❐   139

Performance of Principal Component Analysis and Orthogonal Least Square on Optimized Feature Set in Classifying Asphyxiated Infant Cry Using Support Vector Machine

R. Sahak[1], W. Mansor*[2], Khuan Y. Lee[3], A. Zabidi[4]
[1,2,3,4]Faculty of Electrical Engineering, Universiti Teknologi MARA 40450 Shah Alam, Selangor, Malaysia
[2,3]Computational Intelligence Detection RIG, PLS CORE, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Article Info	ABSTRACT
	An investigation into optimized support vector machine (SVM) integrated with principal component analysis (PCA) and orthogonal least square (OLS) in classifying asphyxiated infant cry was performed in this study. Three approaches were used in the classification; SVM, PCA-SVM, and OLS-SVM. Various numbers of features extracted from Mel-frequency Cepstral coefficient (MFCC) were tested to obtain the optimal parameters of SVM kernels. Once the optimal feature set is obtained, PCA and OLS selected the most significant features and the optimized SVM then classified the selected cry patterns. In PCA-SVM, eigenvalue-one-criterion (EOC), cumulative percentage variance (CPV) and the Scree test (SCREE) were used to select the most significant features. SVM with radial basis function (RBF) kernel was chosen in the classification stage. The classification accuracy and computation time were computed to evaluate the performance of each method. The best method for classifying asphyxiated infant cry is PCA-SVM with EOC since it produces the highest classification accuracy which is 94.84%. Using PCA-SVM, the classification process was performed in 1.98s only. The results also show that employing feature selection techniques could enhance the classifier performance.

Corresponding Author:

W. Mansor,
Faculty of Electrical Engineering,
Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.
Email: wahidah231@salam.uitm.edu.my

1. INTRODUCTION

Asphyxia is referred to respiratory failure, a condition caused by inadequate intake of oxygen [1]. It is important to diagnose asphyxia in infants at early birth since it is a major cause of infant morbidity. If misappropriate treatment is given to the infant, hypoxia will result, which could lead to serious complication, such as damage to infant's brain, organs, tissues or even fatality.

Asphyxia occurs in infants with neurological level disturbance, which is found to affect the sound of the cry produced by the infants [1]. The effect causes the cry signals to have distinct patterns compared to healthy infant cry. These have been proven by previous studies [2]. The researchers successful distinguished between the healthy and asphyxiated infant cry using Computer-Based analysis.

Basically, Computer-Based analysis contains three components; pre-processing, feature extraction and pattern classification. The cry signals were pre-processed first before cry features extracted in feature extraction stage. The popular method used for extracting features from cry signals is Mel-frequency Cepstral Coefficient (MFCC) [3-5]. Once the features are extracted, pattern classifier will classify the cry patterns according to types of cry. However, recent researchers appended a stage after the feature extraction stage,

called feature selection [6-8]. Normally, in feature extraction, the extracted features have a large dimension that sometimes encompasses less significant and redundant features. Hence, feature selection method is required to select the most significance cry feature which then enhances the classification accuracy and computation time as well.

In this study, orthogonal least square (OLS) and principal component analysis (PCA) have been employed to select the most significant cry features which extracted from the analysis of MFCC. OLS has shown to be able to select significant features from MFCC successfully in previous studies [9-11]. Even though PCA has been used in the previous infant cry analysis, application of PCA on both normal and asphyxiated cry only has not been investigated. Furthermore, different approaches of PCA selections known as eigenvalue-one-criterion (EOC), cumulative percentage variance (CPV) and scree test (SCREE) have not been examined yet in the previous study. To classify infant cry signals, a capable support vector machine (SVM) is used. SVM performs classification tasks based on the principle of binary classification. SVM offers more advantages such as globally optimal, small sample-size, good generalization ability and resistant to the over-fitting problem, than other classifiers [12-15].

This paper describes the asphyxiated infant cry classification using SVM with radial basis function (RBF) kernel. Three approaches known as SVM, PCA-SVM and OLS-SVM were used in the classification. The optimization process is carried out to obtain optimal parameters. The optimization of input feature set extracted from MFCC analysis is performed first in the SVM approach. The optimized feature set obtained is then used as a reference input to OLS-SVM and PCA-SVM approaches. Classification accuracy and computation time were computed and compared to achieve the best classifier performance.

