Emotion Recognition using Neighborhood Components Analysis and ECG/HRV-based Features

Hany Ferdinando[1,2][0000-0003-0857-2946], Tapio Seppänen[3][0000-0002-3963-0750], and Esko Alasaarela[1]

[1] Health and Wellness Measurement Research Unit, Opto-electronic and Measurement Technique (OPEM) unit, University of Oulu, Finland
[2] Department of Electrical Engineering, Petra Christian University, Indonesia
[3] Physiological Signal Team Analysis, University of Oulu, Finland
`hany.ferdinando@oulu.fi`

Abstract. Previous research showed that supervised dimensionality reduction using Neighborhood Component Analysis (NCA) enhanced the performance of 3-class problem emotion recognition using ECG only where features were the statistical distribution of dominant frequencies and the first differences after applying bivariate empirical mode decomposition (BEMD). This paper explores how much NCA enhances emotion recognition using ECG-derived features, esp. standard HRV features with two difference normalization methods and statistical distribution of instantaneous frequencies and the first differences calculated using Hilbert-Huang Transform (HHT) after empirical mode decomposition (EMD) and BEMD. Results with the MAHNOB-HCI database were validated using subject-dependent and subject-independent scenarios with kNN as classifier for 3-class problem in valence and arousal. A t-test was used to assess the results with significance level 0.05. Results show that NCA enhances the performance up to 58% from the implementation without NCA with p-values close to zero in most cases. Different feature extraction methods offered different performance levels in the baseline but the NCA enhanced them such that the performances were close to each other. In most experiments use of combined standardized and normalized HRV-based features improved performance. Using NCA on this database improved the standard deviation significantly for HRV-based features under subject-independent scenario.

Keywords: NCA, emotion recognition, ECG, HRV

1   Introduction.

Previous research has reported that applying supervised dimensionality reduction (SDR) significantly enhanced the performance of emotion recognition using ECG from the MAHNOB-HCI database [1]. To be more specific, the Neighborhood Components Analysis (NCA) outperformed the Linear Discriminant Analysis (LDA) and the Maximally Collapsing Metric Learning (MCML), and the SDRs were only applied to features resulted from one method, i.e. the statistical distribution of dominant frequencies

and the first differences after applying the bivariate empirical mode decomposition (BEMD) to ECG signals, which showed its superiority in the absence of the SDR [2].

Apparently, analysis on the previous research has suffered from the number of feature extraction methods. We did not know how well the NCA can enhance the performance of the same system using features from other methods. This is the main research question addressed in this paper because conclusions based on the one feature extraction method may lead to wrong interpretation.

We applied the NCA only to features calculated using other methods, i.e. standard HRV analysis with normalization and standardization, statistical distribution of instantaneous frequency based on Hilbert-Huang Transform (HHT) after applying empirical mode decomposition (EMD) and bivariate empirical mode decomposition (BEMD), while other SDR methods were subject to future works. The results were validated under subject-dependent and subject-independent scenarios using kNN as a classifier.

The paper has been organized in the following way: the first section gives brief introduction, including a gap in the previous research. Literature studies about supervised dimensionality reduction and research in emotion recognition follows it with the main research question appears at the end of this section. The next section discusses detail methods we used in this study, including a block diagram to explain the process visually. Succeeding this section, we present experimental results along with the discussions about the findings. The last section provides conclusions and some future works

2   Literature Studies.

2.1   Supervised Dimensionality Reduction

The SDRs use classes of the samples to guide the dimensionality reduction (DR) process such that distances among points belong to the same class are decreased while increasing the distances among points belong to different class. Some proposed algorithms were, e.g. Neighborhood Components Analysis (NCA) [3], Maximally Collapsing Metric Learning [4], Large Margin Nearest Neighbor (LMNN) [5], Supervised Dimensionality Mixture Model (SDR-MM) [6], Support Vector Decomposition Machine (SVDM) [7], etc.

