

Classification of Cerebral Blood Flow Oscillation

Balázs Benyó

Department of Informatics, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary, E-mail: benyo@sze.hu

Péter Somogyi

Department of Control Engineering and Information Technology, Budapest University of Technology and Economics, Magyar tudósok krt. 2, H-1117 Budapest, Hungary, E-mail: psomogyi@bio.iit.bme.hu

Béla Paláncz

Department of Photogrammetry and Geoinformatics, Budapest University of Technology and Economics, Műegyetem rakpart 3, H-1111 Budapest, Hungary, E-mail: palancz@epito.bme.hu

Abstract: Cerebral blood flow (CBF) oscillation is a common feature of several physiological and pathophysiological states of the brains. In this study the characterization of the temporal pattern of the cerebral circulation has been analyzed. The classification of CBF signals has been carried out by two different classification methods – neural network and support vector machine – employing spectral and wavelet analysis as feature extraction techniques. The efficiency of these classification and feature extraction methods are evaluated and compared. Computations were carried out with Mathematica and its Wavelet as well as Neural Networks Applications.

Keywords: Biomedical Systems, Classification, Neural Network models, Radial base function networks, Support Vector Machine

1 Introduction

Low frequency spontaneous oscillations in cerebral hemodynamics have been observed – and linked to certain physiological and patophysiological states [1],

such as epilepsy [2]. Therefore it is worthwhile to investigate the possibilities of classification of the temporal patterns of this vasomotion. Three classes of CBF signals have been distinguished experimentally [3, 4, 5] in relation to consecutive administration of two different drugs (see Figure 1):

- 1 Normal blood flow signals before applying any drugs, that does not exhibit low frequency oscillations (LFO-s), referenced as class A;
- 2 Slight oscillation after the administration of L-NAME, a NO synthase inhibitor reportedly evoking CBF oscillations, referenced as class B;
- 3 More pronounced oscillation observed after the administration of U-46619 for stimulating thromboxane receptors, having the effect of also inducing LFO, referenced as class C.

