. https://doi.org/10.1016/j.biopsycho.2010.03.010
[3] S. Begum, S. Barua, and M. U. Ahmed, “Physiological sensor signals classification for healthcare using sensor data fusion and case-based reasoning,” *Sensors (Switzerland)*, Vol. 14, No. 7, Pp. 11770–11785, 2014. https://doi.org/10.3390/s140711770
[4] W. Huang, G. Liu, and W. Wen, “MAPD: A Multi-subject Affective Physiological Database,” *Proc. - 2014 7th Int. Symp. Comput. Intell. Des. Isc. 2014*, Vol. 2, Pp. 585–589, 2015. https://doi.org/10.1109/ISCID.2014.247
[5] W. Wen, G. Liu, N. Cheng, J. Wei, P. Shangguan, and W. Huang, “Emotion Recognition Based on Multi-Variant Correlation of Physiological Signals,” *IEEE Trans. Affect. Comput.*, Vol. 5, No. 2, Pp. 126–139, 2014. https://doi.org/10.1109/TAFFC.2014.2327617
[6] A. Marzuki, L. D. Rumpa, A. D. Wibawa, and M. H. Purnomo, “Classification of human state emotion from physiological signal pattern using pulse sensor based on learning vector,” *Int. Semin. Intell. Technol. Its Appl.*, Pp. 129–133, 2016. https://doi.org/10.1109/ISITIA.2016.7828646

[7] L. Hakim, A. D. Wibawa, E. Septiana Pane, and M. H. Purnomo, "Emotion Recognition in Elderly Based on SpO2 and Pulse Rate Signals Using Support Vector Machine," *Proc. - 17th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2018*, Pp. 474–479, 2018. https://doi.org/10.1109/ICIS.2018.8466489
[8] A. Jubran, "Pulse oximetry," *Crit. Care*, Vol. 19, No. 1, Pp. 1–7, 2015. https://doi.org/10.1186/s13054-015-0984-8
[9] M. Costa, A. L. Goldberger, C. Peng, B. Israel, and D. Medical, "Multiscale Entropy Analysis (MSE)," *Entropy*, No. 1, Pp. 1–8, 2003.
[10] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series.," *Phys. Rev. Lett.*, Vol. 89, No. 6, Pp. 68102, 2002. https://doi.org/10.1103/PhysRevLett.89.068102
[11] M. Costa, A. L. Goldberger, and C. K. Peng, "Multiscale entropy analysis of biological signals," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, Vol. 71, No. 2, Pp. 1–18, 2005. https://doi.org/10.1103/PhysRevE.71.021906
[12] A. B. Liu, H. T. Wu, C. W. Liu, C. C. Liu, C. J. Tang, I. T. Tsai, and C. K. Sun, "Application of multiscale entropy in arterial waveform contour analysis in healthy and diabetic subjects," *Med. Biol. Eng. Comput.*, Vol. 53, no. 1, pp. 89–98, 2015. https://doi.org/10.1007/s11517-014-1220-4
[13] A. S. Bhogal and A. R. Mani, "Pattern Analysis of Oxygen Saturation Variability in Healthy Individuals : Entropy of Pulse Oximetry Signals Carries Information about Mean Oxygen Saturation," Vol. 8, No. August, Pp. 1–9, 2017. https://doi.org/10.3389/fphys.2017.00555
[14] E. S. Pane, A. D. Wibawa, and M. H. Purnomo, "Peningkatan Akurasi Pengenalan Emosi pada Sinyal Electroencephalograpy Menggunakan Multiclass Fisher Discriminant Analysis," *J. Nas. Tek. Elektro dan Teknol. Inf.*, Vol. 7, No. 4, Pp. 437–443, 2019. http://dx.doi.org/10.22146/jnteti.v7i4.462
[15] C.-E. Kuo, S.-F. Liang, Y.-H. Shih, and F.-Z. Shaw, "Evaluating the Sleep Quality Using Multiscale Entropy Analysis," *IFMBE Proc.*, Vol. 47, Pp. 166–169, 2015. https://doi.org/10.1007/978-3-319-12262-5_46
[16] T. Murata, M. Kikuchi, K. Takahashi, T. Mizuno, Y. Wada, R. Y. Cho, and T. Takahashi, "Assessment of EEG dynamical complexity in Alzheimer's disease using multiscale entropy," *Clin. Neurophysiol.*, Vol. 121, No. 9, Pp. 1438–1446, 2010. https://doi.org/10.1016/j.clinph.2010.03.025
[17] K. Michalopoulos and N. Bourbakis, "Application of Multiscale Entropy on EEG Signals for Emotion Detection," *IEEE EMBS Int. Conf. Biomed. Heal. Informatics Proc.*, No. 5, Pp. 341–344, 2017. https://doi.org/10.1109/BHI.2017.7897275
[18] J. S. Richman, D. E. Lake, and J. R. Moorman, "Sample Entropy," *Methods Enzymol.*, Vol. 384, No. 1991, Pp. 172–184, 2004. https://doi.org/10.1016/S0076-6879(04)84011-4
[19] M. Welling, "Fisher linear discriminant analysis." Toronto, 2005.
[20] C. Li and B. Wang, "Fisher Linear Discriminant Analysis." Pp. 1–6, 2014.
[21] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Fourth Edition Data Mining: Practical Machine Learning Tools and Techniques*, Fourth. Cambridge: Todd Green, 2017.
[22] V. Vapnik and C. Cortes, "Support Vector Networks," *Mach. Learn.*, Vol. 20, No. 3, Pp. 273–297, 1995. https://doi.org/10.1007/BF00994018
[23] C. Chang and C. Lin, "LIBSVM : A Library for Support Vector Machines," *ACM Trans. Intell. Syst. Technol.*, Vol. 2, Pp. 1–39, 2013.
[24] C.-W. Hsu, C.-C. Chang, and C. Lin, "A Practical Guide to Support Vector Classification," *BJU Int.*, Vol. 101, No. 1, Pp. 1396–400, 2008.
[25] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of Database Systems*, New York: Springer Science+Business Media, LLC, 2009, Pp. 532–538. https://doi.org/10.1007/978-0-387-39940-9