As was mentioned in the Introduction section that this paper focused on NCA only, the following was a brief mathematical background about the NCA, as proposed by Goldberger et al. [3]. The NCA works based on Mahalonobis distance measure

$$\left\| f\left(\mathbf{x}_1\right) - f\left(\mathbf{x}_2\right) \right\|^2 = \left(\mathbf{x}_1 - \mathbf{x}_2\right)^T \mathbf{Q}\left(\mathbf{x}_1 - \mathbf{x}_2\right) \tag{1}$$

within kNN framework, where

$$\mathbf{Q} = \mathbf{A}^T \mathbf{A} \tag{2}$$

is a positive semidefinite (PSD) learning matrix to a certain space. The algorithm aims to find the projection matrix, $\mathbf{A}$, such that the classifiers perform well in the transformed space.

By maximizing a stochastic variant of the leave-one-out (LOO) kNN score on the training data, the NCA makes no assumption about the shape of the class distribution or the boundaries between them. Since the LOO classification errors of kNN suffers from discontinuity, a differentiable cost function based on stochastic ("soft") neighbor assignments in the transformed space was introduced,

$$p_{ij} = \frac{\exp\left(-\left\|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_j\right\|^2\right)}{\sum_{k \neq i}\exp\left(-\left\|\mathbf{A}\mathbf{x}_i - \mathbf{A}\mathbf{x}_k\right\|^2\right)}, \, p_{ii} = 0 \tag{3}$$

Eq. (3) assigns the probability of point i belongs to the class of selected point j, among k points as its neighbor. When point i chooses several neighbors and they might belong to different classes, total probability that point i belongs to class $C_i = \{ j \,|\, c_i = c_j \}$, is defined as

$$p_i = \sum_{j \in C_i} p_{ij} \tag{4}$$

The main idea is to maximize cost function

$$f(\mathbf{A}) = \sum_i \sum_{j \in C_i} p_{ij} = \sum_i p_i \tag{5}$$

The NCA has been implemented in the *drtoolbox*, a Matlab® toolbox for dimensionality reduction [8]. Experiments in this study used this toolbox after slightly modifying the algorithm, see section 3.2.

2.2 Literature Review

The MAHNOB-HCI database [9] is one of the affect recognition databases which includes ECG signals as one of the peripheral physiological signals. Other affect recognition databases which include ECG signals are RECOLA [10], Decaf [11], and Augsburg [12]. DEAP also provide signals from the heart activities but they were quantified as Heart Rate Variability (HRV) measured using Blood Volume Pulse (BVP) on finger [13]. In this paper, we use the MAHNOB-HCI database, which involved 27 subjects (11 males and 16 females) stimulated with pictures and video clips. The data includes the following synchronized signals:

- 32-channel EEG.
- Peripheral physiological signals (ECG, temperature, respiration, skin conductance).
- Face and body videos using 6 cameras.
- Eye gaze.
- Audio.

Many feature extraction methods for ECG-based emotion recognition have been proposed. HRV-based features using the standard HRV analysis were quite popular in many applications. This method requires at least 5 minute or even hours of ECG signal to get reliable analysis [14]. There are a number of large cross-sectional studies which suggest to use other methods, some of them are non-linear point-process [15], wavelet analysis [16], Recurrent Plot [17], and empirical mode decomposition-based techniques [18].

Ferdinando et al. [19] used standard HRV analysis to get features for emotion recognition in 3-class of valence and arousal, to provide baseline for the recognition using ECG signal only from the MAHNOB-HCI database. Using SVM classifier, the achieved accuracies were 43% and 48% for valence and arousal respectively based on 10-fold cross validation. The accuracies were slightly above chance level and close enough to the ones based on all peripheral physiological signals. There was no DR applied to the acquired features even for feature selection.

Apparently, the standard HRV analysis was not suitable for ECG signals from the MAHNOB-HCI database because the signal length varies from 35-117 seconds. Inspired by Agrafioti et al. [18], EMD and BEMD analysis were employed [2]. Using statistical distribution of the dominant frequencies estimated from spectrogram analysis after employing BEMD analysis to ECG as features, the achieved accuracies using kNN were 56% and 60% for valence and arousal respectively based on the subject-dependent scenario. Validated under subject-independent scenario, the accuracies were 60% and 59% for valence and arousal correspondingly. Features based on statistical distribution of instantaneous frequencies estimated using Hilbert-Huang Transform (HHT) achieved less than 50% of accuracies for both valence and arousal. The only DR technique used in these experiments was feature selection.