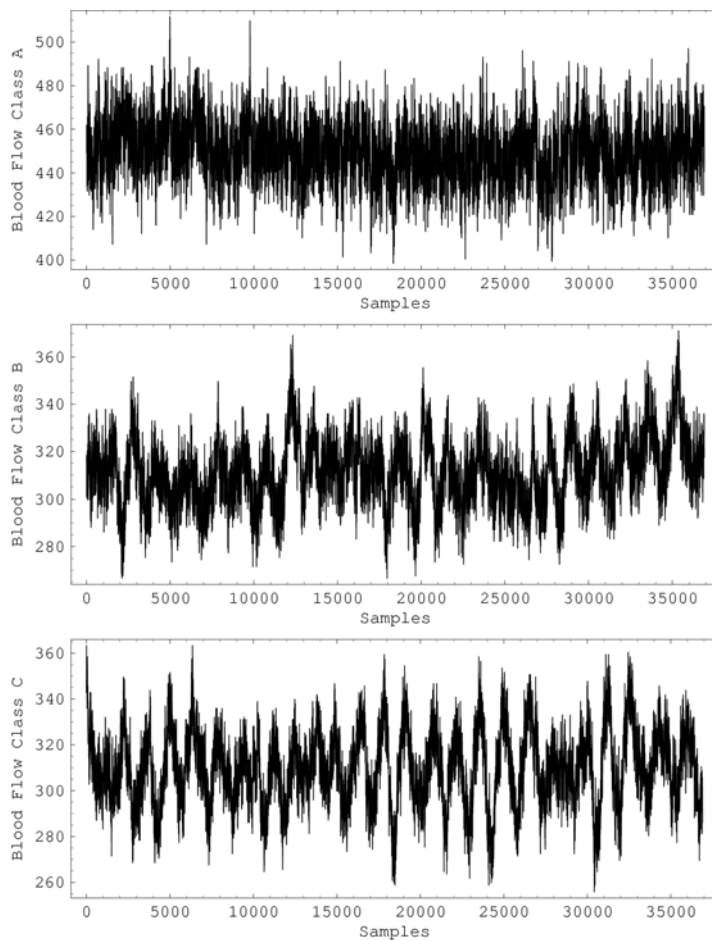


Figure 1

Measured blood flow series of the three states

Recently, to identify different states of CBF oscillation, different classification methods, based on a two-dimensional feature vector – the maximum amplitude and its frequency of the Fourier transform of the time signals – have been employed, using neural network and support vector machine classifiers (SVMC) [6, 7]. However, these approaches were only partly successful because the two-dimensional feature vector could not characterize all the features of the time series. Even the most promising technique, the SVMC suffered from overlearning [8]. The separation of the first class from the two latter has been carried out successfully using two feature vector elements derived from the measured signal. However, the second and third classes cannot be effectively distinguished due to the highly overlapping regions (stars and squares), as seen on the feature map Figure 2. Hence the discrimination of the mentioned classes, or cerebral blood flow states, is the subject of this paper. Two different feature extraction methods have been applied to characterize the given time signals, based on spectral and wavelet subband analysis.

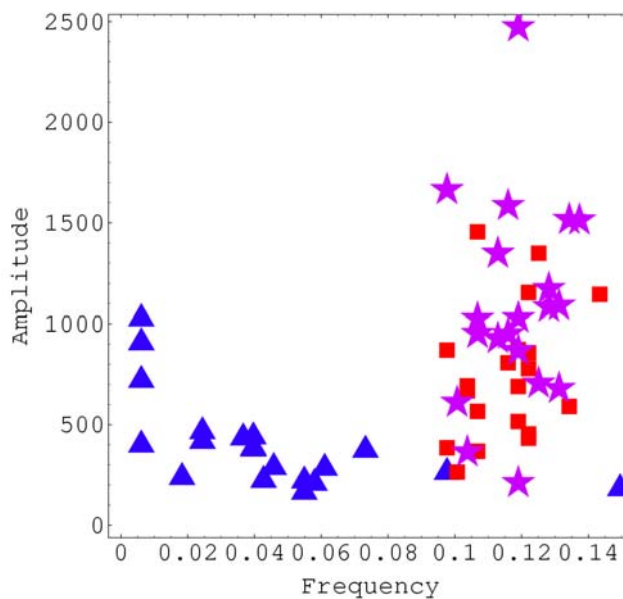


Figure 2

Normalized dimensionless feature map of cerebral blood flow: normal blood flow, class A (triangle), before administration of U-46619, class B (square) and after administration of U-46619, class C (star) from [4]

2 Feature Extraction

2.1 Using Spectral Analysis

In this case, the characterization of the signals to be classified is based on the eigenvalue of the spectral matrix of the signal [9]. In order to obtain the singular values being characteristic of the different states, a matrix has to be derived from the time signal. This is obtained by creating a spectral matrix. Given the time series of data d_i , where $i = [1 \dots 70.000]$ are the sample points, we pick a window size of $n \ll 70.000$ and form $70.000 - n$ window vectors, which we apply to a given range of data points:

$$\begin{aligned}
 \underline{u}_1 &= (d_1, d_2, d_3, d_4, d_5, d_6, \dots, d_n) \\
 \underline{u}_2 &= (d_2, d_3, d_4, d_5, d_6, d_7, \dots, d_n) \\
 &\vdots \\
 \underline{u}_j &= (d_j, d_{j+1}, d_{j+2}, \dots, d_{j+n-1})
 \end{aligned} \tag{1}$$

The matrix is built from these window vectors as columns:

$$\underline{\underline{A}} = [\underline{u}_1^T \ \underline{u}_2^T \ \dots \ \underline{u}_j^T] \tag{2}$$

then our spectral matrix can be computed as $\underline{\underline{S}} = \underline{\underline{A}}^T \underline{\underline{A}}$. In order to find the optimal window size and range, a series of decompositions have been completed, and the reconstructed signals have been compared to the original recordings. A sample of the maximal reconstruction errors can be seen in Figure 3, showing the local minimum and maximum. There is a few percent difference between the local minimum and maximum, therefore for the feature extraction, a window size of 50 and a window range of 5000 samples has been selected.

