Microemboli Classification using Non-linear Kernel Support Vector Machines and RF Signals

1Karim Ferroudji, 1Nabil Benoudjit and 2Ayache Bouakaz

Abstract- Man's intelligent behavior is due in part to his ability to select, classify, and abstract significant information reaching him from his environment by way of his senses. This function, pattern recognition, has become a major focus of research by scientists working in the field of artificial intelligence. Due to its clinical importance, several classification methods have been studied for microemboli detection and characterization. In the human body; emboli can produce severe damage like stroke or heart attack thus the importance of an automatic classification system. In this paper, we propose a new approach to detect and classify microemboli using support vector machine and the backscatter Radio-Frequency (RF) signal.

This short communication demonstrates the opportunity to classify emboli based on a RF signals and support vector machine; the classification rates reached 96.42%.

Keywords: Microemboli, Gaseous embolus, Solid embolus, RF signals, Non linear Kernel, Support vector machines, classification.

1. INTRODUCTION

Classification and prediction are among the major techniques in data mining and widely used in various fields. The goal of these techniques is to build a concise model of the distribution of class labels in terms of predictor features. The resulting classifier is then used to assign class labels to the testing instances where the values of the predictor features are known, but the value of the class label is unknown.

In this article we present a study on how microemboli problems can be solved using classification and prediction techniques. Many techniques were developed in this field in order to ameliorate this process.

An embolus is defined as a foreign body that arises from a location and travel in the bloodstream. It may consist of gas bubbles or biologic aggregates. Emboli signals have been detected in patients with a wide variety of potential embolic sources including carotid artery stenosis, atrial fibrillation, and valvular heart disease [1]. Commonly used Doppler detection techniques have shown their limits to distinguish between emboli and artifacts [2], and more importantly to make the differentiation between solid and gaseous microemboli [3]. An alternative approach would be to examine Radio Frequency (RF) signal instead of Doppler signals. The gas bubble, under specific ultrasound wave, shows a non-linear behavior which is used to differentiate gaseous emboli from solid emboli.

Recently a relatively new classification method, artificial neural network (ANN), K-Nearest Neighbors Rule (KNNR) [4, 5], has been used for the prediction of the embolus nature. We suggest in this experimental study a new approach to detect and classify microemboli using support vector machines (SVM) and the backscatter RF signal.
Support Vector Machine (SVM) is a new promising pattern classification technique proposed recently by Vapnik and co-workers [6-8]. Unlike traditional methods which minimize the empirical training error, SVM aims at minimizing an upper bound of the generalization error through maximizing the margin between the separating hyperplane and the data.

SVMs have been employed in a wide range of real world problems such as text categorization [9], hand-written recognition [10], tone recognition [11], image classification and object detection [12], cancer diagnosis [13], gene expression data analysis [14]. It has been shown that SVM is consistently superior to other supervised learning methods [15].

Experimental results are obtained for two different concentrations of microbubbles and for two different Mechanical Index (MI).

The amplitudes and bandwidths of the fundamental and the second harmonic components were selected as input parameters to the model. Moreover, the frequency bandwidths of the fundamental and the second harmonic spectral echo components were approximated by Gaussian functions, and their coefficients were used as a third input parameter to the SVM classifier.

The rest of this manuscript is organized in four sections. Section 2 describes the basics of support vector machines when used as classifier. Section 3 presents the material and methods used in this experimental study. In Section 4, we discuss the experimental results obtained for two concentrations (5 µl and 10 µl) of microbubbles at low (0.2) and high (0.6) MI (the mechanical index). Section 5 draws the conclusion of this study.

2. SUPPORT VECTOR MACHINE

The Support Vector Machine (SVM) [6-8] is a novel machine learning technique based on a statistical learning theory proposed by the Vapnik group in 1995. It has gained increasing attention in areas that range from pattern recognition to regression estimation, due to its perfect learning performance. Based on the structural risk minimization principle [16], it can perfectly improve the generalization ability of a learning machine. At the same time, an optimization problem can be transformed into a convex quadratic programming problem. The solution of a quadratic programming problem is the unique optimization solution of the whole. Hence, SVM does not have the problems with local extrema that are present for traditional neural networks, which require large numbers of training samples.

The SVM performs pattern recognition for two-classes problems by determining the hyperplane of separation with the maximum distance to the narrowest points of the positioning of formation. These points are called the vectors of support.

Given a training sample set, $W = \{(x_i, y_i), i = 1 \ldots l\}$, an input sample, $x_i \in \mathbb{R}^d$, and class labels, $y_i \in \{-1, +1\}$, for a linearly separable binary classification problem, the separation hyperplane is:

$$x_i \cdot w + b = 0$$

where $w$ is a weight vector and $b$ is a bias.