Using SDRs implemented in *drtoolbox* [8], NCA, MCML, and LMNN, performances of the system using statistical distribution of dominant frequency after applying BEMD analysis to ECG from the same database were enhanced [1]. The NCA outperformed the other method by improving the performance significantly from 56% to 64% and from 60% to 66.1% for valence and arousal respectively in subject-dependent scenario. Under subject-independent scenario, the enhancement only worked for arousal by improving the performance from 59% to 70%.

Although the NCA showed promising results [1], the evidences reviewed in this subsection seem to suggest evaluating how well the NCA can enhance the same system using different features, such as HRV- and HHT-based features. To our knowledge, no previous study has investigated the NCA, which was applied on the exactly same system but using features from different methods. This study can open new finding about phenomena in NCA related to different feature extraction methods as the main research question addressed in this paper.

3 Methods.

Fig. 1 shows the block diagram of our method. ECG signals used in these experiments were downloaded from the database server under "Selection of Emotion Elicitation"

group. Sample from session #2508 was discarded as the visual inspection showed that it was corrupted, leaving 512 samples for further process. All measured signals have a synchronization pulse to separate response and baseline signals. The non-stimulations or relaxation stages are 30 seconds before and after stimulation part, and they must be separated to each other. We applied signal pre-processing methods suggested by Soleymani et al. [9] to remove baseline wandering and power line interference.
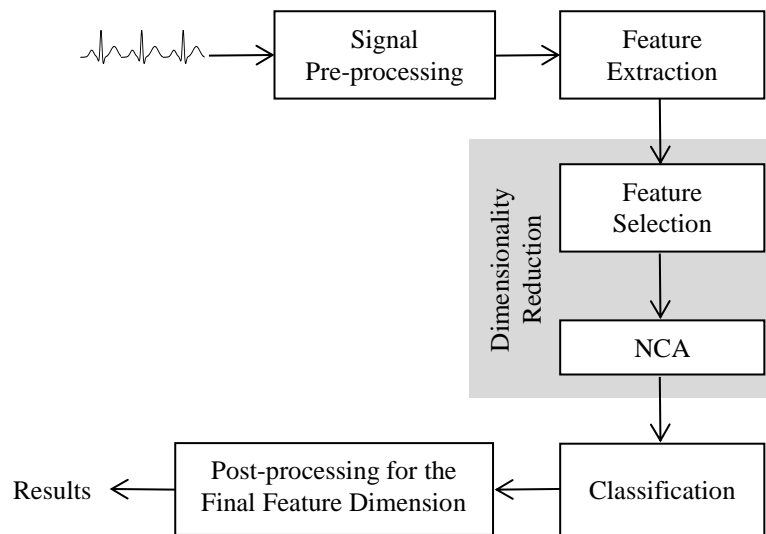

Fig. 1. Block diagram of the experiments

3.1 Feature Extraction

We used the standard HRV analysis to extract features from both baseline and response signal as suggested by Soleymani et al. [9], i.e.

- RMS of the Successive Difference between adjacent R-R intervals (RMSSD).
- Standard Deviation of the Successive Difference between adjacent R-R intervals (SDSD).
- Standard Deviation of all NN intervals (SDNN).
- Number of pairs of adjacent NN intervals differing by more than 50 ms (NN50).
- Number of pairs of adjacent NN intervals differing by more than 20 ms (NN20).
- NN50 count divided by the total number of NN intervals (pNN50).
- NN20 count divided by the total number of NN intervals (pNN20).
- Power spectral density for very low frequency (VLF), low frequency (LF), high frequency (HF), and total power.
- Ratio of HF to LF.
- Poincaré analysis (SD1 and SD2).
- Ratio of response to baseline features.

6

resulted 42 features: 14 from baseline, 14 from response, and 14 from the ratios of response to baseline. We normalized them to [-1,1] and standardized them based on mean and standard deviation to get two sets of features.

Another feature extraction method was based on the instantaneous frequency (IF) calculated using Hilbert-Huang Transform (HHT), see Eq. (6), from the intrinsic mode functions (IMFs) after either EMD or BEMD [2].

$$z(t) = d(t) + jH[d(t)]$$
$$z(t) = y(t)e^{j\theta(t)}$$
$$IF = \frac{1}{2\pi}\frac{d\theta(t)}{dt}$$
(6)

Specific to BEMD, a synthetic ECG signal, synchronized on R-wave event, was generated using a model developed by McSharry et al. [20], as the imaginary part of the complex ECG signals, while the original signal as the real part. Of note, the model only generated one cycle ECG signal as a template. By placing the template according to the R-wave event, a complex ECG signal was formed. This method was faster than generating one cycle ECG signal using the model for each detected R-wave event [2].

