SVM Classification for Diabetics with Various Degrees of Autonomic Neuropathy Based on Cross-Correlation Features

Chuang-Chien Chiu\(^1\)\(^,*\)  Shouou-Jeng Yeh\(^2\)  Yu-Hen Hu\(^3\)  Ken Ying-Kai Liao\(^4\)

\(^1\)Department of Automatic Control Engineering, Feng Chia University, Taichung 407, Taiwan, ROC
\(^2\)Section of Neurology and Neurophysiology, Cheng-Ching General Hospital, Taichung 407, Taiwan, ROC
\(^3\)Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706, U.S.A
\(^4\)Program of Electrical and Communications Engineering, Feng Chia University, Taichung 407, Taiwan, ROC

Abstract

This study investigates the feasibility of using the cross-correlation between mean arterial blood pressure (MABP) and mean cerebral blood flow velocity (MCFBV) as a signature to distinguish diabetics with various degrees of autonomic neuropathy. 54 subjects were recruited. Among them, 15 were healthy adults (normal subjects), 17 were diabetics with mild autonomic neuropathy symptoms, and 22 were diabetics with severe autonomic neuropathy symptoms. The cross-correlation function of pre-filtered spontaneous MABP and MCFBV was computed. The maximum peak and the corresponding standard deviation along with the maximum peak index and the corresponding standard deviation were extracted from three frequency ranges, namely very low frequency (0.015-0.07 Hz), low frequency (0.07-0.15 Hz), and high frequency (0.15-0.40 Hz), in the supine position, giving a total of 12 features for each subject. After feature reduction using a greedy forward selection technique, each subject was classified into one of three possible classes. The feature vectors were classified based on a support vector machine (SVM) classifier with a classification rate of 90.74%. The results indicate that SVM classification based on the cross-correlation between MABP and MCFBV could be an effective approach for discriminating diabetics with various degrees of autonomic neuropathy.

Keywords: Support vector machine (SVM), Diabetes mellitus, Autonomic neuropathy, Cross-correlation function

1. Introduction

Cerebral autoregulation (CA) [1] refers to the intrinsic mechanism of the cerebrovascular subsystem that maintains relatively constant cerebral blood flow (CBF) over large systemic changes (in the range of 50 to 170 mmHg) of the mean arterial blood pressure (MABP). This mechanism ensures constant perfusion of blood to the brain and is essential for normal brain function [2]. Various cerebrovascular diseases may disrupt the CA mechanism [3,4]. For instance, the cerebral autoregulatory capacity may be partially or completely lost after a stroke or subarachnoid hemorrhage. In long-term diabetes mellitus, there may be chronic impairment of CBF autoregulation, probably due to diabetic micro-angiopathy [4].

A number of noninvasive static and dynamic methods have been developed for CA assessment [5]. Recent investigations have shown that the autoregulatory dynamic response can be identified from spontaneous fluctuations in MABP and cerebral blood flow velocity (CBFV) [6]. The dynamic relationship between spontaneous MABP and CBFV has been investigated using transfer function analysis in both normal subjects [7,8] and autonomic failure patients [9]. Some research groups have used spontaneous blood pressure changes as input signals to test CA conditions [10,11]. Spectral and transfer function analyses of CBFV and arterial blood pressure (ABP) were performed using the fast Fourier transform (FFT) in these experiments. However, the stationary property and time resolution are two critical problems in spectral analysis. The dynamic influences of ABP and CO\(_2\) on CBFV were studied in [12], where continuous recordings of spontaneous beat-to-beat fluctuations in MABP and breath-by-breath variability in end-tidal CO\(_2\) (EtCO\(_2\)) were obtained from normal subjects at rest.

Testing potential CA impairment usually requires the introduction of variations in ABP via physiological or pharmacological manipulation. In many studies, head-up tilt has been applied as a reliable method to study CA by inducing defined perturbations to cardiovascular and cerebrovascular systems [11,13]. Assuming the upright posture increases vulnerability to the effects of gravity on circulation in humans. The brain, the organ most susceptible to hypoxia, is in the most disadvantageous location [14,15]. A series of cardiovascular-
regulatory mechanisms or reflexes [16] are activated to offset this assault on the circulation. The main system involved in the maintenance of cerebral perfusion is CA reacting to changes in ABP.