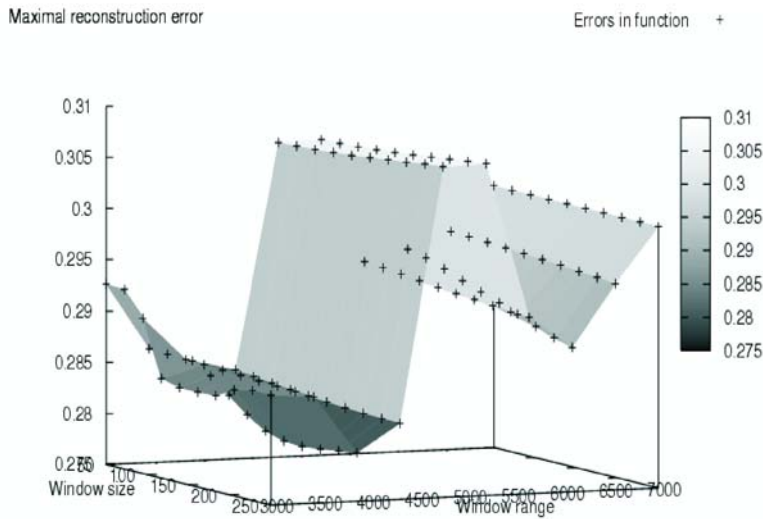


Figure 3

Approximation error in function of window size and range

Employing these window parameters, the eigenvalues of the spectral matrix can be computed. As it can be seen, in the case of a class C signal on Figure 4, the first six values are good candidates to be the elements of the feature vector describing a given signal.

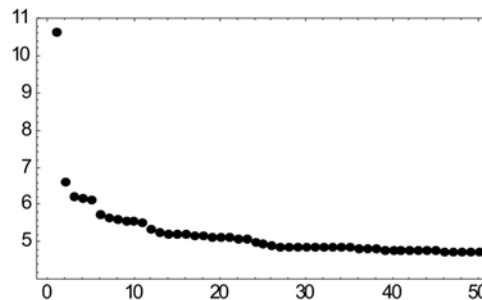


Figure 4

Eigenvalues of the spectral matrix of a class C signal, on a logarithmic scale

2.2 Feature Extraction via Wavelet Transformation

In recent years, feature extraction methods were developed based on wavelet transformation to recognize acoustic signals. They are applicable to the recognition of ships from sonar signatures, cardiopulmonary diagnostics from

heart sounds, safety warnings and noise suppression in factories, and recognition of different types of bearing faults in the wheels of railroad cars and so on [10].

Wavelet-based analysis is similar to the Fourier analysis where sinusoids are chosen as the basis function. The Wavelet analysis is also based on a decomposition of a signal using an orthogonal family of basis functions. Because of the used basis function the wavelets are well suited for the analysis of transient, time-varying signals. The wavelet expansion is defined by a two-parameter family of functions:

$$f(t) = \sum_k \sum_j a_{j,k} \Psi_{j,k}(t) \quad (3)$$

where j and k are integers, the $\Psi_{j,k}(t)$ are the wavelet expansion functions.

The wavelet expansion (or basis) functions based on the mother wavelet of the formula

$$\Psi_{j,k}(t) = 2^{j/2} \Psi(2^j t - k) \quad (4)$$

where j is the translation parameter and k is the dilation parameter.

The expansion coefficients $a_{j,k}$ are called the discrete wavelet transform (DWT) coefficients of $f(t)$. The coefficients are given by the following equation

$$a_{j,k} = \int f(t) \Psi_{j,k}(t) dt \quad (5)$$

The DFT and the DWT are the two most commonly used techniques for the signal transformation to the frequency domain. More details of the transforms can be found in [11, 12].

Let us illustrate this classical technique applying it to a CBF signal. Before the DWT of the time signal can be computed, we drop the beginning and the end of this raw signal, getting a signal of 2^{16} length of samples. This transformation decomposes the data into a set of coefficients in the wavelet basis. There are 16 sublists containing the wavelet coefficients in the orthogonal basis of the orthogonal subspaces. The contributions of the coefficients to the signal at different scales are represented by the phase space plot, see Figure 5. Each rectangle is shaded according to the value of the corresponding coefficient: the bigger the absolute value of the coefficient, the darker the area. The time unit is 5 msec.