This method has two drawbacks, at least. First, the connection between consecutive ECG templates is not smooth but it can be minimized by adjusting the start and end of ECG template very close to zero. However, this discontinuity issue brings small problem if it is kept as small as possible. Second, the synthetic ECG may have different shape at the beginning of the signal because there is no such guarantee of getting a complete PQRST wave at the beginning of the signal, see Fig. 2. For this reason, 256 zeros were inserted at the beginning of the synthetic ECG and discarded them after the whole synthetic ECG was complete.
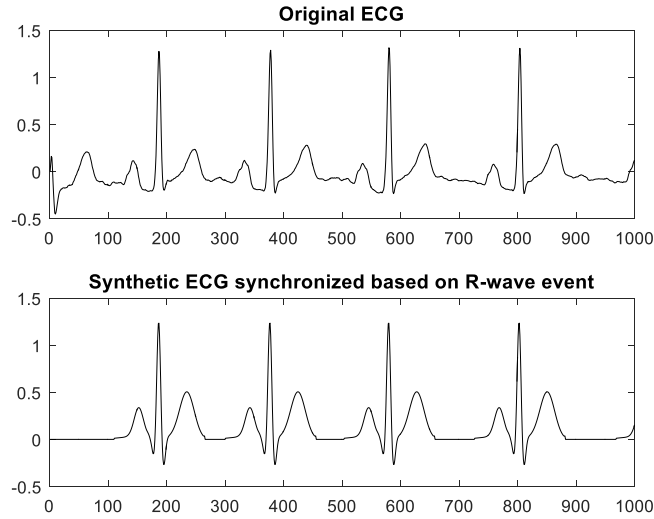

Fig. 2. Synchronized synthetic ECG signal with its original ECG signal [2].

Prior to applying EMD and BEMD, each samples was divided into 5-second segments, because both EMD and BEMD were sensitive to signal length, such they resulted 5-6 IMFs plus residue [18]. Besides, the ECGs in the MAHNOB-HCI database have different length. The IMFs from both EMD and BEMD were subject to HHT to obtain the instantaneous frequencies (IFs) from each segment. Once this process was finished, all five IFs from each segment belonging from the same ECG sample were joined to represent five IFs of that ECG signal. Following this step was to calculate 14 statistical distribution values, i.e. mean, standard deviation, median, Q1, Q3, IQR, skewness, kurtosis, percentile 2.5, percentile 10, percentile 90, percentile 97.5, maximum, and minimum, from IFs and the first differences as the features for classifier. Finally, we had another two groups of features and each group contained statistical distribution of IFs from one to five IMFs, resulting five different sets of features for each EMD and BEMD, see Table 1 for clarity. All acquired features were standardized based on mean and standard deviation (SD).

Table 1. Feature configuration prior to feature selection.

	HHT-based features after EMD	HHT-based features after BEMD
1 IMF	28 features	28 features
2 IMFs	56 features	56 features
3 IMFs	84 features	84 features
4 IMFs	112 features	112 features
5 IMFs	140 features	140 features

3.2   Dimensionality Reduction.

There were two DR processes prior to classification phase, i.e. feature selection using sequential forward floating search and the NCA. Feature selection is the simplest DR technique and requires no projection matrix. It only combines the available features into a new set, with reduced dimensionality, that offers the best performance.

Using the reduced dimensionality from the previous stage, the NCA algorithm was applied to find a projection matrix able to reduce feature dimensionality while enhancing the performance. The initial projection matrix in the *drtoolbox* was set using random number such that each process produced different results and there was no guarantee that the optimum projection matrix could be acquired within single pass. For this reason, the algorithm was modified to be iterative such that it stopped when there was no improvement, validated using leave-one-out, within 200 iterations. The SDR process was applied to the selected features having dimensionality higher than the target, 2D to 9D, as in the previous study [1]. The highest possible dimensionality target, however, was 37 but it was different for each set of features. In particular, the performance analysis of the reduced dimensionality was problematic due to this limitation but we had to keep it similar to the previous one for the sake of equal methodologies [1]. The acquired projection matrices were saved for further processes.