2.   RESEARCH METHOD

The database of normal and asphyxiated infant cry in this study is obtained from the University of Milano–Bicocca [16]. The whole process of classifying asphyxiated infant cry is shown in Figure 1. Initially, pre-processing was carried out where the signals were normalized and sampled at eight kHz and pre-emphasized before dividing them into segments of one second each. From this process, 316 segments of normal cries and 284 of asphyxiated cries were generated.

After pre-processing, MFCC analysis was carried out where the segmented signals were first multiplied with a Hamming window with a width of 25 ms and an overlapped of 50% between successive frames. The output signal was then processed by Fast Fourier Transform (FFT). The resulted spectrum was processed through triangular filter banks to transform it into a Mel-spectrum. Finally, DCT was applied to produce the corresponding coefficients. In this study, ten to 20 coefficients were produced from various numbers of filter banks ranged between 20 and 30 which formed121 feature sets.

There were three methods used in the classification of asphyxiated infant cry. The first method called SVM was performed to obtain the optimal feature set. SVM performs classification task using hyperplane to differentiate between two cases. Thus, it is important to construct an optimal separating hyperplane. The hyperplane is said to be optimal when there is a large distance between hyperplane and data point. By choosing appropriate regularization parameter (C) and kernel function, an optimal separating hyperplane could be obtained [17, 18]. The kernel function is used to transform data points into high dimensional space so that the data can be linearly separated. The hyperplane is then built into these transformed data based on the value of C that has to be selected properly by a user. If C is too large, it may overfit the data and if it too small, it may underfit the data.

The decision function can be expressed as:

$$y(\alpha) = sign\left[\sum_{i=1}^{N} y_i \rho_i K(\alpha_i, \alpha_j) + q\right] \tag{1}$$

where N is the total of support vector number, ρ_i is a Lagrangian multiplier, α_i is a component vector of α and y_i is the label associated with α_i; +1 or -1, α_j is support vector, q is the hyperplane and $K(\alpha_i, \alpha_j)$ is the kernel function.

Figure 1. The process of classifying asphyxiated infant cry

The kernel function can be expressed as:

$$K(\alpha_i, \alpha_j) = \varphi(\alpha_i)^T \varphi(\alpha_j) \tag{2}$$

The common kernel used is radial basis function (RBF) which can be expressed as:

$$K(\alpha_i, \alpha_j) = \exp\left(\gamma \|\alpha_i - \alpha_j\|^2\right) \tag{3}$$

where γ is the kernel width.

In SVM approach, the 121 feature sets extracted from the MFCC analysis were passed through SVM with RBF kernels. In the second method called PCA-SVM, once the optimal feature set was obtained, PCA was employed to the optimized feature set. In PCA-SVM, EOC, CPV, and SCREE algorithms were applied to the optimal feature set. In another method, OLS was applied to the optimized feature set and this approach is called OLS-SVM.

In order to obtain an optimal model of SVM with the polynomial kernel, C was varied from 0.000001 to 0.01 and d was varied from two to four. For SVM with RBF kernel, C and γ were varied from 0.01 to 100 and 0.001 to 0.05 respectively. The range of C, d, and γ was selected based on the previous

studies and also from the current experiments. For reliability results,10-cross-validation was used to separate between training and test datasets.

3.   RESULTS AND ANALYSIS

Figure 2 shows the variation of classification accuracy of SVM with RBF kernel (C = 1 and γ = 0.009) obtained when the numbers of coefficient and filter bank were varied from ten to 20 and 20 to 30 respectively. It is observed that feature set that consists of 20 coefficients gives poor classification accuracy which shows there are many repetitive and fewer significance features. The highest classification accuracy is 93.84% when ten coefficients with 22 and 25 filter banks are used. So, feature set generated from ten coefficients and 22 filter banks is chosen as the best criterion for developing the optimal feature set since this combination produces small features.
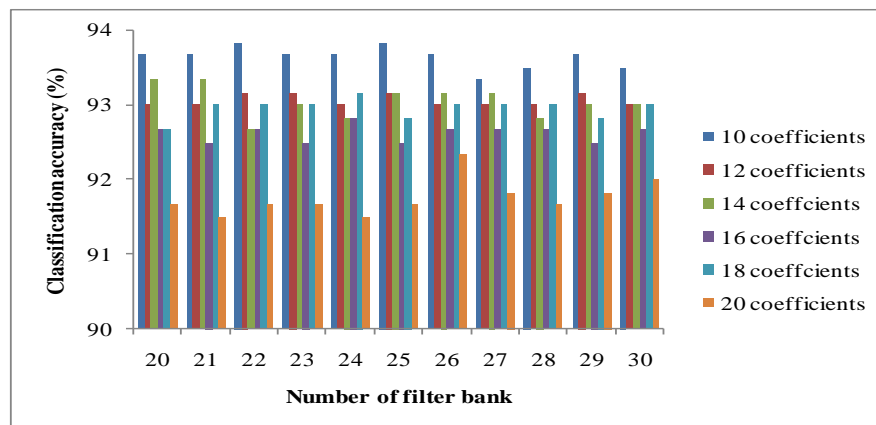