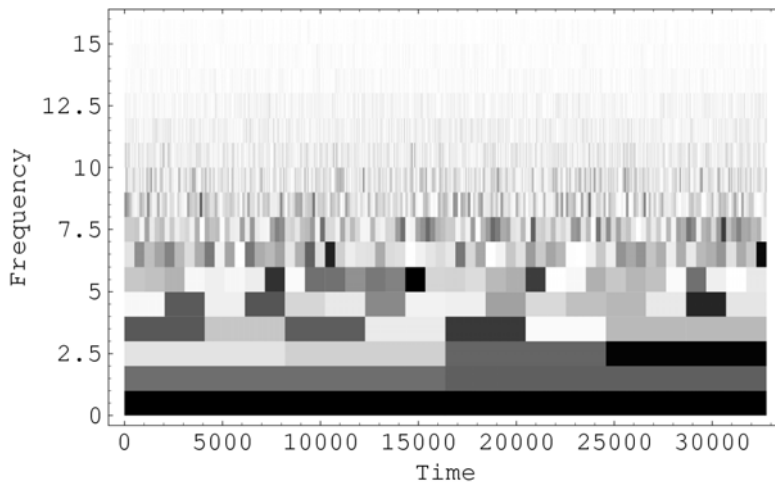


Figure 5

The phase space plot of the DWT of the time signal

Normally, from the wavelet coefficients of each of the 16 resolution levels (subbands) and from sample values of the original time signal, one computes the average energy content of the coefficients at each resolution. There are a total of 17 subbands (16 wavelet subbands and one approximation subband represented by the original signal) from which features are extracted. The i^{th} element of the feature vector is given by,

$$v_i = \frac{1}{n_i} \sum_{j=1}^{n_i} w_{i,j}^2, i = 1, 2, \dots, 17 \quad (6)$$

where $n_1 = 2, n_2 = 2, n_3 = 2^2, \dots, n_{16} = 2^{15}$ and $n_{17} = 2^{16}$, where $w_{i,j}$ is the j^{th} coefficient of the i^{th} subband. In this way, from a time signal having 2^k samples or dimensions, one can extract a feature vector of $k + 1$ dimensions. This technique has been extended for two dimensional signals, for digital images [13]. In order to study the effect of the dimension of the input space on the quality of the classification as well as to save the morphology of DWT, here we employ a different approach. We consider the wavelet coefficients belonging to a given subband as a feature vector based on this given resolution. It can be a reasonable approach, because the approximated signal representation in the orthogonal subspace corresponding to this subband is given by these coefficients [14]. In our case, there are two sets of time signals, representing two classes of CBF states and only 40 patterns (2×20) are at our disposal. Intuitively, it is possible to shatter two points by any linear manner in the one-dimensional space and three points in

two-dimensional space. By analogy, it is possible to shatter $N + 1$ points in the N -dimensional space with the probability of 1. If the patterns to be classified are independent and identical distributed, then in the 2^N patterns are linearly separable in the N -dimensional space [15]. The coefficients of the subbands from $n_2 = 2$ up to $n_6 = 2^5 = 32$ as different feature vector components will be employed. The magnitudes of the wavelet coefficients at these subbands are shown in Figure 6.

To carry out the computations, a second order Daubechies filter has been employed, see Figure 7.

According to [6], the two greatest components of a wavelet decomposition do not represent adequately the signals derived from drug induced oscillations. This means that the very small coefficients belonging to the higher resolutions, $n_1 - n_{16}$ and the very big coefficients of the lowest resolution, (n_1) are not taken into consideration. The previous ones have no contributions; the latest one would suppress all of the others (see Figure 5). With other words, we consider the "measurable" fine structure of the subband coefficients. Figure 8 shows the maximums of the magnitude of the wavelet coefficients of different resolutions, except of those belonging to the first (lowest) one. The omitted first wavelet coefficient has a magnitude of about 74276, being significantly larger than the other wavelet coefficients.

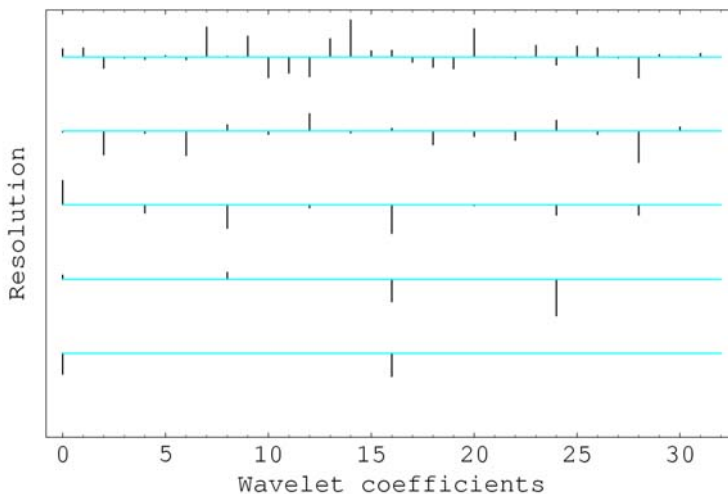


Figure 6

The magnitude of the wavelet coefficients at resolution from (at the bottom) up to (at the top)