3.3   Classifier and Validation Method.

We used kNN as in the previous study [1] to make reliable comparisons before and after applying the NCA. kNN is one of the classifiers which gets benefits from DR because its reliability relies on the sample size. Using DR prior to building model with kNN saves more space for storage. Another reason related to kNN is about the computational speed. Using many samples to build a reliable model, kNN may suffer from the slow speed. This issue, however, was beyond the scope of this study because we only used 512 samples.

Results in this study were validated using subject-dependent and subject-independent scenarios. Within the subject-dependent scenario, 20% of the samples were held out for validation while the rest of them were subject to training and testing using 10-fold cross validation. The model was built based on the projection of the selected features using saved projection matrix from previous stage. The whole validation process was repeated 1000 times, with new resampling in each repetition, to accommodate the Law of Large Numbers (LLN) such that the average was close enough to the real value.

Subject-independent scenario evaluated if the features were ready for a general model where new samples were introduced to the classifier for recognition. Samples from one subject were excluded from building the model and used them to test the model. This process continued for all subjects and the reported performance was the average over all exclusion processes. We called this validation as Leave-One-Subject-Out (LOSO) validation.

3.4   Post-processing for the Final Feature Dimension

Validations tests were designed so that they produced classification accuracy with several dimensionalities to select the best one. However, small differences between the accuracies may not be statistically significant. Specifically, there can be two accuracies close to each other while the feature vector dimensions are different. It would make sense to choose the one that has a lower dimension. The following procedure was therefore used to choose the final feature vector dimension:

1. Find the best accuracy (namely, A1).
2. If the best accuracy is occurred at the lowest dimensionality, then the best result is found (best result = A1).
3. Otherwise, find the second-best accuracy (namely, A2) from the lower dimensionality and compare A1 to A2 using t-test with significance level 0.05.
4. If the difference is statistically significant, then the best results is found (best result = A1).
5. If the difference is not statistically significant, then the second-best turns to the best accuracy. Repeat process from step 2 until it reaches the lowest dimensionality.

4   Results and Discussions.

We experimented with four sets of features extracted using different methods and then compare the results side-by-side. We provided the baseline performances for each set of features, evaluated the difference using t-test with significance level 0.05, and calculated the improvement in percent to answer the main research question in this paper. Results from the previous study [1] were also presented.

4.1   HRV-based features

Table 2 shows the experiment results using standardized HRV-based features under subject-dependent scenario. The highest performances were 60% and 46% for valence 4D and arousal 3D respectively. Using the post-processing procedure for the final feature dimension in section 3, the second highest performance in valence with lower dimensionality was at 2D and p-value as the result of significance test was close to zero, indicating that the difference was significant such that 60.0 ± 4.4 was the best result occurred at the lowest dimensionality. Applying the same rules for arousal, we compared the one at 3D to 2D using t-test and resulted a very small p-value specifying that 46.0 ± 4.1 in 3D was better than the other. Performances improved about 17% and 6% for valence and arousal respectively.

Table 2. Results for standardized HRV-based features for subject-dependent scenario from each dimensionality.

	Baseline	2D	3D	4D	5D	6D
Valence	51.2 ± 4.2	57.8 ± 4.3	57.2 ± 4.2	$\mathbf{60.0 \pm 4.4}$	58.8 ± 4.4	59.4 ± 4.3
Arousal	43.3 ± 4.2	45.2 ± 4.3	$\mathbf{46.0 \pm 4.1}$	44.2 ± 4.3	43.8 ± 4.1	44.2 ± 4.4

Results from standardized HRV-based features within subject-independent scenario are presented in Table 3. For valence, the highest accuracy was at 4D but significance test against the one at 3D gave p-value 0.079 indicating that the difference was not significant such that result at 3D became the best one. Next, we compared result at 3D to 2D, the second highest result, and found that the difference was significant, brought 61.4 ± 4.0 at 3D as the best result. For arousal, comparing the highest performance at 3D, 42.8 ± 4.0, to the second highest one at 2D, 41.7 ± 4.2, emerged p-value 0.0866, such that result from 2D was chosen as the best result. Although the NCA worked well in valence, no evidence was found for improvement in arousal.