Figure 2. Classification accuracy of SVM with RBF kernel for various coefficients and filter bank numbers

In PCA-SVM approach using CPV, 54 principal components are retained as the principal components have about 80% of the cumulative percent of the variance. Using CPV, the dimension of the input feature is reduced from [1240 x 600] to [54 x 600]. When the features are selected by EOC, only 23 principal components are retained since they have eigenvalue greater than one. Note that with EOC, the original dimension is reduced from [1240 x 600] to [23 x 600]. In Scree test, only three principal components were retained as they have the large gap between each other as shown in Figure 3. With this method, the dimension of input feature is reduced from [1240 x 600] to [3 x 600].
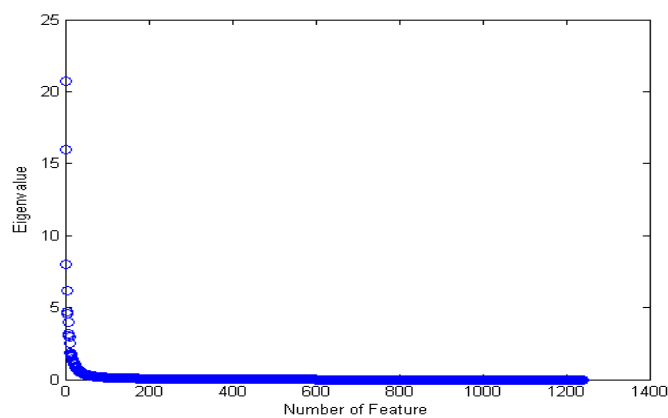

Figure 3. Scree plot of input features for SVM with RBF kernel

Figure 4 shows classification accuracy of PCA-SVM with RBF kernel when C increases from 0.01 to 100 for γ equal to 0.025. As shown in the figure, the highest classification accuracy is obtained when $C = 1$ using EOC algorithm. The rate of increment in classification accuracy is influenced by the dimension of the input feature vector.
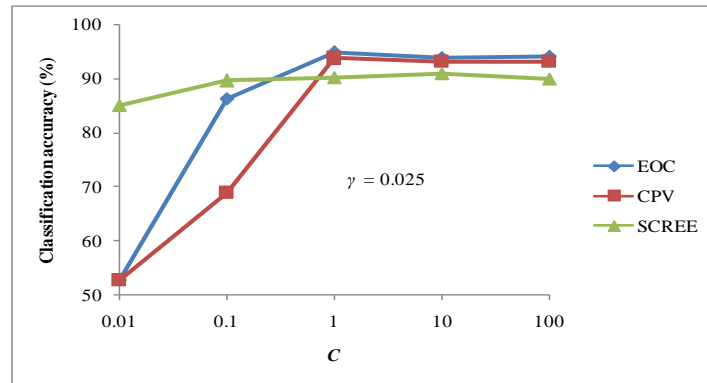

Figure 4. Effect of varying C on the classification accuracy of SVM with RBF kernel combined with EOC, CPV, and SCREE