The time-domain cross-correlation function (CCF) has been recently applied to characterize the dynamics of CA [17-19]. Our previous studies showed that CCF could be a useful tool for assessing CA in normal subjects [20]. CA has been assessed in diabetics with autonomic neuropathy by relating the CCFs obtained in supine and tilt-up positions to orthostatic cardiovascular and cerebrovascular changes [21]. Cross-correlation analysis provides the correlation and the phase relationship between MABP and mean CBFV (MCBFV). Results indicated that the correlation values for normal subjects are generally higher than those for diabetics with autonomic neuropathy [21]. Some investigators presumed that autoregulation failure manifests itself as a zero or nearly zero time delay [22].

In this study, the feasibility of using the CCF between MABP and MCBFV in the supine position as a signature to distinguish diabetic patients with varying degrees of autonomic neuropathy is investigated. Specifically, an effective signal preprocessing and feature extraction method for extracting the CCF signature is proposed. A support vector machine (SVM) algorithm is adopted to develop a pattern classifier to classify the CCF signatures into three classes: normal, diabetic with mild autonomic neuropathy, and diabetic with severe autonomic neuropathy.

2. Materials and methods

2.1 Subjects and measurements

Three groups of subjects were recruited in this study, namely 15 healthy adults as normal subjects (group 1, 8 men, 7 women) with a mean age of 43.47 ± 17.23 years, 17 diabetics with mild autonomic neuropathy (group 2, 8 men, 9 women) with a mean age of 63.00 ± 8.11 years, and 22 diabetics with severe autonomic neuropathy (group 3, 18 men, 4 women) with a mean age of 60.32 ± 10.24 years. The subjects in the healthy group were included only if they had no history of vascular disease, heart problems, hypertension, migraine, epilepsy, cerebral aneurysm, intra-cerebral bleeding, or other pre-existing neurological conditions. None of the subjects received any medication during the time of the study. No food, caffeine, or nicotine was permitted for 3 hours before the experiment.

In this study, the autonomic functions were assessed using a battery of autonomic reflex tests, namely two cardiovagal tests, namely deep breathing heart rate change and Valsava maneuver ratio, one sudomotor test (the sympathetic skin response (SSR) test), and two vasomotor function tests, namely blood pressure (BP) change in tilting and Valsava phase change. Each test was scored as normal (0), borderline (1), or abnormal (2). A composite score ranging from 0 to 10 was used to indicate the severity of autonomic abnormalities with a minor modification of Low’s techniques [25,26]. For pattern classification, patients with less than two abnormal test scores (i.e., a composite score ranging from 0 to 3) were labeled as having borderline autonomic failure; those with two to three abnormal test scores (i.e., a composite score ranging from 4 to 6) were labeled as having mild to moderate autonomic failure, and those with more than three abnormal test scores (i.e., a composite score ranging from 7 to 10) were labeled as having severe autonomic failure.

The CBFV measurements were measured from the right middle cerebral artery using transcranial Doppler ultrasound (TCD, EME TC2020, Nicolet Instruments, Warwick, UK) in conjunction with a 5-MHz transducer fixed over the temporal bones using an elastic headband. Continuous ABP recordings were obtained through the Finapres (Finapres, 2300, Ohmeda, Englewood, CO, USA) device with the cuff attached to the middle finger of the right hand. Data acquisition was started after 10 min of relaxation in the supine position. Spontaneous ABP and CBFV were recorded simultaneously to a personal computer (PC) for off-line analysis. The data were acquired for periods of approximately 5 min (in the supine position) using a custom-developed data acquisition system (see Fig. 1). A PC equipped with a general purpose data acquisition board was used for data acquisition within the LabVIEW® (National Instruments, Austin, TX, USA) environment [20]. The sampling rate was 60 Hz.

Figure 1. Diagram of data acquisition system for acquiring ABP and CBFV signals from Finapres and TCD, respectively.