Table 3. Results for standardized HRV-based features for subject-independent scenario from each dimensionality.

	Baseline	2D	3D	4D	5D	6D
Valence	54.1 ± 11.3	55.2 ± 4.8	$\mathbf{61.4 \pm 4.0}$	62.8 ± 4.9	60.3 ± 4.6	61.5 ± 4.3
Arousal	44.5 ± 8.0	$\mathbf{41.7 \pm 4.2}$	42.8 ± 4.0	36.9 ± 4.2	36.9 ± 4.1	42.0 ± 4.7

The same procedures were applied to all sets of features and the results were summarized in Table 4-7 for valence and arousal within both scenarios. From the second column of those tables, it was evident that NCA improved all performances significantly except for arousal in subject-independent scenario and it was even lower than its baseline. If we compared results from arousal for both scenarios, the second column of Table 5 and 7, they were either similar or even worse than the ones in [19] although the later used neither feature selection nor NCA. It was also shown here that the improvements from the baseline was somehow small.

If we now turn to experiment with normalized HRV-based feature, the third column of Table 4-7, it was apparent that NCA improved all performances significantly as shown by p-values at the last row. Compare to standardized HRV-based features, improvements for normalized HRV-based features were considerably better than the other.

Surprisingly, experiments within subject-independent scenario showed that the SD reduced around 50% after applying NCA, indicating higher consistency among the repetitions than the ones in baselines, but not in subject-dependent scenario. These were unexpected as validation using this scenario usually resulted high variation.

These experiments also presented the fact that whether valence had higher accuracy than arousal and the other way around depended on the normalization method. This finding brought an idea to combine standardized and normalized HRV-based feature, select the most discriminant features, and then apply NCA to evaluate if this combination offers more powerful features than working individually. By combining these two sets of features, result for arousal under subject-independent scenario looked promising, see the fourth column of Table 4-7. Besides, the accuracies were even better than the ones when both sets of features worked individually. The selected features in this scheme were from both parties showing that combining these two set of features was a choice.

4.2   HHT-based features

Performances for recognition in valence and arousal using HHT-based feature after EMD analysis under both scenarios are presented in the fifth column of Table 4-7. Applying NCA to HHT-based feature after EMD analysis enhanced the performance for both valence and arousal in both scenarios significantly, indicated by p-values, with the largest enhancement occurred in arousal under subject-independent scenario, see Table 7.

The sixth column of Table 4-7 displays the summary of experiments using HHT-based feature after BEMD analysis. The improvements were large and enhanced the performance significantly as well. Of note, the baselines of HHT-based feature after BEMD were mostly the smallest among all experiments such that it offered the largest improvement.

Table 4. Compare side-by-side for valence in subject-dependent scenario.

	Standardized HRV-based features	Normalized HRV-based features	Combined standardized and normalized HRV-based feature	HHT-based features after EMD analysis	HHT-based features after BEMD analysis	Spectrogram-based features with BEMD analysis
Baseline	51.2 ± 4.2	48.6 ± 4.5	45.9 ± 4.4	50.5 ± 4.5	45.8 ± 4.3	55.8 ± 7.3
After NCA	60.0 ± 4.4	57.6 ± 4.4	67.1 ± 4.4	50.5 ± 4.5	**68.6 ± 4.4**	64.1 ± 7.4
Improvement	17%	19%	46%	19%	50%	15%
p-value	~0	~0	~0	~0	~0	~0

Table 5. Compare side-by-side for arousal in subject-dependent scenario.

	Standardized HRV-based features	Normalized HRV-based features	Combined standardized and normalized HRV-based feature	HHT-based features after EMD analysis	HHT-based features after BEMD analysis	Spectrogram-based features with BEMD analysis
Baseline	43.3 ± 4.2	53.3 ± 4.6	55.8 ± 4.6	50.2 ± 4.5	47.3 ± 4.1	59.7 ± 7.0
After NCA	46.0 ± 4.1	61.8 ± 4.5	**70.7 ± 4.3**	66.6 ± 4.5	64.1 ± 4.3	66.1 ± 7.4
Improvement	6%	16%	27%	21%	36%	11%
p-value	~0	~0	~0	~0	~0	~0

Table 6. Compare side-by-side for valence in subject-independent scenario.