Changes in the classification accuracy of EOC, CPV and SCREE, when combined with SVM with RBF kernel as γ is varied at the optimal C (which is one) is shown in Figure 5. Note that increasing γ does not yield good classification accuracy for EOC, CPV, and SCREE. The optimal γ is obtained at 0.025 since it produces the highest classification accuracy (94.84%). This optimal value is obtained when the EOC selection is employed. When SCREE is used, the worst classification accuracy is obtained.
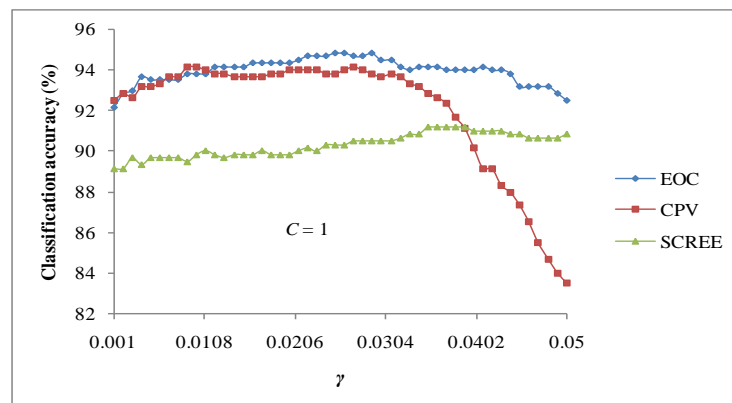

Figure 5. Classification accuracy of SVM with RBF kernel when γ is varied using EOC, CPV, and SCREE

The ten coefficients ranked with the use of OLS algorithm is shown in Table 1. Coefficient with the highest ERR is arranged at the top of the list whereas coefficient with the least ERR is placed at the bottom of the list. The fourth, first, third, seventh, fifth and tenth coefficients contain significant information.

In the optimization process, the optimal C is one for all methods. However, different optimal values of γ are found for SVM (0.009), PCA-SVM (0.025) and OLS-SVM (0.007). The highest classification accuracy (94.84%) is obtained from PCA-SVM (EOC), followed by SVM (93.84%) as shown in Figure 6. However, OLS-SVM (eight coefficients) provides comparable accuracy (93.34%) to SVM since the difference between their classification accuracies is about 0.503% only. It can be concluded that the feature selection algorithms affect the classification accuracy based on the fact that the lowest support vector produces the highest classification accuracy.

Table 1. Ranking of coefficients using OLS

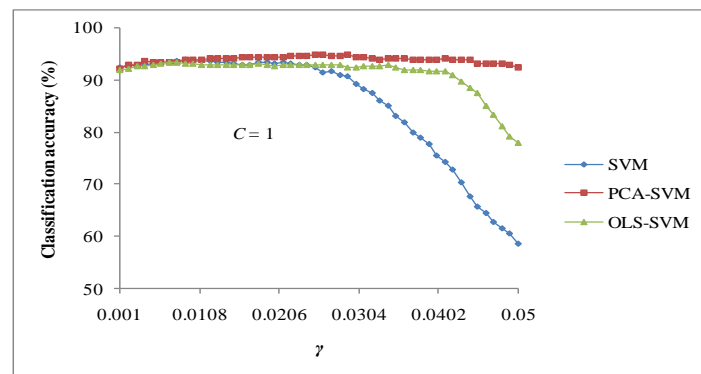| Rank	Coefficient Number	ERR
1	4	0.178
2	1	0.161
3	3	0.155
4	7	0.020
5	5	0.020
6	10	0.011
7	6	0.011
8	2	0.009
9	8	0.007
10	9	0.005

Figure 6. Classification accuracy of SVM, PCA-SVM, and OLS-SVM with RBF
kernel for various γ

Table 2 shows the accuracy and computation time between SVM, PCA-SVM, and OLS-SVM with RBF kernel in classifying cry patterns. The PCA-SVM takes only 1.9 s which is the shortest time to classify the infant cries. The computation time of OLS-SVM is shorter than SVM. It can be concluded that combining feature selection algorithms with SVM improves the classification accuracy and computation time.

Table 2. Classification accuracy and computation Time of SVM, PCA-SVM,
and OLS-SVM with RBF Kernel

Methods	Classification Accuracy	Computation Time (s)
SVM	93.84%	62.877
PCA-SVM	93.84%	1.982
OLS-SVM	93.34%	36.894

4.   CONCLUSION

An investigation into the classification performance of three approaches, SVM, PCA-SVM and OLS-SVM in classifying asphyxiated infant cries with SVM combined with feature selection techniques has been discussed. RBF kernel was employed and optimization process was carried out to obtain optimal parameters.