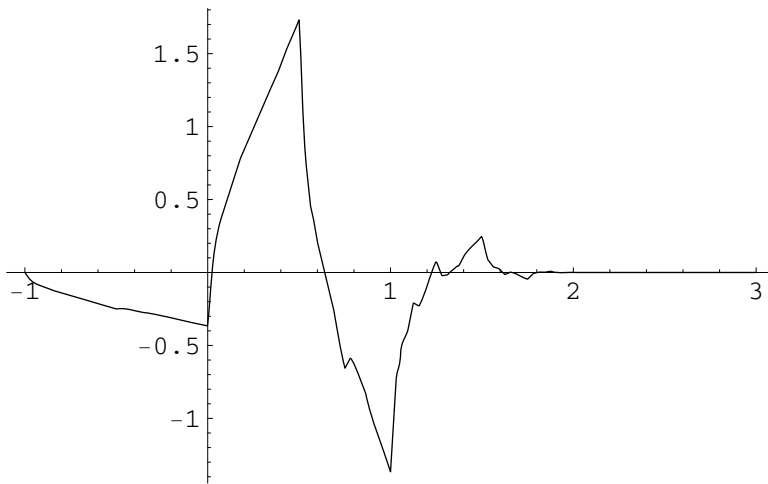


Figure 7
Daubechies filter with $\psi(t)$

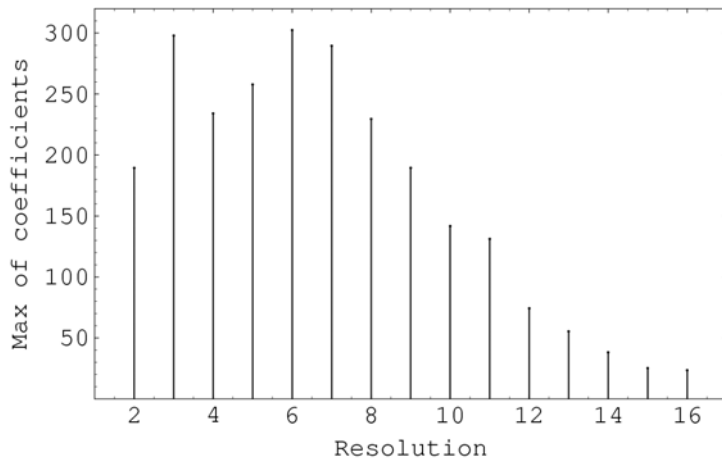


Figure 8
The maximal magnitudes of the wavelet coefficients of different resolutions

3 Classification

3.1 Using Radial Basis Function with Artificial Neural Networks

Figure 9 illustrates an RBF network with inputs x_1, \dots, x_n and output \hat{y} . The arrows in the figure symbolize parameters in the network. The RBF network consists of one hidden layer of basis functions, or neurons. At the input of each neuron, the distance between the neuron centre and the input vector is calculated. The output of the neuron is then formed by applying the basis function to this distance. The RBF network output is formed by a weighted sum of the neuron outputs and the unity bias shown.

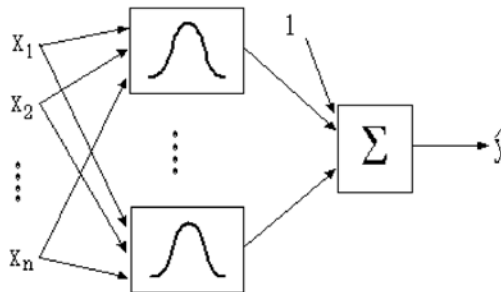


Figure 9
Illustration of an RBF network

The RBF network in Figure 9 is often complemented with a linear part. This corresponds to additional direct connections from the inputs to the output neuron.