	Standardized HRV-based features	Normalized HRV-based features	Combined standardized and normalized HRV-based feature	HHT-based features after EMD analysis	HHT-based features after BEMD analysis	Spectrogram-based features with BEMD analysis
Baseline	54.1 ± 11.3	50.0 ± 11.4	44.7 ± 10.4	48.3 ± 10.0	43.0 ± 9.5	59.2 ± 11.4
After NCA	61.4 ± 4.0	59.7 ± 4.5	**70.7 ± 4.9**	61.2 ± 11.9	62.8 ± 12.1	61.7 ± 14.1
Improvement	13%	19%	58%	27%	46%	4%
p-value	~0	~0	~0	~0	~0	0.1873

Table 7. Compare side-by-side for arousal in subject-independent scenario.

	Standardized HRV-based features	Normalized HRV-based features	Combined standardized and normalized HRV-based feature	HHT-based features after EMD analysis	HHT-based features after BEMD analysis	Spectrogram-based features with BEMD analysis
Baseline	44.5 ± 8.0	53.6 ± 9.7	55.7 ± 11.0	47.0 ± 13.5	44.1 ± 10.9	58.7 ± 9.1
After NCA	41.7 ± 4.2	62.7 ± 4.0	**73.5 ± 4.4**	65.7 ± 12.6	63.8 ± 11.9	69.6 ± 12.4
Improvement	-6%	17%	32%	40%	45%	18%
p-value	0.0432	~0	~0	~0	~0	~0

4.3   Summary of the experiments

Table 4-7 present comparisons side-by-side for each emotional label under subject-dependent and subject-independent scenarios. Generally, the NCA could improve the performances in both valence and arousal within both scenarios. An exception occurred for valence in subject-independent scenario using spectrogram-based features after BEMD [1], see the most right column of Table 6, and arousal in the same scenario using standardized feature, see the second column of Table 7, as the NCA failed to make it.

Although the baselines for different set of features had large differences, the results after applying NCA were quite good by neglecting the exceptional cases above. The lowest accuracies before NCA corresponded to the highest improvements and vice versa. These facts were interesting as the NCA could make the end results almost close to each other no matter the feature extraction method applied to ECG signals. To our knowledge, this interesting phenomenon has not been exposed before.

Related to computational cost, HRV-based feature extraction offered the lightest one with moderate performances after feature selection process and good results after applying NCA. On the other hand, a method utilizing spectrogram after BEMD had high cost because of the spectrogram analysis parameters, i.e. four values of window size and nine values of overlap parameters, and process related to BEMD [2]. Moreover, the feature selection process and NCA must search for all parameter combinations. Furthermore, the end results from spectrogram-based feature after BEMD analysis could not beat the ones from combined standardized and normalized HRV-based features although the former won the competition on the baseline.

Experiments under subject-independent scenario expect higher variances as the classifiers never learn the structure from training data [1, 2]. However, NCA was able to successfully reduce the SDs significantly for HRV-based features. Even more interesting, the NCA lowered them such that the values were close to the other scenario. There was no such improvement from the other feature extraction methods. This finding was also interesting but it needs more studies with other databases and feature extraction methods, and is left for future work.

5   Conclusions.

Enhancements of ECG-based emotion recognition on the MAHNOB-HCI database, processed by several feature extraction methods, using NCA were presented. Generally, NCA could successfully enhance the performance on this database significantly and provided new baselines. Results using combined standardized and normalized HRV-based features were superior, except for valence in subject-dependent scenario. Although spectrogram-based features after BEMD analysis outperformed the other feature extraction methods when NCA was not applied [2], the results were completely different after applying NCA as shown in Table 4-7.

Different feature extraction methods had different classifier performances but the NCA could make the results from different methods closer to each other. This fact was interesting to note because so far feature extraction methods were very critical. However, this observation needs more elaboration with other databases and feature extraction methods.

Spectrogram-based features after BEMD analysis had a heavy computational cost and the performances after NCA were not as good as in the baseline. On the other hand, HRV-based features had a light computational cost but offered better results after NCA. Having higher baseline accuracy brought no guarantee that applying NCA would result in as good improvement as with lower baseline levels.

The NCA reduced the SDs around 50% from the baseline on experiments using HRV-based features under subject-independent scenario. To our knowledge, such results have not been explored before in many experiments using NCA. However, confirmation using other databases but MAHNOB-HCI database remains as future work.