2.2 Preprocessing and feature extraction

The blood volume in the finger artery under an inflatable finger cuff was measured with an infrared plethysmograph. The cuff pressure was adjusted automatically at regular intervals using a built-in servo adjustment mechanism to keep the blood volume at a preset value. The artifacts caused by the regular servo adjustment in the continuously acquired ABP pulse signal are referred to as the “servo component”. The time-domain CCF used for characterizing the dynamics of CA was calculated using a 64-beat-wide moving window [20] from simultaneously acquired ABP and CBFV signals. As such, the direct removal of servo artifacts from the ABP recording was not appropriate because it would result in different time durations in comparison to that of the CBFV signal. Instead, a signal relaxation algorithm was developed to compensate for the servo component. In the calculation of the CCF for each 64-beat-wide moving window, at most 2-4 out of 64 beats in the ABP pulse signal required such compensation.

An example of servo component signal relaxation is illustrated in Fig. 2. The upper plot shows the continuously acquired ABP pulse signal over a 5-min duration (256 beats). The lower part is an enlargement of a short segment containing
between the MABP and MCBFV time series is calculated. MABP and MCBFV are normalized using their mean values. The normalized MABP and MCBFV time series are denoted as \( f(n) \) and \( g(n) \), respectively. The \( f(n) \) and \( g(n) \) signals are then bandpass-filtered in the very low-frequency (VLF, 0.015-0.07 Hz), low-frequency (LF, 0.07-0.15 Hz), and high-frequency (HF, 0.15-0.40 Hz) ranges before the CCF is applied. A Chebyshev filter is used to minimize the error between the idealized filter characteristics and the actual characteristics over the range of the filter, but with ripples in the passband. There are two types of Chebyshev filter, Type I (with equal ripple in the passband, monotonic in the stopband) and Type II (with equal ripple in the stopband, monotonic in the passband) [27]. In this study, a third-order digital-bandpass Type I Chebyshev filter was applied. The filter sampling period was set equal to the mean heart period of each subject. The low cutoff frequency and high cutoff frequency were assigned at the beginning and ending passband frequencies, respectively, at each bandpass frequency range of interest. The passband ripple error of each bandpass Chebyshev filter was limited to 0.1 dB. The bandpass-filtered \( f(n) \) and \( g(n) \) time series are denoted as \( \hat{f}(n) \) and \( \hat{g}(n) \), respectively. The CCF is calculated as:

\[
CCF_i(k) = \frac{R_{\hat{f}\hat{f}}(k)}{\sqrt{R_{\hat{f}\hat{f}}(0)R_{\hat{g}\hat{g}}(0)}}
\]

where \( R_{\hat{f}\hat{f}}(k) \) is an estimate of the cross-covariance in the \( i \)th time window and defined as:

\[
R_{\hat{f}\hat{f}}(k) = \frac{1}{W} \sum_{j=0}^{W-1} \left( \hat{f}(j) \hat{f}(j+k) \right), k = 0, 1, 2, ...
\]

Also, \( R_{\hat{f}\hat{f}}(0) = \frac{1}{W} \sum_{j=0}^{W-1} \hat{f}^2(j) \) and \( R_{\hat{g}\hat{g}}(0) = \frac{1}{W} \sum_{j=0}^{W-1} \hat{g}^2(j) \). \( N \) is the total number of cardiac cycles, \( W \) is the window width, and \( k \) is the time lag. \( CCF_i(\cdot) \) is the result of the CCF between \( \hat{f}(n) \) and \( \hat{g}(n) \) in the \( i \)th time window. Mean CCF patterns were obtained for each subject and for the entire population.

For linear systems and long observation times, the information provided by cross-correlation analysis is directly related to that obtained with cross-spectral analysis. However, cross-correlation analysis can be performed over much shorter observation times than those for traditional cross-spectral analysis [28]. Our previous studies [20] adopted a relatively narrow moving window of the CCF to study the properties of the linear transient relationship between MABP and MCBFV. A 64-beat moving window was applied. The total number of cardiac cycles used for both MABP and MCBFV to compute the CCF was 256 beats. Thus, the total number of CCFs included for each subject was 193. The MABP and MCBFV signals were bandpass-filtered in the VLF, LF, and HF ranges before features were extracted from the CCFs for studying the effect of dynamic CA. The filtering ranges were selected on the basis of those previously defined for heart rate variability in
short-term recordings [29]. These specific frequency bands of interest were determined in our previous study [20]. The maximum peak and index were calculated for each CCF, and the mean and standard deviation for the maximum peak and index were calculated from the 193 CCFs as the features to represent each subject. Due to the possibility of noise and artifacts, only peaks found between -10 and +5 were included in the mean and standard deviation calculations. Therefore, there are 12 features for each subject. A block diagram of the preprocessing before feature extraction is shown in Fig. 3.