Mathematically, the RBF network, including a linear part, produces an output given by

$$\hat{y}(\theta) = g(\theta, x) = \sum_{i=1}^{nb} w_i^2 e^{-\lambda_i^2 (x - w_i^1)^2} + w_{nb+1}^2 + \chi_1 x_1 + \dots + \chi_n x_n \quad (7)$$

where nb is the number of neurons, each containing a basis function. The parameters of the RBF network consist of the positions of the basis function w_i^1 , the inverse of the width of the basis functions λ_i , the weights in output sum w_i^2 and the parameters of the linear part $\chi_1 x_1 + \dots + \chi_n x_n$. In most cases of function approximation, it is advantageous to have the additional linear part but it can be excluded when not necessary.

Considering N patterns of measured CBF signals representing the two overlapping classes, we have

$$x_i \in R^M \quad (8)$$

feature vectors derived from time series samples, where $i = 1 \dots N$ are the samples, and M is the dimension of the feature vectors, consisting of first M dominant eigenvalues. In our case the number of the measurements were $N = 40$. In order to obtain the minimum size of the feature vector which is required to produce reliable results, up to six eigenvalues were used. The goal of the classification problem is to assign new, previously unseen patterns to their respective classes based on previously known examples: in our case to assign input signals to class B or class C. Therefore the output of our unsupervised learning algorithm is a set of discrete class labels corresponding to the different CBF states. The labelled patterns corresponding to classes B and C, were to be classified. This means, that we are looking for a decision function; the output of this estimating function is interpreted as being proportional to the probability that the input belongs to the corresponding class. To carry out the systematic classification of CBF signals, an RBF was used. Radial basis function has the characteristic feature, that the response increases or decreases monotonically according to the distance from a central point. The Gaussian RBF function is used in a single layer network, consisting of two input nodes in the input layer, seven nodes in the hidden layer, and two nodes in the output layer. Several lengths of feature vectors have been fed to the classifier – producing fewer misclassifications as the number of components of the input feature vector increased. The output of the classifier was accepted, if the rounded value of the output nodes corresponded to the proper class, correctly classifying the input pattern, otherwise the classification of that particular input signal was registered as a misclassification. Because the number of the samples are limited ($N = 40$), the dimension of the feature vectors, as well as the number of nodes of the network should be constrained, in order to ensure a reliable teaching process of the network.

3.2 Support Vector Machine (SVM) Classifier

To study the effects of higher dimensional feature vectors on the classification process, an SVM classifier, having no restrictions regarding input vector dimensions, has also been applied.

Let us consider a set of training samples,

$$S = ((x_1, y_1), \dots, (x_m, y_m)) \quad (9)$$

with labels $y_i = 1$ or -1 , respectively and use the feature space implicitly defined by the kernel $K(x, z)$. We suppose that the parameters are the solutions of the following quadratic optimization problem,

$$\text{maximize } W(\alpha) = \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m y_i y_j \alpha_i \alpha_j \left(K(x_i, x_j) + \frac{1}{c} \delta_{i,j} \right) \quad (10)$$

$$\text{subject to } \sum_{i=1}^m y_i \alpha_i = 0, \alpha_i \geq 0, i = 1 \dots m \quad (11)$$

$$\text{Let } f(x) = \sum_{i=1}^m y_i \alpha_i^* K(x_i, x) + b^* \quad (12)$$

where b^* is chosen so that

$$y_i f(x_i) = 1 - \frac{\alpha_i^*}{c} \quad (13)$$

for any i with

$$\alpha_i^* \neq 0 \quad (14)$$

Then the decision rule given by $\text{sign}(f(x))$ is equivalent to the hyperplane in the feature space implicitly defined by the kernel $K(x, z)$, which solves the optimization problem, where the geometric margin is

$$\gamma = \left(\sum_{i \in sv} \alpha_i^* - \frac{1}{c} \langle \alpha^*, \alpha^* \rangle \right)^{\frac{1}{2}} \quad (15)$$

and the set sv corresponds to indexes i , for which $\alpha_i^* \neq 0$,

$$sv = \{i : \alpha_i^* \neq 0; i = 1 \dots m\} \quad (16)$$

Training samples, x_i , for which $i \in sv$ are called support vectors contributing to the definition of $f(x)$. The geometric margin, γ can indicate the quality of the classification [16], greater the γ , more reliable the classification is.