For the first approach, which is SVM, the optimal feature set was generated from ten coefficients with 22 filter banks. The highest classification accuracy is 93.84% for RBF kernel. In the second approach, which is PCA-SVM, the highest classification accuracy for RBF kernel is 94.84% when EOC selection was employed on the optimized feature set.

In OLS-SVM approach, the highest classification accuracy achieved is 93.34% and the highest classification was achieved when eight coefficients that selected from OLS analysis was employed. For the

results obtained in this study, it can be shown that the classification accuracy is obtained within the shortest computation time after OLS and PCA were employed. The RBF kernel is suitable for classifying asphyxiated infant cry.

REFERENCES

[1] Laberge M. Infancy through Adolescence Gale. *Farmington Hills*, MI: Gale Group, Thomson Gale, 2006.
[2] Lederman D. Automatic Classification of Infants' Cry. M.Sc. Degree, Ben-Gurion University of the Negev, 2002.
[3] Bănică I, Cucu H, Buzo A, Burileanu D, Burileanu C. *Automatic Methods for Infant Cry Classification*, 2016 International Conference on Communications (COMM), 2016; 51 – 54.
[4] Srijiranon K, Eiamkanitchat N. *Application of Neuro-fuzzy approaches to recognition and classification of infant cry*, IEEE Region 10 Conference, 2014; 1 – 6.
[5] Rosita YD, Junaedi H. *Infant's cry sound classification using Mel-Frequency Cepstrum Coefficients feature extraction and Backpropagation Neural Network*. 2nd International Conference on Science and Technology-Computer (ICST), 2016; 160 – 166.
[6] *Uyun S*. Feature Selection Mammogram Based on Breast Cancer Mining, *Indonesian Journal of Electrical and Computer Engineering*, 2018;8 (1): Accepted for publication.
[7] Salmam F Z, Madani A, Kissi M. Emotion recognition from facial expression based on fiducial points detection and using Neural Network, *Indonesian Journal of Electrical and Computer Engineering*, 2018;8 (1): Accepted for publication.
[8] Fauzi M A, Zainal Arifin A, Gosaria S C. Indonesian News Classification Using Naïve Bayes and Two-Phase Feature Selection Model. *Indonesian Journal of Electrical and Computer Engineering*, 2017, 8 (3): Accepted for publication.
[9] Sahak R, Mansor W, Lee Y K, Yassin A I M, Zabidi A. *An Orthogonal Least Square Approach to Select Features of Infant Cry with Asphyxia*. 6th International Colloquium on Signal Processing and Its Applications (CSPA), 2010; 1-4.
[10] Sahak R, Mansor W, Lee Y K, Yassin A I M, Zabidi A. *Orthogonal Least Square Based Support Vector Machine for the Classification of Infant Cry with Asphyxia*. International Conference on Biomedical Engineering and Informatics, 2010; 986-990.
[11] Sahak R, Mansor W, Lee Y K, Yassin A I M, Zabidi A. *Optimized Support Vector Machine for Classifying Infant Cries with Asphyxia using Orthogonal Least Square*. International Conference on Computer Applications & Industrial Electronics, 2010; 692-696.
[12] Zhou S, Wu L, Yuan X, Tan W. *Parameters Selection of SVM for Function Approximation Based on Differential Evolution*. International Conference on Intelligent Systems and Knowledge Engineering, 2007; 7.
[13] Vapnik V. An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*. 1999; 5: 988-99.
[14] Chan W C, Cheung K C, Harris C J. On the Modelling of Nonlinear Dynamic System Using Support Vector Neural Networks. *Engineering Applications of Artificial Intelligence,* vol. 14, 105-113.
[15] Orozco J & C A, Reyes-Garcia C A. Detecting Pathologies from Infant Cry Applying Scaled Conjugate Gradient Neural Networks. *Proceedings of the 2003 European Symposium on Artificial Neural Network,* 349-354.
[16] Reyes-Garcia, O F. [online database]. Available: *http://ingenieria.uatx.mx/~orionfrg/cry/*. 2009.
[17] Wang, W J, Xu Z B, Lu W Z. Determination of the Spread Parameter in the Gaussian Kernel for Classification and Regression. *Neurocomputing*. 2003; 55: 643-663.
[18] Zhu G Q, Liu S R & Yu J S. Support Vector Machine and Its Applications to Function Approximation. *Journal of East China University of Science and Technology*. 2002; 5: 555-559.