References

1. Ferdinando, H., Seppänen, T., Alasaarela, E.: Enhancing Emotion Recognition from ECG Signals using Supervised Dimensionality Reduction. In: Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM. pp. 112–118. Scitepress, Porto, Portugal (2017).

2. Ferdinando, H., Seppanen, T., Alasaarela, E.: Comparing features from ECG pattern and HRV analysis for emotion recognition system. In: 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). pp. 1–6. IEEE, Chiang Mai, Thailand (2016).

3. Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood Components Analysis. Adv. Neural Inf. Process. Syst. 17. 1–8 (2004).

4. Globerson, A., Roweis, S.: Metric Learning by Collapsing Classes. Neural Inf. Process. Syst. 18, 451–458 (2005).

5. Weinberger, K.Q., Saul, L.K.: Distance Metric Learning for Large Margin Nearest Neighbor Classification. J. Mach. Learn. Res. 10, 207–244 (2009).

6. Orlitsky, A.: Supervised dimensionality reduction using mixture models. In: Proceedings of the 22nd international conference on Machine learning (ICML). pp. 768–775. ACM, Bonn, Germany (2005).

7. Pereira, F., Gordon, G.: The Support Vector Decomposition Machine. In: Proceedings of the 23rd international conference on Machine learning (ICML). pp. 689–696. ACM, Pittsburgh, Pennsylvania, USA (2006).

8. van der Maaten, L.: Matlab Toolbox for Dimensionality Reduction – Laurens van der Maaten, https://lvdmaaten.github.io/drtoolbox/.

9. Soleymani, M., Lichtenauer, J., Pun, T., Pantic, M.: A Multimodal Database for Affect Recognition and Implicit Tagging. IEEE Trans. Affect. Comput. 3, 42–55 (2012).

10. Ringeval, F., Sonderegger, A., Sauer, J., Lalanne, D.: Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture

Recognition, FG 2013 (2013).

11. Abadi, M.K., Subramanian, R., Kia, S.M., Avesani, P., Patras, I., Sebe, N.: DECAF: MEG-Based Multimodal Database for Decoding Affective Physiological Responses. IEEE Trans. Affect. Comput. 6, (2015).

12. Wagner, J., Jonghwa Kim, Andre, E.: From Physiological Signals to Emotions: Implementing and Comparing Selected Methods for Feature Extraction and Classification. In: 2005 IEEE International Conference on Multimedia and Expo. pp. 940–943. IEEE.

13. Koelstra, S., Muhl, C., Soleymani, M., Jong-Seok Lee, Yazdani, A., Ebrahimi, T., Pun, T., Nijholt, A., Patras, I.: DEAP: A Database for Emotion Analysis Using Physiological Signals. IEEE Trans. Affect. Comput. 3, 18–31 (2012).

14. Task-force: Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur. Heart J. 17, 354–381 (1996).

15. Valenza, G., Citi, L., Lanatá, A., Scilingo, E.P., Barbieri, R.: Revealing real-time emotional responses: a personalized assessment based on heartbeat dynamics. Sci. Rep. 4, 4998 (2014).

16. Konar, A., Chakraborty, A.: Emotion recognition : a pattern analysis approach. John Wiley and Sons, Inc., Hoboken, New Jersey (2015).

17. Valenza, G., Lanata, A., Scilingo, E.P.: The Role of Nonlinear Dynamics in Affective Valence and Arousal Recognition. IEEE Trans. Affect. Comput. 3, 237–249 (2012).

18. Agrafioti, F., Hatzinakos, D., Anderson, A.K.: ECG Pattern Analysis for Emotion Detection. IEEE Trans. Affect. Comput. 3, 102–115 (2012).

19. Ferdinando, H., Ye, L., Seppänen, T., Alasaarela, E.: Emotion Recognition by Heart Rate Variability. Aust. J. Basic Appl. Sci. Aust. J. Basic Appl. Sci. 8, 50–55 (2014).

20. McSharry, P.E., Clifford, G.D., Tarassenko, L., Smith, L.A.: A Dynamical Model for Generating Synthetic Electrocardiogram Signals. IEEE Trans. Biomed. Eng. 50, 289–294 (2003).