This kernel based classifier can be trained on any size of training set, while neural networks should have so many input nodes as the dimension of the input space and need definitely more training patterns than the number of these input nodes. Employing kernels, a classification problem can be transferred in a higher

dimensional space, where the linear separability is more likely. In addition, the quality of the classification in any dimension can be measured by the geometric margin of the SVM classifier [16]. Here we used the feature vectors produced by the wavelet subband analysis. Twenty of these vectors represent one CBF state, the other twenty represent the other state. As an example let us load the coefficients of the fifth subband, $n_s = 2^4 = 16$, for all of the 40 patterns, giving us 40 feature vectors of dimension of 16. First, these data should be standardized; to be transformed so that their mean is zero and their unbiased estimate of variance is unity. Let us employ Gaussian kernel, with parameter $\beta = 5$, as seen on Figure 10.

$$K(u, v) = e^{-\beta(u-v)^T(u,v)} \quad (17)$$

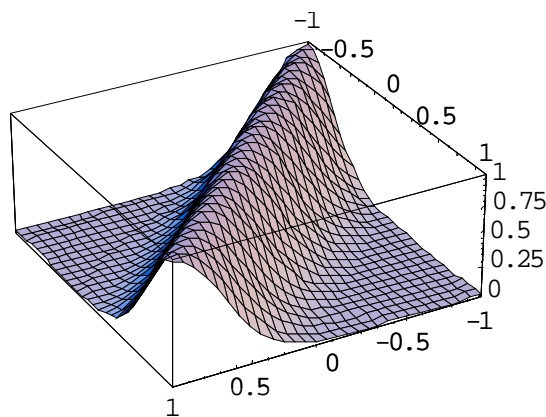


Figure 10

Gaussian RBF universal kernel with $\beta = 5$, in case of $x, y \in \mathbb{R}^1$

Let the value for the control parameter of regularization be $c = 100$. To carry out the training of the support vector classifier, we shall employ the algorithm embedded into the function, *SupportVectorClassifier* developed for *Mathematica* [17]. A sample pattern can be considered as support vector, if its contribution (its weighting coefficient α_i) to the decision function is greater than 1% of the maximal contribution.

These computations were carried out for different feature vectors based on the coefficients of the different subbands.

4 Results

4.1 Result of the SVM Classification

Table 1 shows, that by decreasing the number of the wavelet coefficients, the Gram matrix is getting ill-conditioned, the geometric margin is becoming narrower and probability of the misclassification of patterns is increasing, although the classification with four wavelet coefficients is just acceptable. Let us employ the traditional feature extraction method, when the elements of the feature vector are computed as the average of squares of the wavelet coefficients belonging to the same subband, plus the same contribution of the original signal as additional subband. Consequently, the dimension of the feature vector is $16 + 1 = 17$. Table 2 shows the result for this case. These results correspond with the results of the classification carried out with the eight dimensional feature vectors based on subband level 4, however now the dimension of the feature vectors is 17 instead of 8. The robustness of the SVM classifier has been also proved by successful classification of noisy samples.

Table 1
The results of the SVM classification with different feature vectors

Subband level	Number of wavelet coefficients	Determinant of Gram matrix	Condition number of Gram matrix	Number of support vectors	Geometric margin	Number of misclassified patterns
6	32	1.	1.	40	0.159695	0
5	16	1.	1.	40	0.159695	0
4	8	0.999	1.040	40	0.159701	0
3	4	0.005	69.374	40	0.113922	0
2	2	1.9310^{-39}	1.1510^7	25	0.083355	4

Table 2
The results of the SVM classification employing traditional feature extraction technique

Determinant of Gram matrix	Condition number of Gram matrix	Number of support vectors	Geometric margin	Number of misclassified patterns
0.994	1.170	40	0.159374	0

4.2 Result of the ANN Classification

Columns 1 and 2 of Table 3 show the result of the ANN classification results using eigenvalue-based feature extraction. It can be clearly seen from the numbers, that it makes no sense to use more than 6 eigenvalues. Comparing different feature extraction methods and classification algorithms by taking different numbers of eigenvalues as feature vectors, the results are very close to that obtained when using wavelet decomposition, see columns 3 and 4 of Table 3. In any case, it is clear, that a merely two element feature vector is insufficient for reliable results; at least a five element feature vector was needed to differentiate class B from class C in the case of ANN classification with eigenvalue-based feature extraction, while in the case of the SVM classification with wavelet-based feature extraction, at least 8-dimensional feature vector should be used.