The optimal hyperplane that separates the data into two classes minimizes:

$$\phi(w) = \frac{1}{2}||w||^2 = \frac{1}{2}(w \cdot w)$$

subject to the constraint:
In many practical situations, a separating hyperplane does not exist. To allow for the possibilities of violating the constraint equation (3), slack variables, $\xi_i > 0$, are used to get:

$$y_i[(x_i \cdot w) + b] \geq 1 - \xi_i, i = 1, \ldots, l$$  \hspace{1cm} (4)

The optimization problem now becomes:

$$\phi(w, \xi) = \frac{1}{2}(w \cdot w) + c \left[ \sum_{i=1}^{l} \xi_i \right], i = 1, \ldots, l$$  \hspace{1cm} (5)

Where $c$ is a user-defined constant, called a trade-off factor, which determines the balance between the maximization of the margin and the minimization of the classification error.

Introducing Lagrange multipliers $\alpha_i$ and using the Karush–Kuhn–Tucker theorem of optimization gives the solution as follows:

$$w = \sum_{i=1}^{l} \alpha_i y_i x_i$$  \hspace{1cm} (6)

Only a few of the $\alpha_i$ coefficients are nonzero. The corresponding $x_i$ values are known as support vectors, and they also define the decision boundary. At the same time, all other training samples with zero $\alpha_i$ values are now rendered irrelevant. Finally, the decision function can be obtained as follows [17]:

$$f(x) = \text{sgn} \left( \sum_{i=1}^{l} y_i \alpha_i (x_i, x) + b \right)$$  \hspace{1cm} (7)

SVM was originally designed for linear binary classification. In practice, many applications of SVM are nonlinear classification problems. For nonlinear classification problems, we can use nonlinear transforms. A transformation, $\Phi(x)$, maps the data from the input space to a feature space that allows linear separation. We then seek the optimization separation plane in feature space. One only needs an inner product operation in feature space. The inner product operation may be implemented by a certain function (called a kernel function). In SVM, we introduce a kernel function $K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ to perform the transformation. Then the basic form of SVM can be obtained [17]:

$$f(x) = \text{sgn} \left( \sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b \right)$$  \hspace{1cm} (8)

The commonly used kernel functions are presented in Table 1 [18].

<table>
<thead>
<tr>
<th>Kernel function type</th>
<th>Expressions</th>
<th>Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In our case, radial basis kernel was used to differentiate between gaseous and solid embolus.

More details about recent developments of SVM can be found in [19].

6. EXPERIMENTAL SET-UP

The experimental set-up consists of a nonrecirculating flow phantom containing a 0.8mm diameter vessel, see figure 1. A continuous flow carries the Sonovue microbubbles through the insonified vessel. The concentration of contrast agent microbubbles is chosen to obtain comparable fundamental scattering amplitude than non perfused tissue used to simulate the response of a solid embolus.

![Experimental set-up](image)

The ultrasound waves were generated by a VF13-5 probe connected to a Siemens Antares scanner. The acquisitions were performed at 1.82MHz transmitting frequency in Tissue Harmonic Imaging (THI) mode. Two concentrations of the contrast agent were used, 5µl and 10µl. The microbubbles were administered into a 200 ml volume of Isoton [4]. Three different types of input parameters were selected: The bandwidth of the scattered RF signals around the fundamental and the second harmonic central frequencies respectively.

- For every RF signal scattered from the microbubbles (gas emboli) or surrounding tissue (solid emboli), its FFT is calculated.
- The values of the amplitudes of the spectral components at the fundamental and the second harmonic central frequencies respectively are also used as input parameters.
- The spectra of the scattered RF signal of the fundamental and the second harmonic were approximated by a Gaussian shape function using the following equation:

\[ g(x) = a_i \exp\left[-\frac{(x-b_i)^2}{c_i}\right] \]  

(9)
- The Gaussian coefficients $a_1$, $b_1$ and $c_1$ were used as a third input parameter.

Figure 2 Examples of grey scale images acquired: A. $\text{MI} = 0.2$, B. $\text{MI} = 0.6$ for two microbubbles concentrations.
Figure 3 Examples of RF signals and their corresponding frequency spectrum represented with the Gaussian approximation (dashed line): A. MI= 0.2, B. MI= 0.6.

Figure 2 shows the positions where RF signals corresponding to a solid embolus and gaseous embolus are extracted.

Figure 3 displays two examples of RF signals extracted from the obtained experimental grayscale images. Panel A presents a RF signal backscattered by a solid and a gaseous embolus at low MI (0.2).

We note that the acoustic pressure is not sufficient to start nonlinear microbubbles oscillations. Panel B displays the RF signal of each type of embolus at a higher MI (0.6). These specific signals were selected on purpose to illustrate the case where both frequency spectra exhibit a nonlinear component.