**2.3 Support vector machine**

Support vector machine (SVM) is a very useful technique for pattern classification. The goal is to produce a model that predicts pattern class labels in the test set given the features. The training patterns are mapped into a higher dimensional space and SVM finds a linear separating hyperplane with the maximum margin in this higher dimensional space to classify different groups [30]. Assuming that the training patterns are linearly separable, consider the training data set

\[ S = \{(x_i, y_i), \ldots, (x_l, y_l)\}, x_i \in \mathbb{R}^n, i = 1, 2, \ldots, l, y_i \in \{+1, -1\} \]  

(5)

where \( x_i \) is the \( i \)th training pattern, \( y_i \) is the class label of the \( i \)th training pattern, \( l \) denotes the total number of training data, and \( n \) denotes the dimensions of the feature spaces. As shown in Fig. 4, the function \( \langle w \cdot x \rangle + b = 0 \) is a dividing line that separates the training data into two categories. However, the training data may not be completely separable by a hyperplane. Slack variables, denoted by \( \xi_i \), \( \xi_i \geq 0, i = 1, 2, 3, \ldots, l \), can be introduced to relax the constraints. An error penalty parameter, denoted by \( C, C > 0 \) is also used to add a cost to the misclassified training data points so that the usage of slack variables will be minimized, resulting in a hyperplane that allows for small errors in training by penalizing these errors. Therefore, the SVM requires a solution for the following optimization problem:

\[
\min_{w, b, \xi} \frac{1}{2} ||w||^2 + C \sum_i \xi_i \text{ subject to } y_i \langle w \cdot x_i \rangle + b \geq 1 - \xi_i, i = 1, \ldots, l
\]  

(6)

\( K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \) is called the kernel function. One of the main novelties of SVMs is their ability to use a kernel function to map the feature space into a higher dimensional space, where the classes are linearly separable, and find the support vectors in that higher dimensional space. SVMs use the following four basic kernels [30]: linear, polynomial, radial basis function (RBF), and sigmoid kernels. The RBF kernel, \( K(x_i, x_j) = \exp(-\|x_i - x_j\|^2), \gamma > 0 \), is applied here. With the RBF kernel, the parameter gamma \( (\gamma) \) can be adjusted to map the data onto the best higher dimensional space for the classification.

\[
\langle w \cdot x \rangle + b = 0
\]

\[
\langle w \cdot x \rangle + b = +1
\]

\[
\langle w \cdot x \rangle + b = -1
\]

Figure 4. Example of linear SVM.

**3. Results and discussion**

Autonomic functions are assessed generally based on Low’s noninvasive techniques [24,25]. The severity of autonomic abnormalities can be indicated by a composite score [26]. It would be helpful to develop a simple, easy, and effective identification of the severity of neuropathy in diabetes patients. This study uses an SVM classification algorithm to categorize a cross-correlation of features related to only MABP and MCBFV to identify levels of autonomic neuropathy in diabetics. No special pharmaceutical maneuvers or the Valsalva maneuver, such as the one used in Low’s technique, are needed for the measurement. The measurement used is thus less time-consuming, affordable, and entirely noninvasive compared to the battery of tests that are currently used for diagnosis.

Previous studies comparing the CA capacity of healthy young subjects to that of healthy aged subjects between 50 and 75 years old found no evidence that aging affects dynamic CA [6]. It was also reported [23] that dynamic CA during orthostatic stress is unaffected by aging. Yam et al. [24] calculated the correlation index between spontaneous BP and CBFV, and found no correlation between age and the correlation index in healthy people. Thus, the age difference between the healthy group and the diabetic group should not be of any concern.