Table 3
Misclassification rate

Wavelet coefficient (subband level)	Number of eigenvalues	Wavelet feature extraction & SVM classification, misclassification number	Eigenvalue feature extraction & ANN classification, misclassification number
16 (5)	6	0	0
8 (4)	5	0	0
4 (3)	4	0	3
-	3	-	3
2 (2)	2	4	6

Conclusions

Two feature extraction and classification methods are presented. First an Artificial Neural Network using a radial based function, combined with a spectral matrix based feature extraction was shown. Secondly, a Support Vector Machine Classifier with wavelet subband analysis as feature extraction method was employed. The two methods can successfully differentiate cerebral blood flow classes B and C, and although the approaches described in this paper are very different, they still produced comparable results for this classification problem.

Acknowledgement

The research was supported by the Hungarian National Research Fund, Grants No. OTKA F046726, T042990, and by the Hungarian Ministry of Economics and Transportation Grant No. RET-04-2004.

References

- [1] H. Nilsson and C. Aalkjær, "Vasomotion mechanisms and physiological importance.", *Molecular Interventions*, 7:59-68, 2003
- [2] e. a. B. Diehl, "Spontaneous oscillations in cerebral blood flow velocities in middle cerebral arteries in control subjects and patients with epilepsy.", *Stroke*, 28:2457-2459, 1997
- [3] e. a. G. Lenzser, "Nitric oxide synthase blockade sensitizes the cerebrocortical circulation to thromboxane-induced CBF oscillations.", *Journal of Cerebral Blood Flow and Metabolism*, 23:88, 2003
- [4] e. a. Z. Lacza, "The cerebrocortical microcirculatory effect of nitric oxide synthase blockade is dependent upon baseline red blood cell flow in the rat.", *Neuroscience Letters*, 291:65-68, 2000
- [5] e. a. Z. Lacza, "NO synthase blockade induces chaotic cerebral vasomotion via activation of thromboxane receptors.", *Stroke*, 32:2609-2614, 2001
- [6] e. a. B. Benyó, "Characterization of the Temporal Pattern of Cerebral Blood Flow Oscillations", In Proc. of 2004 International Joint Conference on Neural Networks, pp. 468-471, July 2004
- [7] e. a. B. Benyó, "Classification of Cerebral Flood Flow Oscillation using SVM classifiers with different kernels", *Intelligent Systems at the Service of Mankind*, 2:145-151, 2005
- [8] e. a. B. Benyó, "Characterization of Cerebral Blood Flow Oscillations Using Different Classification Methods", In Proc. of 2005 IFAC, 2005. electronic publication #04987
- [9] W. Shaw and J. Tigg, "Applied Mathematica", Addison-Wesley, 1994
- [10] J. Goswami and A. K. Chan, "Fundamentals of Wavelets. Theory, Algorithms, and Application", Wiley, 1999
- [11] e. a. Li, "Detection of ECG Characteristic Points Using Wavelet Transforms", *IEEE Trans. on Biomedical Eng.*, 1995, Vol. 42, pp. 21-28
- [12] e. a. L. Senhadji, "Comparing Wavelet Transforms for Recognizing Cardiac Patterns", *IEEE EMBS Magazine*, Vol. 14, pp. 167-173
- [13] B. Paláncz, "Wavelets and their Application to Digital Image Processing", *Publications in Geomatics*, 2004, Győr, ISSN 1419-6492
- [14] Aboufadel and Schlicker, "Discovering Wavelets", Wiley-Interscience, 1999
- [15] e. a. L. Zhang, "Wavelet Support Vector Machine", *IEEE Trans. Systems, Man and Cybernetics - Part B: Cybernetics*, 4(1):34-39, Februray 2004
- [16] N. Cristianini and J. Shawe-Taylor, "An introduction to Support Vector Machines and other kernel-based learning methods", Cambridge University Press, 2000
- [17] B. Paláncz, "Support Vector Classifier via Mathematica", Wolfram Research Mathematica Information Center, 2004, <http://library.wolfram.com/infocenter/mathsource/5293>