

# Support vector classification of proteomic profile spectra based on feature extraction with the bi-orthogonal discrete wavelet transform

Frank-Michael Schleif · Mathias Lindemann ·  
Mario Diaz · Peter Maaß · Jens Decker ·  
Thomas Elssner · Michael Kuhn · Herbert Thiele

Received: 21 December 2006 / Accepted: 10 December 2007 / Published online: 7 March 2008  
© Springer-Verlag 2008

**Abstract** Automatic classification of high-resolution mass spectrometry data has increasing potential to support physicians in diagnosis of diseases like cancer. The proteomic data exhibit variations among different disease states. A precise and reliable classification of mass spectra is essential for a successful diagnosis and treatment. The underlying process to obtain such reliable classification results is a crucial point. In this paper such a method is explained and a corresponding semi automatic parameterization procedure is derived. Thereby a simple straightforward classification procedure to assign mass spectra to a particular disease state is derived. The method is based on an initial preprocessing stage of the whole set of spectra followed by the bi-orthogonal discrete wavelet transform (DWT) for feature extraction. The approximation coefficients calculated from the scaling function exhibit a high peak pattern matching property and feature a denoising of the spectrum. The discriminating coefficients,

selected by the Kolmogorov–Smirnov test are finally used as features for training and testing a support vector machine with both a linear and a radial basis kernel. For comparison the peak areas obtained with the *ClinProt-System*<sup>1</sup> [33] were analyzed using the same support vector machines. The introduced approach was evaluated on clinical MALDI-MS data sets with two classes each originating from cancer studies. The cross validated error rates using the wavelet coefficients were better than those obtained from the peak areas.<sup>2</sup>

**Keywords** Bi-orthogonal wavelet transform · Mass spectrometry · Clinical proteomics · Support vector machine

## 1 Introduction

After the completion of the Human Genome Project, the focus of scientific interest has moved towards functional products of gene expressions. In particular, investigation of the protein and peptide-composition of cells and tissues may help to enlighten cell function, developmental processes and emergence of diseases. Proteomic pattern diagnostics is a promising tool for early disease detection [30,46] and may help to reduce the number of invasive medical procedures, such as biopsies, in the future. Matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) is a well suited method for acquisition of proteomic profile spectra from biological fluids. Magnetic bead-based enrichment of serum peptides and proteins has been shown to be a perfect tool for sample preparation prior to MALDI-TOF

---

Communicated by G. Wittum.

---

F.-M. Schleif  
Department of Mathematics and Computer Science,  
University of Leipzig, Leipzig, Germany  
e-mail: schleif@informatik.uni-leipzig.de

M. Lindemann · M. Diaz · P. Maaß  
Zentrum für Technomathematik, FB 3,  
University of Bremen, Bremen, Germany  
e-mail: pmaass@math.uni-bremen.de

J. Decker · T. Elssner · H. Thiele (✉)  
Bruker Daltonik GmbH, Fahrenheitstrasse 4,  
28359 Bremen, Germany  
e-mail: Herbert.Thiele@bdal.de

M. Kuhn  
LightTrans GmbH, Jena, Germany  
e-mail: michael.kuhn@lighttrans.com

---

<sup>1</sup> Bruker Daltonik GmbH, Bremen, Germany

<sup>2</sup> In this contribution the classifications were calculated using LIBSVM<sup>®</sup> (Version 2.8, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>) and the wavelet transform using the wavelet package of Matlab<sup>®</sup> (<http://www.mathworks.com>)