The phase shift analysis between MABP and MCBFV depicted the phenomenon of dynamic autoregulation. The negative lag of a CCF peak between MABP and MCBFV is the result of the phase-lead property and indicates that CA is a highpass filter. The increasing time lag in the CCF peak shows the autoregulatory disturbance effect. The cross-correlation analysis indicates that changes in MCBFV usually preceded...
those in MABP in most subjects. However, the standard deviations of the corresponding peak CCF value time lag in three diabetic groups were usually larger than those in the normal group. Note that time lags between MABP and MCBFV closer to the origin imply less efficient CA. In general, data for diabetics with severe autonomic neuropathy were more consistent and their standard deviations for CCF peak time lags are larger than those in other groups. The data in normal subjects were relatively scattered.

The SVM software LIBSVM, developed by C. C. Chang and C. J. Lin [31], was used for the classification task. The SVM used in LIBSVM for multiclass classification is based on the “one-against-one” approach [32], in which \(k(k-1)/2\) classifiers are constructed, where \(k\) is the number of classes, and each one trains data from two different classes. A voting strategy is used as follows. Each binary classification is considered to be a vote, where votes can be cast for all data points. In the end, the data point is designated to be in a class with the maximum number of votes. In case two classes have identical numbers of votes, the one with the smaller index is selected. In our experiments, the best \(C\) and gamma parameters were selected by testing the SVM using leave-one-out cross-validation for all data using a grid search with base 2. The best \(C\) was \(2^0 = 1\), and the best gamma was \(2^{-1} = 0.5\). Normalizing the features made no significant difference. In machine learning, bad features have a large negative impact on classification accuracy. This phenomenon also appears in SVMs, as demonstrated by Weston et al. [33]. Therefore, a greedy forward selection feature algorithm was applied to remove the features that reduced classification accuracy. The following three features were applied to SVM: 1) the mean CCF index from the VLF frequency range, 2) the standard deviation of the CCF index from the VLF frequency range, and 3) the standard deviation of the CCF index from the LF frequency range. The mean and standard deviations of these features are shown in Fig. 5, respectively. Using leave-one-out cross-validation to classify the 54 subjects into the three groups, 90.74% accuracy (49 out of 54 subjects were classified correctly) was achieved, as shown in Table 1. In these 3-class classification results, the misclassified cases occurred in diabetics with mild autonomic neuropathy and diabetics with severe autonomic neuropathy, with classification of diabetics with mild autonomic neuropathy being the worst. This could be due to the fact that when autonomic functions were assessed in clinical practice, patients were classified based on how they scored on a battery of tests. Sometimes the patients would be severe in one test, but show mild autonomic neuropathy in other tests. However, the physician classified them as having mild or severe autonomic failure based on their total score. These results show that an SVM model can be helpful in the classification of patients without any pharmaceutical maneuvers, especially between normal and having diabetic autonomic neuropathy.

Using the proposed model, a group which was classified as diabetics who had borderline autonomic neuropathy (scoring mild or severe on one test, scoring zero on other tests, or completely scoring zero) was tested. The results of this classification are shown in Table 2. The results are highly encouraging, as most patients in this group actually had an autonomic score that was greater than 0, indicating that they had failed at least one autonomic test. The two patients that were classified as normal subjects both had an autonomic score of zero, which means that they passed all the autonomic tests successfully, which indicates that they truly had no autonomic neuropathy.

4. Conclusion

This study used the time-domain cross-correlation approach to test the effects of changes in MABP on MCBFV measured in the supine position. 90.74% classification accuracy...
was achieved using leave-one-out cross-validation to classify three groups. The results indicate that using features extracted from CCFs and an SVM classifier can be a simple, easy, and effective method for classifying dynamic CA in diabetics with autonomic neuropathy. The cross-validation results showed a clear line between normal subjects and patients with diabetic autonomic neuropathy. Tests with this model on a borderline autonomic neuropathy group produced promising results, which means that the model could be used for initial screening for autonomic neuropathy in people with diabetes. Our future work involves reproducing our findings in a larger population, including patients that might show the absence or impairment of pressure autoregulation. This paper used a very simple feature selection technique (greedy forward selection). Future research could adopt more sophisticated feature selection techniques that have been demonstrated to be more effective.

Acknowledgments

The authors would like to thank the National Science Council, Taiwan, R.O.C., for supporting this research under grants NSC 96-2628-E-035-083-MY3, NSC 98-2918-I-035-001, and NSC 99-2221-E-035-054-MY3.

References