![Diagram](image)

**Figure 4 General bloc diagram of the Support Vector Machine Classification model.**

Figure 4 shows the general bloc diagram of the support vector machine classification model used in our study. For each input parameter, a vector containing the values of the parameters (amplitude or bandwidth or the Gaussian coefficients) is constructed. This vector is used as an input to the SVM model which provides in its output a value of 1 or -1 for gaseous or solid emboli respectively.

**5. 5. RESULTS AND DISCUSSION**

Table 2 and 3 summarize the percentage of correct classification of microemboli using the SVM as a function of the different input parameters and the mechanical index respectively for two concentrations of microbubbles (5µl and 10µl).
The bandwidths of the linear and non linear components for the two concentrations do not provide a correct classification at low and high MI where the average rate reached only 50% for both concentrations ($C = 5\mu l$ and $C=10 \mu l$).

When amplitudes of the fundamental and the second harmonic are introduced into the SVM model, the correct classification of microemboli at high MI ($C = 10\mu l$) is 57.14% and 85.71% respectively for both types of emboli (Gaseous and solid).

When the Gaussian coefficients are used as input parameters, we reach a correct classification rate of 96.42% at low and high MI for the concentration $C=5\mu l$ and 92.85% at low and high MI for the concentration $C=10\mu l$.

**TABLE 2** Correct classification rate of gaseous and solid emboli with concentration of microbubbles ($5\mu l$) at low MI ($0.2$) and high MI ($0.6$) for three different inputs parameters: the bandwidths, the amplitudes of the fundamental and the second harmonic, and the Gaussian parameters issued form Equation (9).

<table>
<thead>
<tr>
<th>Results with</th>
<th>Bandwidths</th>
<th>Amplitudes</th>
<th>Gaussian Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low MI</td>
<td>High MI</td>
<td>Low MI</td>
</tr>
<tr>
<td>Gaseous Emboli</td>
<td>0%</td>
<td>0%</td>
<td>85.71%</td>
</tr>
<tr>
<td>Solid Emboli</td>
<td>100%</td>
<td>100%</td>
<td>78.57%</td>
</tr>
<tr>
<td>Average rate</td>
<td>50%</td>
<td>50%</td>
<td>82.14%</td>
</tr>
</tbody>
</table>

**TABLE 3** Correct classification rate of gaseous and solid emboli with concentration of microbubbles ($10\mu l$) at low MI ($0.2$) and high MI ($0.6$) for three different inputs parameters: the bandwidths, the amplitudes of the fundamental and the second harmonic, and the Gaussian parameters issued form Equation (9).

<table>
<thead>
<tr>
<th>Results with</th>
<th>Bandwidths</th>
<th>Amplitudes</th>
<th>Gaussian Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low MI</td>
<td>High MI</td>
<td>Low MI</td>
</tr>
<tr>
<td>Gaseous Emboli</td>
<td>0%</td>
<td>0%</td>
<td>100%</td>
</tr>
<tr>
<td>Solid Emboli</td>
<td>100%</td>
<td>100%</td>
<td>71.42%</td>
</tr>
<tr>
<td>Average rate</td>
<td>50%</td>
<td>50%</td>
<td>85.71%</td>
</tr>
</tbody>
</table>

These high classification rates might be ascribed to the fact that the coefficients of the spectral envelopes contain additional information about the bandwidths and the amplitudes of the linear and non linear components of the backscattered signals from both solid and gaseous emboli.
Figure 5 Embolus classification using SVM with the concentrations of microbubbles 5 µl and 10 µl (A) at low MI; (B) at high MI.

Figure 5 shows the best model obtained for embolus classification using SVM with the two concentrations of microbubbles and for both low and high MI. In this work, we choose the threshold equal to 0 between the both classes. An output value of {1} corresponds to the gaseous embolus and a value of {-1} at the output corresponds to the solid embolus.

For example, Figure 5B, corresponding to the concentrations of microbubbles 10 µl and high MI (0.6) shows that the SVM classification model used in this experiment succeeded in classifying 14 gaseous embolus out of 14 and 11 solid embolus out of 14 also.

6. CONCLUSIONS

This study has demonstrated that the use of radio frequency signal processing could offer better classification of microemboli as solid or particulate matter than the commonly used Doppler processing. Using the support vector machines as a classifier and the fundamental and second harmonic components information’s contained in the RF signal backscattered by an embolus as inputs parameters allows the classification of embolus with a sensitivity of 96.42%.

The technique proves effective improving classification; it still needs a clinical trial and verification to demonstrate the additional benefit.
The main message of this study is to validate this strategy in a simple and a controlled experimental environment before further pre-clinical and clinical validations are undertaken.

ACKNOWLEDGMENT

The authors would like to acknowledge the support (datasets) of UMR Inserm U930 CNRS ERL 3106, Tours (France).

REFERENCES


